Oncogenes in *KRAS* Wild Type Pancreatic Cancer

Somayeh Ahmadloo

ORCID ID: # 0000-0002-1906-7649

Submitted in total fulfilment of the requirements of the degree of
MR-PHILMED - Master of Philosophy - MDHS (Medicine)

April 2020

Faculty of Medicine, Dentistry and Health Sciences
Department of Clinical Pathology
The University of Melbourne

Abstract

Pancreatic ductal adenocarcinoma (PDAC) is an invasive cancer, ranked the fourth most prevalent cause of cancer related death. Somatic genetic alterations are primary drivers of PDAC. 93% of patients have activating mutations in the master oncogene, *KRAS*. Several studies have investigated the mutational landscapes of pancreatic cancer. However, comprehensive studies of *KRAS* wild type pancreatic tumours are limited. Hence the process of initiation and progression of this cancer remains to be discovered and may be associated with genes that have not been identified.

The current project aims to identify oncogenes in *KRAS* wild type Pancreatic cancer. It also aims to identify whether these oncogenes are from the MAPK pathway or independent of it. The genomic and transcriptomic landscapes of *KRAS* wild type PDAC were verified. In the absence of *KRAS* mutation, alternative oncogenes were found. In the genomic data analysis, two cohorts were analysed including 70 samples in the *KRAS* wild type cohort and 571 in the *KRAS* mutant cohort. In the absence of *KRAS* mutation, tumours were found to be rare as were (i) Oncogenic *BRAF* in-frame deletions, hotspot alterations and oncogenic fusions (known and novel) (frequency of 17%), oncogenic *GNAS* hotspot mutation (frequency of 12%) and somatic alterations in *RET* (7% frequency). Recurrent copy number gains (CNV >4) were observed in *MYC* (23% frequency), *CDK6* (16% frequency), *AKT2* (16% frequency), *KDM6A* (14% frequency), EGFR (12% frequency), *RICTOR* (12% frequency), *MET* (11% frequency), *FGFR1*(10% frequency), *FGF3* (9% frequency) and *FGF4* (9% frequency) in the *KRAS* wild type cohort. Other low frequency fusion events in the MAPK pathway include: *RET-CCDC6; ROS1-SLC4A4; BRAF-SND1; BRAF-SDK1; TRIM24-BRAF; STK4-SLC13A3; ARHGAP24-MAPk10; BRAF-BRAF; STMN1-CDK5RAP3,* and*; SLC4A4-RASGRF1*.

Two independent differential expression analyses were performed on the RNA-seq of KRAS wild type versus KRAS mutant Pancreatic Adenocarcinomas, generated by the Australian and Canadian ICGC- pancreatic cancer Consortium consisting of RNA-seq from 88 and 224 bulk tumour samples respectively. Pathway analysis showed that the Calcium signalling pathway was over-expressed in both the Australian and Canadian wild type.

This up-regulation of the Calcium signalling pathway in the whole cohort of *KRAS* wild type is consistent with *GNAS* mutation in the genomic analysis of the *KRAS* wild type cohort. Additionally, the MAPK signalling pathway shows no difference throughout the whole cohort of *KRAS* wild type. Together, these findings at the genomic level reveal that the MAPK signalling pathway is the dominant pathway in the *KRAS* wild type cohort. The results of comparing RNA expression in two groups of *KRAS* wild type and *KRAS* mutant analysis suggest that one oncogene has been substituted for another oncogene in the MAPK pathway, creating an interruption in the MAPK pathway. The lack of differences between *KRAS* mutant and *KRAS* wildtype in the MAPK pathway could suggest that the MAPK pathway is up-regulated in both sets, resulting in a lack of difference in the expression.

Keywords: Pancreatic ductal adenocarcinoma; *KRAS* wild type; Oncogenes; MAPK signaling pathway; Bioinformatics; Genomics; transcriptomics.

Declaration

This is to certify that:

1. The thesis comprises only my original work towards the masters except where indicated in the Preface,
2. Due acknowledgement has been made in the text to all other material used,
3. The thesis is less than 15,000 words in length, exclusive of text in images, tables, bibliographies and appendices.

Somayeh Ahmadloo, April 2020

Acknowledgements

I would like to thank my supervisors Professor Sean Grimmond and Associate professor Oliver Hofmann for their guidance and support throughout the duration of this project. Their abilities to expand my view to help me to see the big picture of biomedical science have been of great importance during the entire proceeding of the research project.

Thanks to my good friends, who have supported my wish to pursue this project and my life for the past two years.

I would like to extend my gratitude to the staff of UMCCR, Writing Academic Skill at the University of Melbourne for their teaching, advice and assistance throughout the project.

Special thanks to my family who have been my emotional shelter, support on my wishes and endeavour during my whole life. Thanks to my brother, Afshin who has been my teacher, role model and support like a father in many aspects of my life.

I wish to thank the university of Melbourne and MDHS for the scholarship.

Preface

All of this research project was conducted by myself at the Centre for Cancer Research at the Department of Clinical Pathology, University of Melbourne under the supervisions of
Professor Sean Grimmond and Associate Professor Oliver Hofmann.

Table of Contents

List of Figures

List of Tables

Abbreviations

CGI	Cancer Genome Interpreter
CNSM	Copy Number Somatic mutations
CNV	Copy Number Variants
CPM	Count Per Million
DE	Differential expression
DEG	Differentially Expressed Genes
ExAC	Exome Aggregation Consortium
FCS	Functional Class Scoring
FDR	False Discovery Rate
GSEA	Gene Set Enrichment Analysis
ICGC	International Cancer Genome Consortium
IMA	Invasive Mucinous Adenocarcinoma
IPMN	Intraductal Papillary Mucinous Neoplasm
ITPN	Intraductal Tubulopapillary Neoplasm
KEGG	Kyoto Encyclopaedia of Genes and Genomes
MAF	Mutation Annotation Format
MAPK	Mitogen Activated Protein Kinase
MCN	Mucinous Cystic Neoplasm
PanIN	Pancreatic Intraepithelial Neoplasia
PC	Pancreatic Cancer
PCA	Principal Component Analysis
PDAC	Pancreatic Ductal Adenocarcinoma
RLE	Relative Log Expression
SNV	Single Nucleotide Variants
SV	Structural Variants
TMM	Trimmed Mean of Mvalues
TSG	Tumour Suppressor Gene
WGS	Whole Genome Sequencing

List of Appendices

Chapter 1

1 Introduction

Pancreatic cancer is an aggressive disease causing an estimated incidence of 420,000 worldwide and the mortality of 410,000 by 2020 [1]. Pancreatic cancer was ranked the fifth and fourth largest cause of cancer related deaths in Australia and the United States for 2018. Pancreatic cancer is predicted to be the second most prevalent cause of cancer deaths by 2030 [2, 3]. Pancreatic cancer is asymptomatic in the majority of patients, and at the time of diagnosis it is unresectable for approximately 80% of patients leaving only, 20% of patients whose tumours may be surgically resected. However, surgical resection of tumours will not prevent metastasis which lead to a five-year survival rate at around 8% [4]. Furthermore , pancreatic cancer patients show significant resistance to radiotherapy, chemotherapy and targeted therapy due to different resistance mechanisms, resulting from genetic and epigenetic variations, and the tumour microenvironment [5]. Pancreatic cancer (PC) is a neoplasm of the pancreas and can be subdivided into exocrine and endocrine neoplasms. Since nearly 90% of all pancreatic neoplasms are pancreatic ductal adenocarcinomas (PDACs) the term 'pancreatic cancer' will refer mainly to PDACs in this thesis.

Pancreatic cancer is mostly driven by accumulation of somatic variations and mutations possibly resulting from the activation of oncogenes and loss of tumour suppressor genes (TSGs). through which a handful of genes are altered at high frequency [6].The most common genetic variation harbours activating mutation in *KRAS* occurring in over 90% of invasive adenocarcinomas [7, 8]. Less than 10% of patients lack mutation in *KRAS* [9]. Previous studies undertaken on *KRAS* wild type pancreatic cancer were conducted on small sample sizes. Moreover, their identified oncogenes had identified genetic heterogeneity including mutations in *GNAS, BRAF* and *RAS* pathway genes as potential oncogenic drivers [10]. Hence, this project aimed to analyse *KRAS* wild type samples with bigger sample size to show which alternative oncogene or oncogenes take the role of oncogenes in the *KRAS* wild type cohort, analysing whether these oncogenes seated in the MAPK signalling pathway and considering whether the result shows some similarity between the *KRAS* wild type pancreatic cancer cohort and previous *KRAS* wild type in pancreatic cancer or other tumours with *RAS* wild type.

1.1 Thesis Structure

Having introduced Pancreatic cancer from the biological perspective, genetic alterations, review of previous studies on *KRAS* wild type in Pancreatic cancer and *RAS* wild type in other types of cancers, the remaining chapters of this thesis take the following structure:

- Chapter 1outlines the aims and scope of this project, detailing what this project wants to achieve.
- Chapter 2 includes literature review and hypotheses and aims of the thesis
- Chapter 3 describes the methods and the specific tools that were used in this project.
- Chapter 4 provides the results of the study in Genomic analysis, including SNV, CNV, fusion and transcriptomic analyses.
- Chapter 5 summarises the main findings, conclusions and future directions.

In addition to the above chapters, this thesis also includes a number of appendices and supplementary tables detailing some of the more technical aspects of the work.

Chapter 2

2 Literature Review

2.1 Cancer Biology

Pancreatic ductal adenocarcinomas (PDAC) originate from the ductal epithelial cell of the pancreas and cause pancreatic cancer in 95% of patients [11]. The chance of developing this cancer at some stage of life is 1 in 64. It is slightly more common in men (1 in 63) than it is in women (1 in 65) [12]. The median age for developing pancreatic cancer is 73 years, and it is rarely diagnosed in individuals younger than 45 years of age [13] and is age related [14]. PDAC has poor prognosis and the lethality of PDAC relates to rapid and aggressive metastasis and limited response to standard care. The median survival is less than 11 months [12].

2.1.1 Symptoms

The initial stage of pancreatic cancer is silent. Signs and symptoms are rarely noticed until the disease metastasises to surrounding organs [15]. Symptoms such as asthenia, jaundice, abdominal pain, venous thrombosis, and weight loss are diagnosed at a late stage of the disease. Digestive system problems occur in patients whose cancer blocks the route of pancreatic fluids into the intestines [15, 16].

2.1.2 Prognosis

Clinical staging of pancreatic cancer is in accordance with the American Joint Committee on Cancer's tumour–node–metastasis classification. The classification has three stages: resectable; locally advanced, and; metastatic disease (Table 1) [17]. Contrast-enhanced CT scanning can detect resectable tumours with an accuracy of 80-90% [17].

Table 1. Staging of Pancreatic Cancer [17].

Stage	Tumor Grade	Nodal Status	Distant Metastases	Median Survival†	Characteristics
				mo	
IA	T1	N0	M0	24.1	Tumor limited to the pancreas, ≤2 cm in longest dimension
IB	T2	N0	M0	20.6	Tumor limited to the pancreas, >2 cm in longest dimension
IIA	T3	N0	M0	15.4	Tumor extends beyond the pancreas but does not involve the celiac axis or superior mesenteric artery
IIB	T1, T2, or T3	N1	M0	12.7	Regional lymph-node metastasis
III	T4	N0 or N1	M0	10.6	Tumor involves the celiac axis or the superior mesenteric artery (unresectable disease)
IV	T1, T2, T3, or T4	N0 or N1	M1	4.5	Distant metastasis

* N denotes regional lymph nodes, M distant metastases, and T primary tumor.
† Data are from Bilimoria et al.[45]

2.1.3 Risk Factors

Exposure to tobacco smoke is associated with an increased relative risk of 2 to 3, with the risk being higher for heavy smokers [18]. Lifestyle factor such as cigarette smoking accounts for 25-29% incidence of pancreatic cancer with an odds ratio of 1.6 to 5.4. Cigarette smoking accumulates circulating carcinogens which bring inflammation and mutations in the proto-oncogene *KRAS* and the tumour suppressor gene *TP53* [19]. Other factors include increasing age, long-standing chronic pancreatitis, obesity, and high caloric intake.

Since obesity and high caloric intake increase inflammatory processes, it is involved in PDAC and pancreatitis. A high-fat diet in correlation with *KRAS* activation increases tumorigenesis [20]. Patients with type 2 diabetes have double the risk of PDAC compared to non-diabetics. Additionally, each 0.56 mmol/l addition of fasting blood glucose level results in a 14% increase of the incidence of PDAC [21].

Heavy alcohol consumption increases the risk of pancreatic cancer by 60% and is known to decrease the survival rate [22, 23]. Low to moderate alcohol consumption has not been found to increase the risk of pancreatic cancer. High alcohol use significantly increases the pancreatic cancer risk (OR 4.16, 95% CI 1.86-9.31) [24, 25]. In addition, pancreatic cancer is the number one cancer in terms of the frequency of *KRAS* mutation and the risk of mutation in *KRAS* increased three times in those who drink alcohol, compared to those

who do not. There is a mutational signature known to occur in heavy drinkers that correlates with alcohol consumption in pancreatic cancer [26]. Hence, *KRAS* mutations can correlate with lifestyle and environmental factors [27].

One of the identified risk factors for pancreatic cancer is chronic pancreatitis, with relative risk ranging up to 13.3. According to Chhoda et al, "Chronic inflammation is believed to promote premalignant cell survival, autocrine stimulation of a protumorigenic environment, and desmoplasia" [18]. Finally, Wolpin et al. suggests a pancreatic cancer risk increase of 17% for patients with non-O blood group [28].

2.1.3.1 Genetic Risk Factors

90% of pancreatic cancer is sporadic. 5-10 % is attributed to familial aggregation, and 3% is attributed to genetic syndromes or genetic predisposition to chronic diseases [18]. Inheritable syndromes that increase the risk of PDAC include hereditary pancreatitis (HP), Peutz-Jeghers syndrome (PJS), familial adenomatous polyposis, familial atypical multiple mole melanoma syndromes, Li-Fraumeni syndrome, hereditary breast and ovarian cancer syndrome, and ataxia telangiectasia [18, 23]. Having a pancreatic cancer patient as a first degree relative increases the chance of developing the disease by 2-5 fold [29]. Moreover, several genetic syndromes and germline mutations are associated with pancreatic cancer such as Peutz-Jeghers syndrome with *STK11* mutation, [30] hereditary pancreatitis with *PRSS1* mutation [31], hereditary nonpolyposis colorectal cancer syndrome with *MLH1* mutation [32] and breast cancer with *BRCA2* mutation [33]. These rare syndromes and germline mutations contribute less to the overall incidence of pancreatic cancer, but their germline variation can provide some information for understanding the pattern of pancreatic cancer formation.

2.2 Histopathology of Pancreatic Cancer

PDAC originates from non-invasive precursor lesions including cystic lesions (mucinous cystic neoplasm (MCN), intraductal tubulopapillary neoplasm (ITPN) and intraductal papillary mucinous neoplasm (IPMN)) and a noncystic lesion (pancreatic intraepithelial neoplasia (PanIN)) [34].

The World Health Organisation classifies PDAC precursor lesions (PanIN, IPMN, and MCN) into three grades of dysplasia: low, intermediate, and high grade [35]. Clinical investigation has shown a lower risk of malignancy for low and intermediate grade lesions while there is a high risk of metastatic tumour for high-grade lesions, with a corresponding need for surgical intervention [34].

2.2.1 Pancreatic Intraepithelial Neoplasia

PanINs are small mucinous lesions in the pancreatic ducts with a size of less than 5 mm. According to Ren, et al, PanINs of low grade have “flat, mucinous epithelium with bland to atypical nuclei, nuclear stratification, crowding, and hyperchromasia while high-grade PanIN include flat to a papillary, micropapillary, or cribriform formation with severe nuclear atypia, loss of polarity, macronucleoli, and abnormal mitotic figures” [34] (Figure 1B). In some patients, the distinction between PanIN and dilated neoplastic glands of PDAC is difficult. Hence, in high-grade PanINs, sufficient sampling and detailed histological examination are highly recommended [34].

2.2.2 Intraductal Papillary Mucinous Neoplasm

IPMNs are characterised as cystic neoplasms with the size of greater than 1 cm. From a radiologic perspective, they are cystic lesions, involving the pancreatic ducts. The main and/or branch ducts are dilated, while atrophy is found in surrounding pancreatic parenchyma. According to the contribution of pancreatic ducts, IPMNs are categorised into the main-, branch-, and mixed-duct types. According to the type of mucin that IPMNs express, they are categorised into different subtypes [34] (Figure 1 C-F).

2.2.3 Mucinous Cystic Neoplasm

MCN encompasses a single multilocular cyst. MCN differs from IPMN in many aspects. MCNs are separated from pancreatic duct system. In addition, they are located at the distal part of the pancreas. Radiologic characteristics of MCNs contain one cyst with peripheral calcifications [34].

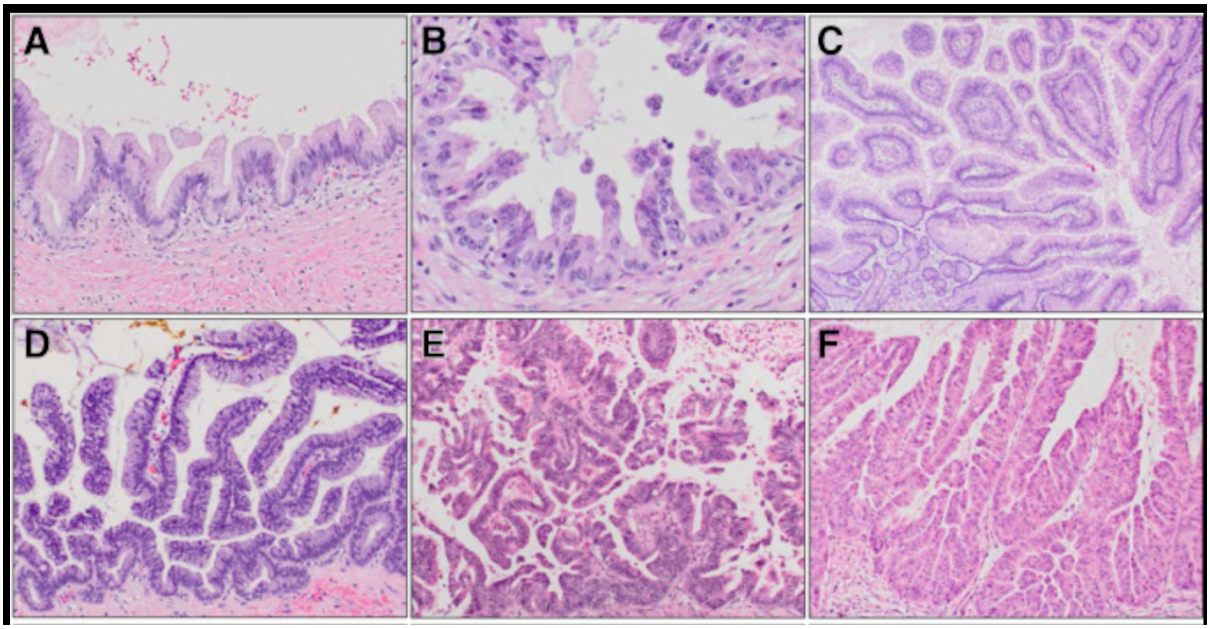


Figure 1. PanIN and IPMN subtypes. A and B: PanIN: low grade PanIN (A) high grade PanIN (B). C–F: Different Subtypes of IPMN according to their origin: gastric (C), intestinal (D), pancreatobiliary (E), and oncocytic (F) types [34].

2.2.4 Intraductal Tubulopapillary Neoplasm

ITPN is characterised as an intraductal epithelial neoplasm which occurs rarely. ITPN and IPMN can not be diagnosed by imaging. Generally, 95% of ITPNs originate from the main duct of the pancreas. ITPNs have a bottle-cork shape in MRI cholangiopancreatography and endoscopic investigation. The average size of ITPN is 3 cm and range from 1 to 15 cm [34].

2.3 Somatic Alterations in Pancreatic Cancer Precursor Lesions

Mutation in driver genes is not enough for cancer development. Those cells which escape from apoptosis and immune suppression and have initiating driver mutations, start to divide. The division creates a clonal expansion of primary variation. Clonal expansion occurs through either stepwise evolution or punctuated evolution model [36].

PDAC occurs from non-invasive precursor lesions. The precursor lesion PanIN is more common than any other in pancreatic cancer, and it is graded according to the degree of architectural and cytologic alterations [37, 38]. As PanIN grade increases, *KRAS* mutation also increases. In other words, low-grade PanIN is a combination of *KRAS* wild type and *KRAS* mutant cells [39]. Inactivation of *CDKN2A* has been found in more than 70% of high-grade PanIN, while not being common in low-grade PanIN. High-grade PanIN and invasive PDAC are often found to have mutations in *SMAD4* and *TP53* [37]. Aside from somatic point mutations, telomere shortening occurs at an early stage of tumorigenesis and is found at low-grade PanIN frequently [40]. Copy number changes occur frequently in high-grade PanIN and in well-known driver genes. Both low and high-grade PanINs have been found to contain chromothripsis-like locations, suggesting that it is involved in the progress to PDAC metastasis [41] (Figure 2).

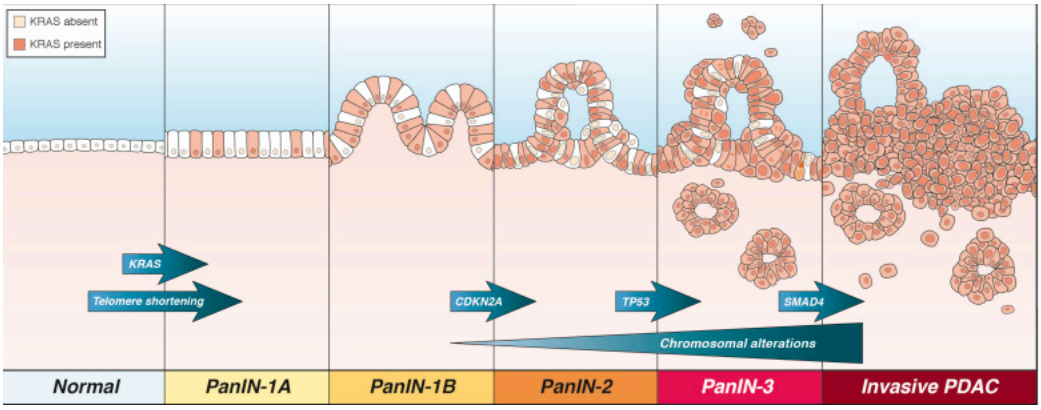


Figure 2. The PanIN progression model [42].

Pancreatic cancer can also occur from IPMN. IPMNs include the pancreatic duct system with a size greater than 1cm [43]. In addition to alterations in *KRAS, CDKN2A, TP53*, and *SMAD4* which are restricted to IPMN-invasive, IPMNs have hotspot mutation in *GNAS* and inactivating mutation in *RNF43. GNAS* mutation happens at an early stage of IPMN and becomes more common in invasive IPMNs. Mutation of *RNF43* in IPMN tumorigenesis has not been established [44, 45]. The gain of function *GNAS* mutations have also been observed in the early stage of IPMNs. *GNAS* mutations were found in invasive adenocarcinomas originate from IPMN. *GNAS* mutations have been found in low-grade, high-grade and invasive adenocarcinomas. It seems that *GNAS* mutations contribute to initiate the pathogenesis in IPMN [46-48] (Figure 3).

Intraductal tubulopapillary neoplasms (ITPNs), lead to 0.9% of all exocrine neoplasms and 3% of intraductal neoplasms [49]. This type of neoplasm has a specific molecular profile, in particular, a rare mutation in *KRAS, NRAS,* and *GNAS*, but have mutations in *PIK3CA* and *AKT* [50]. In some cases, *TP53* overexpression and loss of *CDKN2A* and *SMAD4* have been reported [49].

Mucinous cystic neoplasms (MCNs) have some genetic similarity with IPMN with regard to *KRAS* mutation but lack *GNAS* mutation. In addition, they have a mutation in *RNF43* at a later stage. MCNs have a specific ovarian-type stroma [45].

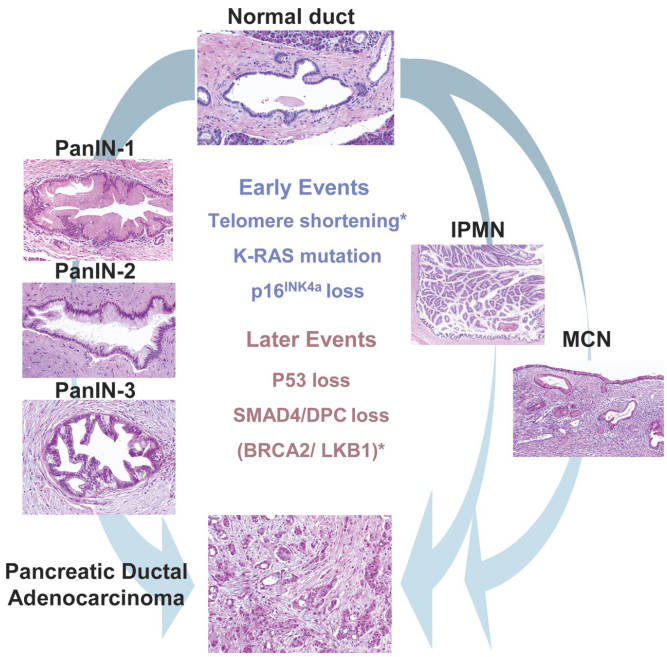


Figure 3. Pancreatic cancer precursor lesions. Pancreatic cancer can originate from IPMN, PanIN or MCN [51].

The tumour microenvironment (TME) also makes a critical contribution to the pathogenesis of this cancer. Pancreatic tumours are composed of tumour cells, fibroblasts, immune cells, and endothelial cells. Genetic variations are not common in the stromal cells. Therefore, it is likely epigenetic modifications affect the phenotype of fibroblasts due to evidence that the tumour cells are capable of inducing DNA methylation in the suppressor of cytokine signaling1 (SOCS1) gene in cancer-associated fibroblasts (CAFs) which promote cell growth [52]. It is likely that oncogenic signalling mediated by *KRAS* needs pancreatic stroma for being involved in pathogenesis [53]. Also, activated *KRAS* in tumour cells has been shown to stimulate Hedgehog signalling in the surrounding stroma, thereby mediating fibroblast activation [54].

2.4 Genetic Alterations in PDAC

The accumulation of somatic alterations drives pancreatic cancer. Loss of tumour suppressor genes (TSGs) and activation of oncogenes contribute in tumorigenesis of all cancers. While only a handful of genes are altered at high frequency in PDAC, recent whole-genome surveys demonstrate genetic heterogeneity in PDAC with the occurrences of rare somatic mutations in many genes among different patients [55].

2.4.1 Map Kinase Pathway Drivers

The *KRAS* gene is located at 12p. It is the most commonly mutated gene in PDAC, occurring over 90% of invasive adenocarcinomas. Mutations in *KRAS* have been shown to frequently occur in pancreatic premalignant lesions (from PanIN1 stage). The most common *KRAS* mutations are single amino acid substitutions from Gly to Asp at codon 12 that result in a gain of function mutation [56]. *KRAS* binds to GTP and functions as a molecular switch for numerous signalling pathways. Activated *KRAS* binds to GTP and further initiates an array of downstream effectors including the RAF/MEK/ERK, phosphoinositide-3 kinase/mTOR and RalGDS pathways. These pathways are involved in inhibition of apoptosis, progression of the cell cycle and cytoskeletal alterations. In the absence of exogenous signalling, *KRAS* hydrolyses GTP to GDP, subsequently turning itself off. However, activating *KRAS* mutations impair GTP hydrolysis resulting in constitutive activity [57].

BRAF, located on chromosome 7q, is a *KRAS* downstream effector. *BRAF* encodes a serine/threonine protein kinase that is involved in regulating the Ras-Raf-MEK-ERK-MAP kinase pathway which affects cell division, differentiation, and secretion. *BRAF* and *KRAS* mutations are mutually exclusive in PDAC where 33% of the PDAC with wild type *KRAS* contain mutations in the *BRAF* gene [58]. Although *BRAF* mutations do not represent a high-frequency defect in PDAC, it has been suggested that *BRAF* functionally substitutes for *KRAS* mutations during PDAC initiation. Recent emerging evidence shows that expression of mutant BRAF (V600E) in the mouse pancreas leads to PanIN lesions, further emphasising the role of *BRAF* mutations in pancreatic cancer [59].

The catalytic subunit of p110 in phosphoinositide 3-kinases (PI3K) is another *KRAS* downstream effector. Activated PI3K induces a phosphorylation series that involves the Ser/Thr kinase AKT/protein kinase with major roles in cell survival and cell growth. Observations further confirm the importance of the PI3K-AKT pathway in pancreatic cancer that cause constant activation of this pathway [60]. Moreover, amplification of the AKT2 kinase, a downstream effector of PI3K complex, is documented in 32% of PDAC patients [61].

Another Ras effector, RalGDS is a guanine nucleotide exchange factor, which activates Ral-GTP proteins, RalA and RalB. Activated Rals initiate downstream signalling events that regulate many cellular processes including cytoskeleton remodelling, vesicular transportation and transcriptional activation [62]. RalA is activated in pancreatic cancer. RalA and RalB appear to have distinct roles in pancreatic carcinogenesis, with RalA promoting tumorigenicity and RalB promoting metastasis [63].

2.4.2 Overexpression of Receptor Tyrosine Kinases

Transmembrane receptor tyrosine kinases have four family members including *ErbB1* or *EGFR, ErbB2, ErbB3* and *ErbB4.* Alteration in the expression of ErbB receptors, especially ErbB1 and ErbB2 have an essential role in development and malignancy in a variety of human cancers [64]. Signalling cascades resulting from homo- and heterodimerisation of RTK receptors ultimately activate downstream of many signalling pathways such as PI3K/PTEN/Akt/mTORC1/GSK-3, Ras/Raf/MEK/ERK, Jak /STAT and much other signalling [65].

EGFR overexpression has been reported in 40-70% of PDAC tissue as compared to normal [66] while HER2 overexpression varies from 16-69% across different studies, it is a prognostic factor and its overexpression correlates with a lower survival rate in PDAC patients. Amplifications of *HER2* or *EGFR* have been reported in less than 2% PDACs and do not correlate with patient survival, suggesting that HER2 and EGFR deregulation takes place either on the transcriptional or translational level in pancreatic cancer [67].

2.4.3 *GNAS* and G-protein Coupled Receptor

GNAS is located on chromosome 20q13 [68]. It encodes the Gα of heterotrimeric G protein. The G-protein mediates signal transduction from the G-protein-coupled receptor [69]. Attachment of ligand causes stimulation of GPCRs that lead to the exchange of GDP to GTP on Gα. Activated Gα affects Adenylate cyclase causing conversion of ATP to cAMP. cAMP in turn leads to activation of protein Kinase A(PKA). *GNAS*-CAMP signalling keeps quiescence or cellular differentiation in many organs [69].

One of the pathways that become activated via cAMP is calcium signalling [68]. Ca^{2+} has a major function in pancreatic cells. It regulates the secretion of exocrine and endocrine cells [70]. In addition, Genome-wide association by Tang et al. showed in pancreatic cancer, that the calcium signalling pathway correlates with *GNAS* mutation. *GNAS* mutation relates to diabetes-associated pancreatic cancer, through the alteration in cAMP signalling pathway or increasing resistance to insulin [71].

2.4.4. Loss of Tumour Suppressors *p16/CDKN2A, SMAD4* and *P53*

p16/CDKN2A is the most frequently inactivated TSG in PDAC. The consequence of p16/CDKN2A inactivation is increased cell proliferation due to aberrant G1-S progression. *CDKN2A* negatively regulates G1-S cell cycle progression through inhibiting *CyclinD/CDK4/6* complexes that phosphorylate retinoblastoma protein (RB). The loss of p16/*CDKN2A* results in increased RB protein phosphorylation and subsequent release of the E2F transcription factors that fast-forwards progression through the cell cycle. p16/CDKN2A is inactivated in PDAC through homozygous deletions (40%), intragenic mutations with the heterozygous loss (40%) or hypermethylation of the p16 promoter (15%) [72, 73]. *p16/CDKN2A* aberrations are first seen in 30% of early PanIN lesions, increase in frequency with disease progression and are reported in >90% of invasive adenocarcinomas [74]. p16/*CDKN2A* loss during the PanIN progression suggests a synergy with *KRAS* activating mutations. Indeed, evidence from mouse models demonstrates that p16/*CDKN2A* inactivation and activating *KRAS* mutations cooperatively promote PDAC progression and metastasis [75].

SMAD4 is a tumour suppressor gene located on the long arm of chromosome 18. In response to TGFβ signalling, *SMAD4* is activated by transmembrane receptor kinase and forms heterodimer complexes with *SMAD2* and *SMAD3*. *SMAD* complexes are then translocated to the nucleus to regulate transcription of target genes. The consequence of *SMAD4* loss is attenuated growth inhibition through the loss of pro-apoptotic signals or loss of cell cycle arrest [76]. *SMAD4* loss is first seen in 31% of PanIN3 lesions and 55% of invasive adenocarcinoma. Loss occurs through homozygous deletions in 30% of patients or single allelic loss coupled with intragenic mutations in 25% of patients [77]. Mouse models of PDAC have depicted SMAD4 loss in the context of *KRAS* activating mutation significantly accelerates the progression of mouse PanINs [78].

The tumour suppressor gene, *TP53* is also altered in up to 85% of human PDAC, mostly through a different mechanism of nonsense mutations, frameshifts and homozygous deletion [79]. *TP53* controls programmed cell death and cell cycle arrest through several key molecules such as 14-3-3σ (G2 arrest), p21 (G1-S progression), PUMA and NOXA (apoptosis). Since *TP53* protein protects cells from DNA damage and uncontrolled cell growth, it has been known as a "guardian" of the genome. It is the most common somatic mutation in human cancer [80]. Thus, uncontrolled cell growth, increased cell survival, and genomic instability are all possible consequences of p53 aberration in pancreatic cancer. *TP53* loss in mouse pancreas expressing mutant *KRAS* accelerates the PDAC progression through suppression of Ras-induced senescence. In addition, compared to knockout *TP53* only, mice expressing mutant P53R172H develop metastatic pancreatic cancer, suggestive of differential roles of TP53 mutations in PDAC carcinogenesis [81].

2.5 Genetics of *KRAS* Wild Type Pancreatic Cancer

KRAS wild type tumour analysis has established considerable genetic heterogeneity. Some studies have found that this group of patients have copy number amplifications or fusions in oncogenes such as *RET, ROS1, HER2, NRG1, BRAF* and *ALK* [14, 37, 82, 83].

The absence of *KRAS* mutation has been identified in some of the PDAC studies including Raphael et al. [84], Barret et al.[85], Witkiewicz et al.[86], Shimada et al.[87], Sausen et al.[88], Heining et al.[89], Petricoin et al. [82] and Wolpin et al. [90]. However, there is

not consistency between all of these studies and their analysis was based on small numbers of *KRAS* wild type samples.

One study by Raphael [84] et al and another by Witkiewicz [86]et al have reported the following as alternative oncogenes of KRAS: *BRAF*, *GNAS, PI3KCA,* and *NRAS*. Further, Raphael and his colleagues have suggested that the *KRAS* wild type group of samples are significantly enriched in the TSC/ mTOR signalling pathway compared with *KRAS* mutant samples (10/150, p = 0.0007) [91]. It is important to note that *PIK3CA* and *BRAF* were the first oncogenes identified in the study by Witkiewicz et al., and in the absence of these mutations, other cancer genes were suggested as drivers, such as *STK11, GNAS, CHEK2* and *RB1* [86]. Pishvaian et al. analysed 616 PDAC samples via cancer gene panel analysis among which, 81 samples were *KRAS* wild type. These KRAS wildtype samples had a mutation in other MAPK signalling pathway genes including four *NRAS* and fourteen *BRAF* alterations. Interestingly, 37 KRAS wildtype samples lacked adenocarcinoma histology and 33 a further 33 had no mutation in other components of the MAP kinase pathway. [82]. Recently Grinshpun et al. reported 3 cases of *KRAS* wild type with mutations in *RET* and *BRAF* [92].

Oncogenic Fusions are well known contributors in the pathogenesis of *KRAS* wild type. Heining and his colleagues recently identified oncogenic *NRG1* fusions in all four *KRAS* wild type patients of the PDAC cohort while *KRAS* mutated tumours lack these kinds of rearrangements [89]. The exciting study by Shimada et al. discovered two rare fusion aberrations from the analysis of 100 cases of PDAC ALK (DCTN1-ALK) and *RRAS* (Q87L) that drive pancreatic cancer pathogenesis independent of *KRAS* mutation [87] and therefore implementing KRAS wildtype involvement.

Copy number amplification in some oncogenes is important in the *KRAS* wild type group. Wolpin et al. investigated 71 PDAC samples. Seven samples had no *KRAS* mutation, but showed variation and copy number changes in MAPK pathway oncogenes, including amplification in *FGFR1*, structural variations in *ROS1* and *BRAF* variations [90].

MAP Kinase and Wnt signalling are two pathways that were suggested to be activated in the *KRAS* wild type group. In the absence of *KRAS* mutation, in *KRAS* wild type PDAC, hotspot mutation in *BRAF* [93] and *ERBB2* [94] have been shown to activate the MAPK

signalling pathway. Additionally, Raphael et al. have shown *KRAS* wild type PDAC to have oncogenic *CTNNB1* mutations and highlight the importance of Wnt signalling in this group of PDAC [84].

The above findings demonstrate alternative oncogenes in the absence of *KRAS* mutation, but in all of the previous studies the sample sizes were small and there was little similarity between introduced genes. Therefore, we have recruited more samples from a published database collated by the International Cancer Genome Consortium [95] to increase the power and hence robusticity of conclusions regarding the cohort of *KRAS* wild type.

The above mentioned, ICGC, has been used to create the atlas for many cancer types. Creating a public database has facilitated diagnosis, treatment, and prevention of different cancer types. In 2008 the International Cancer Genome Consortium (ICGC) started a collaborative project between multiple institutes to identify and characterise somatic mutation in 50 types and subtypes of cancer. The release 28 in March 2019, contain 86 Cancer projects, 22 Cancer primary sites, 22,330 Donors with molecular data in DCC, 24289 total donors, 81,782,588 simple somatic mutations and 57,905 mutated genes. The data from March 2019 release of ICGC Data Portal was used in this thesis.

2.5.1 *KRAS* Wild Type Mutational Status Confers Overall Survival Advantage

KRAS mutation status can be applied as a prognostic biomarker for the clinical status of PDAC. Studies by *Kim* et al. [96], Shroff et al. [97], Strumberg et al. [98], and Boeck et al. [99] have revealed that *KRAS* wild type patients have demonstrated a better objective response, longer overall survival, and a lower risk of death compared to *KRAS* mutant patients. In general, this longer survival rate and improved prognosis in *KRAS* wild type patients suggests that clinical trials in pancreatic cancer should consider the *KRAS* mutation status when utilising novel therapy [97]. This finding has some similarity to other types of cancer such as lung cancer. Guan et al. demonstrate that *KRAS* wild type groups have longer survival rate than *KRAS* mutant [100].

2.6 *RAS* Wild Type in Other Cancers

The absence of *RAS* mutation has been demonstrated in other cancers such as melanoma and lung cancer. Fusions in some oncogenes are important in driving tumorigenesis. Shin et al. found *NRG1* fusions in the absence of known driver mutations in samples with invasive mucinous adenocarcinoma (IMA) [101]. *NRG1* is one of the ligands that bind to *ERBB*3 and cause activation of ERBB RTKs [102]. In the study by Shi et al., it was found that 24 Gastrointestinal stromal tumour (GIST) samples did not have any mutation in KIT/PDGFRA/RAS pathways. Instead, two tumours had fusions in *FGFR1* genes including *FGFR1–HOOK3, FGFR1–TACC1* and *ETV6–NTRK3* fusion [103]. *CCDC6-RET* fusion is rare fusion which was reported in 2 cases of the study by Le Roll et al. in the absence of other well known driver mutations in colorectal cancer such as *KRAS, NRAS, PIK3CA* and *BRAF* [104].

Somatic alterations such as SNV, SV and CNV in the components of the MAP kinase pathway have been shown to drive tumour progression. Nikolaev et al. and Nissan demonstrated alternative oncogenes from the MAP kinase signalling pathway such as *MAP2K1, MAP2K2*, and loss of *NF1* [105, 106]. Rau et al. reported that half of the traditional serrated adenoma (TSA) patients with *KRAS* wild type have mutation in *BRAF*.[107]

Copy number amplification in MAPK components causes activation of the MAPK signalling pathway. El-Deiry et al. reported a *HER2*/neu amplification in *KRAS* wild type tumours of colorectal cancers [108]. Another study by Ross et al. reported that more than 5% of left colonic tumours with *RAS* wild type have *ERBB2* amplification [109].

Teng et al. identified *FAM47C* mutations in *KRAS* wild type colorectal cancer (4/10 tumours). The frequency of *FAM47C* mutation in the COSMIC and TCGA databases were 5.71% and 5.41%, respectively [110]. *FAM47C* encodes a protein which belongs to a family of proteins whose function is unknown [111].

Taken together, these studies have demonstrated alternative oncogenes in the absence of *RAS* mutation in other types of cancers. Moreover, some of the introduced genes were mapped in the MAP kinase signalling pathway. Therefore, analysis of *KRAS* wild type

samples may reveal which alternative oncogene or oncogenes are responsible for tumourigenesis in the *KRAS* wild type cohort. Analysis should also consider whether the result shows some similarity between the *KRAS* wild type pancreatic cancer cohort and other tumours with *RAS* wild type.

Other wildtype tumours such as RAS wildtype have been associated with the oncogenes *BRAF* and *RET* including thyroid and colorectal cancers. Hence examination of similarities between the KRAS and RAS wildtypes may contribute further evidence to the role of these oncogenes in role in wildtype PDAC.

2.7 Hypothesis and Aims

The previous whole genomic sequencing studies that identified a number of novel genetic mutations in *KRAS* wild type PDAC were small in scale (less than 12 wildtype tumours). This study is the largest meta-analysis of *KRAS* wild type pancreatic tumours to date (n=70 tumours from ICGC and TCGA studies) through integration of the previous studies. Therefore, the current study is more comprehensive than the previous studies. Additionally, of those genes identified, mutations have not been separated into either "driver" mutations with candidate roles in *KRAS* wild type PDAC or into functionally non-consequential "passenger" mutations. The current study assumes that there is a distinct mutational landscape associated with *KRAS* wild type in pancreatic cancer.

2.7.1 Research Question

How do the mutational landscapes of *KRAS* wild type and *KRAS* mutant Pancreatic Adenocarcinoma differ?

2.8 Hypotheses

Addressing the challenges mentioned above raises the following hypotheses that will form the core of this thesis:

1. An alternative candidate oncogenic event drives tumorigenesis in PDAC lacking an activating *KRAS* mutation.
2. This candidate oncogenic event will involve another component of the MAPK pathway.
3. The mutation will occur in a MAPK independent pathway.

2.9 Aims

The current study was designed to use different bioinformatic tools to analyse somatic mutations, copy number variations, structural variants and RNA sequencing by integrating data from multiple sources to dissect the biological significance of oncogenic drivers, fusion genes and relating pathological pathway in *KRAS* wild type Pancreatic cancer.

The aims of the current project include:

1. To determine what candidate oncogenic events drive initiation and progression of tumours in *KRAS* wildtype by looking at somatic mutations, copy number variations, and structural variants through integrating data from multiple sources.

2. To characterise the underlying Pathway driving the *KRAS* wild type.

2.10 Research Methods Outline

Determine the characteristic mutational landscape associated with *KRAS* wild type vs. *KRAS* mutant pancreatic tumours

1. Review the pattern of *KRAS* mutations in published PDAC studies by integrating data from multiple sources.

2. Determine the mutational landscape of *KRAS* wild type vs. *KRAS* mutant tumours.

3. Determine the pattern of driver vs. passenger mutations in each *KRAS* wild type tumour.

4. Identify whether gain of function MAPK pathway mutations are enriched in *KRAS* wild type tumours.

Chapter 3

3 Methods

3.1 Patients and Samples

Seventy cases of pancreatic cancer were selected from a *KRAS* wild type cohort with the inclusion criteria of no activating mutations in the *KRAS* gene. They were identified in a publicly available study from the International Cancer Genome Consortium. Thirty-four samples were derived from the Australian-ICGC pancreatic cancer study, twenty-two were from the Canadian-ICGC study and fourteen samples were from ICGC- US cohort. In the study of the Australian sample sets, three samples were not in the ICGC database. Hence, I received the input files for analysis of SNV, CNV and SV for the Australian group from my supervisors. The input was uploaded onto GitHub. A contrasting *KRAS* mutant cohort having mutation in *KRAS* codons 12, 13, 61 was compiled from the Australian, and Canadian ICGC pancreatic cohorts including 335 samples from the Australian cohort and 236 from the Canadian cohort. These two cohorts were used to assess relative enrichment of somatic events between *KRAS* wild type and mutant pancreatic cancers.

Other than three Australian samples, all data was extracted from a cohort of ICGC databases. The IDs of the Australian and Canadian samples in *KRAS* wild type and *KRAS* mutant are in appendix 1-2. The analysis was accomplished in R package version 3.5.1 (2018-07-02).

The script for analysis of SNV and CNV was deposited on GitHub (https://github.com/umccr/SNV) and (https://github.com/umccr/CNV).

The following section describes the tools that were used and the inputs and outputs that were generated from the tools.

3.2 Maftools

To compare SNVs between *KRAS* wild type vs. *KRAS* mutant, the first step was to apply the Maftools package [112]. Maftools (Mutation Annotation Format) is one of the Bioconductor Packages that analyses somatic variants. This tool provides multiple analysis, summary, annotation and visualisation of data. It requires a standard tab delimited text file which contains all somatic variations in the cohort [113]. This tool provides multiple analysis, summary, annotation and visualisation of data [112]. It requires a standard tab delimited text file which contains all somatic variations in the cohort. In the analysis of *KRAS* mutant, the simple somatic mutation file for *KRAS* mutant was extracted from the ICGC database for the Australian and Canadian cohorts as a tsv file and underwent Maftools analysis. In addition, the simple somatic mutation file for the Canadian and American groups of *KRAS* wild type cohort was extracted from the ICGC database as a tsv file and analysed with Maftools. A script by Jacek Marzec was used for making the Maf files and making plots which were then deposited on GitHub (https://github.com/umccr/MAF-summary).

3.3 BEDTools

Another approach in the analysis of CNV and SV is BEDTools. BEDTools which was developed by Quinlan et al. to compare, manipulate and annotate genomic features. The formats that are widely used in this software for genomic features are Browser Extensible Data (BED) and General Feature Format (GFF) [114]. This tool is used to find overlap between coordinates and genes. In order to find coordinates the UCSC Genome Table Browser was employed [115]. In addition, if two genomic features had at least one base pair in common, it was nominated as ‘intersect’ or ‘overlap’[114].

In identification of copy number variation (CNV) and structural variants (SV) from the ICGC, the genomic position was sourced from the input file that was obtained from the copy number somatic mutations (CNSM) and structural somatic mutation (STSM) sections of the database. This data was obtained on 25 March 2019 from ICGC. In order to find the genes that coordinated with that genomic position, the data from UCSC genomic Table browser build 37 was extracted on 25 March 2019 as BED format (https://genome.ucsc.edu/cgi-bin/hgTables) and then overlapped with CNV and SV

coordinates obtained from ICGC database. All overlaps that had at least one base pair were retained and then used as input for the next step analysis. The command line of BEDTools was run on Linux. (https://bedtools.readthedocs.io/en/latest/content/tools/intersect.html). The command line is provided in Appendix-3.

3.4 Cancer Genome Interpreter (CGI)

One of the tools that was used in the present research is CGI. CGI webtool reviews somatic point mutations, small insertions/deletions, copy number alterations and/or gene fusions of an individual and predicts those likely to be drivers versus passengers. It also summarises any evidence of the mutations influencing drug response. CGI identifies all known and likely genomic alterations including the analysis of variants of unknown significance. CGI predicts driver mutations on the contribution of every gene and its region in specific cancer types. The prediction is obtained from the analysis of a large cohort of tumours. In addition, when there is no available information for specific cancer types, CGI uses Pan-Cancer information on the corresponding gene [116]. That is to say, CGI works on the individual sample and takes each sample one at a time and then for every event within the patient, it screens the entire set of variants and identifies the genes that are likely to be drivers based on the type of mutation and prior knowledge. Moreover, the identification of known driver alterations and the biomarkers of drug response are applied by making a match between the tumour type in which they have been described and one of the samples under analysis. CGI interprets mutations from any sequencing platform including exome, whole genome, gene panel sequencing, and VCF [116]. CGI screens the entire set of variants and identifies the genes that are likely to be drivers or passengers based on type of mutation and prior knowledge and application of the Catalogue of Validated Oncogenic Mutations. CGI also assesses the tumorigenicity of variants of unknown significance via OncodriveMUT [116]. The details about how Oncodrive MUT classify TIER1 and TIER2 are summarised in Table 2. OncodriveMUT uses prior information about each mutation from large tumour analysis and location of mutation for these classifications.

Table 2. OncodriveMUT classification of variants as a potential oncogene. OncodriveMUT use prior knowledge about each mutation from large tumour analysis and location of mutation for these classifications. [117].

Consequence type	Gene Category	Condition	Driver Prediction	Description
Missense	Tumour driver, other tumours driver	Mutation in a gene cluster	Tier 1	Custer-missense
Missense	Tumour driver	CAAD > 25	Tier 1	Functional- missense
Missense	Other tumours driver	CAAD > 30	Tier 1	Functional- missense
Missense	Other tumours driver	CAAD > 25	Tier 2	Functional- missense
Missense	Tumour driver, other tumours driver	Mutation in a delicate domain and CAAD > 20	Tier 2	Functional- missense
Disrupting	Tumour driver, other tumours driver	LOF gene and the mutation is not in the distal protein portion	Tier 1	Loss of function-disrupting
Disrupting	Tumour driver, other tumours driver	Gene with ambiguous role and the mutation is not in the distal protein portion	Tier 2	Loss of function-disrupting
Disrupting	Tumour driver, other tumours driver	Mutation in a gene cluster	Tier 2	cluster -disrupting
Inframe indel	Tumour driver, other tumours driver	LOF (or ambiguous) gene and mutation in a delicate domain and CAAD > 25	Tier 2	loss of function-inframe
Inframe indel	Tumour driver, other tumours driver	LOF (or ambiguous) gene and mutation in a delicate domain and CAAD > 20	Tier 2	loss of function-inframe
Inframe indel	Tumour driver, other tumorous driver	Mutation in a gene cluster	Tier 2	Cluster- inframe

3.4.1 SNV Analysis

After extracting the simple somatic mutations from a tsv file of the Canadian and American *KRAS* wild type group from ICGC and a MAF file of the Australian group from my supervisors, the three SNV inputs were merged in R to make a SNV dataframe. The columns ‘chromosome,’ ‘reference allele,’ ‘alteration allele,’ ‘chromosome position’ and ‘sample name’ were extracted and entered into CGI as input data. In the *KRAS* mutant cohort, 571 *KRAS* mutant simple somatic mutations were extracted from the ICGC database in tsv format. The data was imported into R and the same columns were extracted for the *KRAS* mutant cohort and run separately in CGI.

3.4.1.1 Pearson's Chi-Squared Test

Pearson's chi-squared test is a test widely used for comparing association between two categorical variables [118]. Pearson's X^2 is applied to analyse whether the categorical variables in one group are distributed independently from other groups. Pearson's test has

two hypotheses including null and alternative. The null hypothesis or H_0 is when there is independency and the H_1 hypothesis is when there is dependency [119]. This test was applied in one section of SNV analysis to compare whether the frequency of top list of candidate genes in *KRAS* wild type is significantly different from *KRAS* mutant.

3.4.2 Copy Number Variation Analysis

CNV analysis by CGI requires gene symbols and an indication of whether they are amplified or deleted. Analysis was conducted at high gain copy number > 4. Also, homozygous deletion (copy number = 0) of genes was analysed using the cancer genome interpreter [116]. The coordinated gene was identified from the chromosomal positions by using BEDTools that was explained in the previous section (4.3).

For the *KRAS* wild type cohort, the samples including the Canadian and American groups as well as the *KRAS* mutant Australian and Canadian cohort were extracted from the ICGC database as Copy Number Somatic Mutation files. Additionally, the input CNV file for the Australian group that I received from my supervisors was used for the analysis.

In the file that was extracted from ICGC, the dataframe of homozygous deletion was created by subsetting the 'mutation type' column to 'loss.' The dataframes of copy number amplification were created by subsetting of column 'mutation type' to 'gain.' In the next step, copy number gain was converted to data frames to those that have copy number > 4.

In order to find the genes that were coordinated to each chromosomal location, the columns of 'chromosome,' 'chromosome start,' 'chromosome end' and 'sample id' were extracted as input and run on Linux by using the command line (Bedtools intersect , Appendix-3) . After providing the gene name for coordinated position from Linux command line, the columns of 'gene name,' 'sample' and additional column of 'AMP/DEL' were added to the analysis of the Australian, Canadian and American cohorts. The input containing 'sample id,' 'gene name' and 'amplification or deletion' columns was run on CGI.

3.5 Fusion Hub

In the Fusion analysis, one of the web platforms used was Fusion Hub. Fusion Hub is a web platform for annotating and visualising fusions. It was developed on the Linux platform using Perl, R, PHP, and HTML. In addition, Fusion Hub offers annotation and visualisation of gene fusion events by using three ways of visualisation: a circular view, a domain architecture view, and a network view. In the 'FusionSearch' section, a Fusion database exists which is a collection of 10 public domain databases and 14 datasets from literature. In total, 24 datasets are available in the analysis by Fusion Hub. [120]. The circular view section visualises intra- as well as inter-chromosomal fusion by providing a chromosome map. The domain architecture illustrates the way head and tail genes integrate showing exon and functional domains. The network section indicates dynamic partner gene networks. The summary table is a tab delimited file which is downloadable. The results of Fusion Hub predictions come from the reports of gene fusions from the prior literature [120].

To obtain the combined Australian and Canadian *KRAS* wild type cohort, the tsv structural somatic mutation file for the Canadian *KRAS* wild type cohort was extracted from ICGC database. The American cohort was not included because it lacked fusion in their data sets. The types of structural variants that are important for fusion analysis are inversion and translocation, hence in the tsv file the 'variant type' column was filtered to 'translocation' and 'inversion' categories. After preparing the structural data frame, the file contained the 'break point from' and 'break point to' position without coordinated genes. From the original data frame, two data frames were created. The first data frame included the 'breakpoint_from' data frame, containing the 'chr_from,' 'chr_from_bkpt,' 'icgc_donor_id' and 'sv_id' columns. In addition, the second data frame included data frame 'breakpoint_to,' which included the 'chr_to,' 'chr_to_bkpt,' 'icgc_donor_id' and 'sv_id' columns. In order to find 'coordinated gene,' an additional column was necessary. In other words, one base pair was added to the 'break_point to' and 'breakpoint from' number positions to create the range for each data frame. In total, five columns were in each data frame. The columns 'chr_from,' 'chr_from_bkpt,' 'chro_from_bkpt+1,' 'icgc_donor_id' and 'sv_id' were in data frame 'break point from.' For data frame 'break point to' the columns were 'chr_to,' 'chr_to_bkpt,' 'chr_to_bkpt+1,' 'icgc_donor_id' and 'sv_id'. The two input data frames were run on Linux command line using Bedtools

(Appendix -3) to assign genes to fusion breakpoints. After finding the coordinated genes, the two breakpoints were merged together by the 'row names' column. The resulting merged data frame was the input database for further analysis.

In the next step, the fusion hub dataset was downloaded from the database (https://fusionhub.persistent.co.in/out/global/Fusionhub_global_summary.txt) and imported into R. This database was compared to the input dataset that was generated previously. These were then filtered to find the fusions that were shared between the two data frames '*KRAS* wild type' and 'fusion Hub.'

For the *KRAS* mutant cohort the structural somatic mutation for the combined Australian and Canadian cohort was extracted from the ICGC database and the same approach as described above was used for the *KRAS* mutant cohort to find the coordinated gene. The script for fusion was deposited in GitHub (https://github.com/umccr/Fusion_update).

3.6 Oncofuse

The second tool that was used in the fusion analysis part is Oncofuse. " Oncofuse applies a Bayesian classifier to assign a functional predictive score to previously identified fusion genes responsible for tumorigenesis "[121]. Oncofuse was also applied to detect known and novel gene-fusion driver events in *KRAS* wildtype and *KRAS* mutant pancreatic tumours. One of the advantages of this tool is that it focuses on fusion sequences and reports the probability that any example of fusion is a "driver" in an oncogenic context. It can be downloaded at www.unav.es/genetica/oncofuse.html [122]. Oncofuse uses TICdb and DAVID for gene annotation. TICdb [123] is a catalogue of translocation breakpoints for cancer tumours. DAVID is the Database for Annotation, Visualisation and Integrated Discovery, which is used to interpret biological meaning from the list of genes at any platform [124]. Six functional families that were significantly enriched as driver fusion genes were applied in Oncofuse. These six functional families were transcription factor (TF), kinase activity, transcription cofactor, GTPase (G), helicase/histone modifiers (H) and protein binding (P).

Oncofuse is a post-processing tool that takes fusions called by other programs (such as STAR) as input and predicts oncogenic potential, i.e. the probability of being 'driver' events to fusion sequences. The basis for driver probability prediction is on Bayesian classifier by using Weka machine learning package [125]. Oncofuse uses 24 features for classification of driver mutations. This classification is based on data about known oncogenic fusions hallmarks [126] to discover novel genes to be considered as potential drivers.

Oncofuse produces 36 columns of output. For the purpose of comparison, the columns '5_FPG_GENE_NAME,' '3_FPG_GENE_NAME,' 'SPANNING_READS,' 'ENCOMPASSING_READS,' 'GENOMIC P_VAL_CORR' and 'DRIVER_PROB' were selected. The following descriptions apply to these columns:

1. FPG stands for fusion pair gene.
2. P_val_corr is the Bayesian probability of fusion being a passenger (class 0), given as Bonferroni-corrected p-value.
3. Driver_Prob refers to Bayesian probability of fusion being a driver (class 1).

After preparing the input of Canadian *KRAS* wild type cohort from the ICGC and the Australian input that I received, the variant types that were used for fusion were inversion and translocation. The input file for fusion was created by subsetting the columns 'chr_from,' 'chr_from_bkpt,' 'chr_to,' 'chr_to_bkpt' and 'icgc_donor_id.' The type of tissue, that is EPI, and the type of input, that is coord, were added to the input data frame as extra columns.

The Oncofuse software package was downloaded from https://github.com/mikessh/oncofuse/releases/tag/1.1.1 then it was run on the Linux command line (Appendix–4). The output was a tab-delimited file with classifier p-values (probability of belonging to Class 0 as a passenger, or Class 1 as a driver). In order to make the fusions 'in frame,' the filtration criteria for the output file were column 'frame difference' with a value of 0, and the p-value of driver probability with a cut-off point near 1 (>=0.90).

3.7 RNA Sequencing Data Analysis

Gene expression technologies are used to understand the transcriptional activity of cells. Using transcriptomic profiles make it possible to understand the molecular and genetic mechanisms underlying cancer. Differentially expressed genes (DEG) can also be counted as markers for diagnosis, disease classification, prognosis prediction and targets for therapy.

The edgeR package [127] in R/Bioconductor [128] is a popular tool for analysis of differential expression analyses of replicated count data from RNA sequencing, SAGE or other similar technologies. The software is used to define changes between two or more groups while at least one of the defined groups has replicated counts [129].

The gene expression data for 224 samples from the Canadian and 88 samples from the Australian data sets were obtained from the ICGC data portal. Among these, 205 samples from Canadian and 80 from Australian were *KRAS* mutant. The remainder, which were 19 samples from the Canadian and 8 samples from the Australian data sets were *KRAS* wild type. The analysis for each cohort was done separately. The list of sample IDs of *KRAS* wild type and *KRAS* mutant in Australian and Canadian cohorts are in Appendix 5-8.

In the Canadian cohort, after plotting the samples on the basis of 'Icgc_donor_id', some of the samples had several transcripts for each gene. Therefore, we summarised transcripts at the gene level. The median expression of all transcripts for each gene were taken. The reason we decided to take the median was that the median is more robust to outliers compared to the mean.

After this step, the Ensembl IDs were converted to gene symbols using the biomaRt package [130] from Bioconductor, which annotates the Ensembl IDs to corresponding genes.

For checking differentially expressed genes, the Australian and Canadian datasets were analysed separately and the *KRAS* wild type samples were extracted. For this analysis, the raw read counts were used. The package that was used for DE analysis was edgeR.

After annotation of read counts and applying RLE on the data, data was analysed for finding differentially expressed genes according to the process outlined in the edgeR package. Two design matrices were created for the Australian and Canadian cohorts. A volcano plot, histogram of p-value, MDS plot and Heatmap were created for each cohort. Based on DE, the top genes were selected, and KEGG pathway analysis was conducted on the top selected genes.

3.7.1. Patients and Samples

The *KRAS* mutant cohort data set sourced from the International Cancer Genome Consortium including Australian and Canadian cohorts was analysed. Criteria for selection of data was a mutation in one of *KRAS* codons 12, 13 or 61 and extracted using whole genome sequencing. The data is downloadable through the 'available data type' section, choosing sequencing-based gene expression (EXP-S). The link for the entire Australian cohort is (https://tinyurl.com/yx9jhwhr) and the Australian *KRAS* mutant cohort is (https://tinyurl.com/yy67mb73). The Australian *KRAS* mutant cohort has 80 samples. Additionally, the Canadian *KRAS* mutant cohort has 205 samples. The link for the entire Canadian cohort is (https://tinyurl.com/yyy3dhuo) and the Canadian *KRAS* mutant cohort is (https://tinyurl.com/y5k9czg9).

The *KRAS* wild type cohort includes 8 samples within the Australian cohort and 19 samples within the Canadian cohort. As previously described, the *KRAS* wild type cohort includes the group of samples that have no *KRAS* mutation in codons 12, 13 or 61. These samples were obtained by deducting the *KRAS* mutant cohort from the whole *KRAS* group. Subsetting the *KRAS* wild type cohort was conducted in R version 3.5.1 (2018-07-02).

It is important to note that for relative expression and pathway analysis the 'raw read_counts' were used. For RNA expression the edgeR package [127] was used. In addition, for detection of expressed genes, the cut-off points of log fold change > 1.5 and less than < -1.5, and FDR of < 0.01, were considered. The analysis was conducted for the Australian and Canadian cohorts separately. The script for RNA and pathway analysis was deposited in GitHub (https://github.com/umccr/Update_RNA).

3.7.2 Principal Component Analysis

Principal component analysis (PCA) is a statistical algorithm that decreases the dimension of the data while keeping most variations in the data. It reduces the dimension by defining directions, called principal components, and shows how much variation exists in the data. It is possible to plot samples to visualise their similarities, differences and to show the possibility of grouping the samples together [131]. In other words, PCA is a common approach to find the patterns in the high dimensional data and to get insight into the general structure of a data set [132].

After making a matrix from the log of data, a PCA plot was conducted for each cohort. Up to 3 components of PCA were plotted, showing frequency for each component. The most variation in the cohort was found with up to three dimensions.

3.7.3. RLE Plot

A Relative Log Expression (RLE) plot is one of the most powerful plots that facilitate checking and visualisation of unwanted variants in gene expression data [133]. The RLE plot is one of the outputs from the EDASeq software package [134]. It is used to see the difference between the distributions of reading counts within samples.

For each sample, it demonstrates boxplots from log ratios of gene level of reading counts from the reference sample which is the median within the samples. Generally, a good RLE plot is centred around the zero line, with minimal variation. If there is a deviation in the plot, it shows the existence of outlier samples, and it requires normalisation.

The normalised data was used for RLE analysis. The median of the normalised data was deducted from the normalised data for each sample. The process of RLE analysis was in the RNA script that was deposited in GitHub (https://github.com/umccr/Update_RNA).

3.7.4. Differential Expression Analysis

For differential expression analysis, the raw read counts were used. EdgeR was employed again for the DE analysis. It has statistical tools for obtaining differential expression in RNA sequencing [127].

After annotation of the read counts and applying RLE on the data, the standard edgeR procedures were followed. Two design matrices were created for the Australian and Canadian cohorts based on the two groups *KRAS* wild type and *KRAS* mutant in each cohort.

The next step was to make list-based data in the 'DGElist' format, followed by the 'calcNormFactors' function which was used to normalise RNA. The method that was used was a trimmed mean of 'Mvalues' (TMM) [127].

For exploration of data and checking the outliers, the multi-dimensional scaling (MDS) plot was created from the previous step. This plot can demonstrate the discrepancies among the expression profiles of samples [127].

The following step was to estimate dispersions. Taking a 'DGEList' object and a design matrix, to evaluate the dispersions, the 'estimateGLMCommonDisp' function from edgeR was used [127].

Finally, to perform the likelihood ratio test to clarify differentially expressed genes, the 'glmFit' and 'glmRT' functions were performed [127].

3.8 Pathway Analysis

Pathway analysis methods are used to analyse high throughput data obtained from microarray and RNA sequencing experiments. Khatri and colleagues classified three generations of pathway analysis techniques [135]. In the first-generation approach, a list of genes which were differentially expressed was obtained from the experiment. The genes were then examined to determine which pathways were over- or under-represented. In the second-generation approach as described by Khatri and colleagues [135], called

functional class scoring (FCS) or gene set enrichment analysis (GSEA), values related to all the genes are considered. Subsequently, the gene level and pathway level are calculated. The third-generation of pathway analysis as defined by Khatri et al. [135] called pathway topology. In this approach, the list of genes related to each pathway, pathway component interactions and the location of the interactions are considered for analysis.

Ontology and pathway enrichment analysis was completed using the R package 'KEGG' [136]. This package uses second generation pathway analysis tools. After finding the top list of genes that were differentially expressed, the list of up- and down-regulated genes was identified by using cut off points of log fold greater than 1.5 [137, 138] for the list of up-regulated genes, less than -1.5 for down-regulated genes and FDR less than 0.01[139, 140]. In order to interpret the results of the gene expression analysis, the KEGG pathway analysis was performed. The function for that analysis uses 'kegga' from the edgeR package. This function uses 'NCBI RefSeq' for annotation, hence the 'Entrez Gene id' was provided. Finally, the list of top 50 enriched KEGG pathways was identified by using the 'topKEGG' function [127].

Chapter 4

4 Results

4.1 Patients and Samples

In total, 70 samples in the *KRAS* wild type cohort were analysed. The samples were from three different cohorts: Australian, Canadian and American.

The *KRAS* mutant cohort was compiled from the remainder of the Australian, and Canadian ICGC pancreatic cohorts (n=571). This was used to assess the relative enrichment of somatic events between *KRAS* wildtype and mutant pancreatic cancers.

4.2 Identifying Candidate Genes Driving Tumorigenesis from SNVs

For analysing SNVs, the first step is extracting "Simple Somatic SNV" as tsv format from ICGC. The tsv files were annotated by the ICGC consortium. The *KRAS* mutant cohort from the ICGC contains Australian and Canadian samples. In addition, the *KRAS* wild type cohort was analysed separately as the US, Australian and Canadian groups.

After extracting "somatic mutation tsv" files from ICGC, one of the tools that was used in somatic SNV analysis was Maftools [112]. This tool provides multiple analysis, summary, annotation and visualisation of data. The function of Oncoplot in Maftools demonstrates the mutational landscape of each cohort. In other words, an Oncoplot contains a matrix with rows demonstrating genes and columns representing the samples [112]. One of the functions in Maftools is oncodrive [141]. Oncodrive can identify cancer driver genes in a MAF file. The algorithm for detection of driver genes in oncodrive is called oncodriveCLUST [142].

Comparing the two cohorts of *KRAS* wild type and *KRAS* mutant, the main oncogene in the *KRAS* mutant group is *KRAS*. In the analysis by Maftools, 3% of noncoding mutations were incorrectly mapped by Maftools. The *KRAS* wild type group had different oncogenes with different hits. In the analysis of the American *KRAS* wild type group, the first oncogene is *GNAS*, with 29% frequency, followed by *BRAF* (14%) and *RET* (14%). In the Australian, the first oncogene is *BRAF* (18%) followed by *GNAS* (15%). In the

Canadian *KRAS* wild type group, the first oncogene is *BRAF* with 14% frequency followed by *RET* (9%). These data suggest that the dominant oncogenes in the *KRAS* wild type cohort are *BRAF, GNAS*, and *RET* and the frequency of these candidate oncogenes in the *KRAS* mutant cohort is less than 4%.

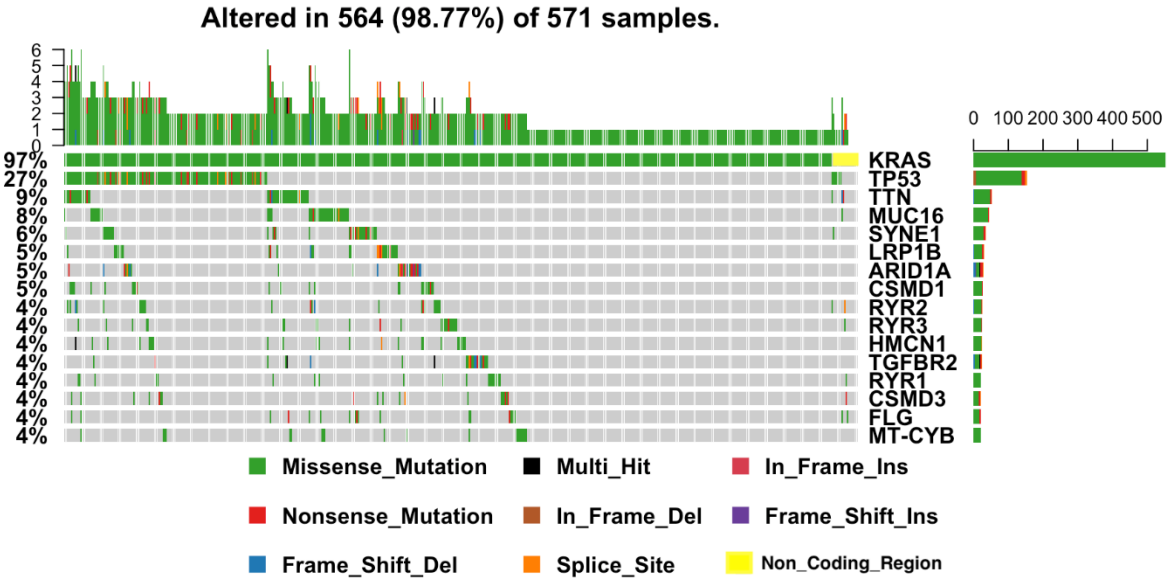


Figure 4. Oncoplot of Australian and Canadian combined *KRAS* mutant cohort. The figure displays the top genes according to the frequency in the *KRAS* mutant cohort and the type of variation in the cohort with missense mutation type. The X axis represents the samples and the Y axis shows different genes with hierarchical frequency. The green colour shows missense mutation which is the main type of mutation in the oncoplot.

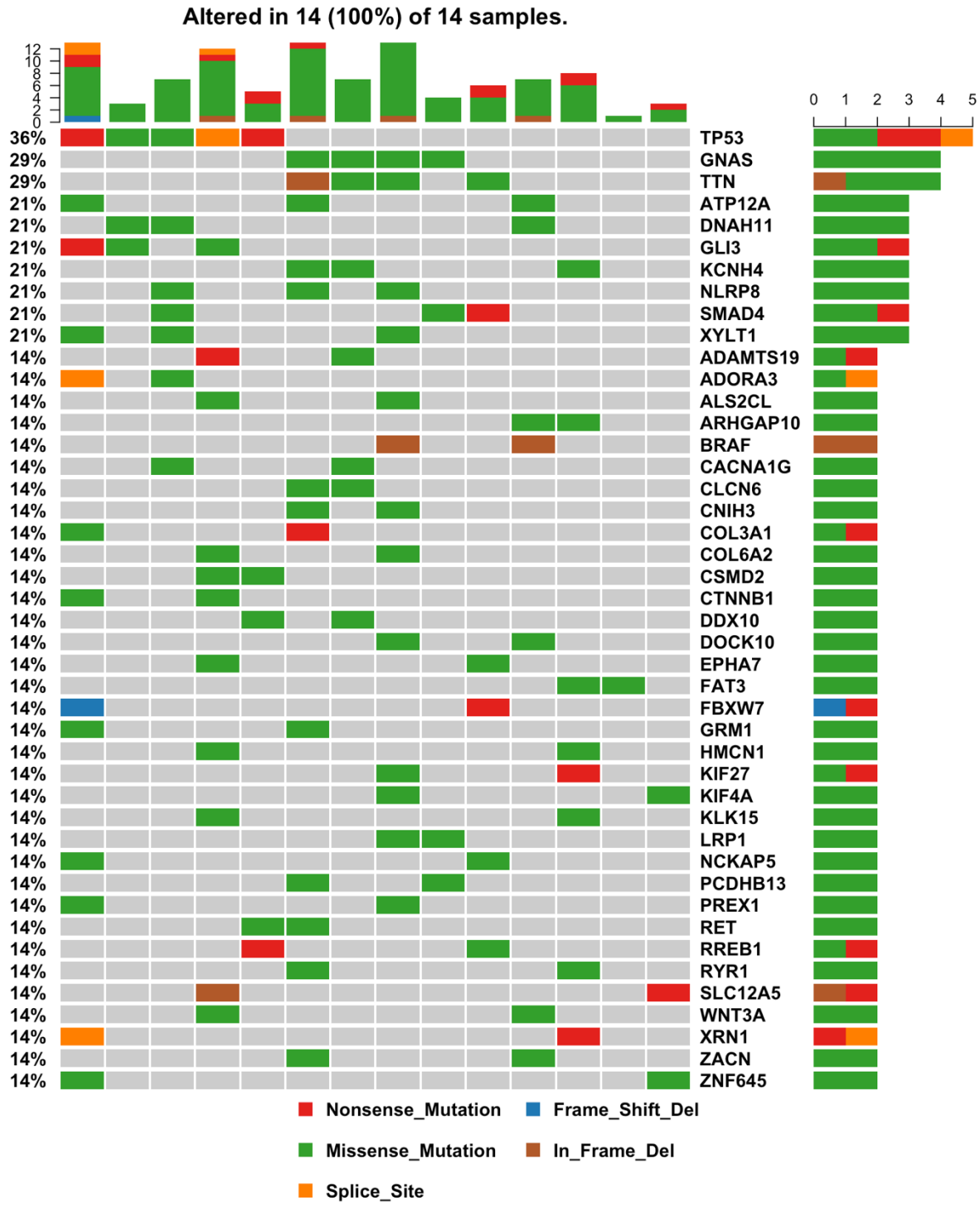


Figure 5. Oncoplot for American *KRAS* wild type samples. The figure displays the top genes according to the frequency in the Australian and Canadian *KRAS* wild type cohort and the type of variation in the cohort. The X axis represents the samples and the Y axis shows different genes with hierarchical frequency. The first oncogene is *GNAS* with 29% frequency with missense mutation, followed by *BRAF* with 14% frequency with mainly in-frame deletion mutation and *RET* with 14% frequency with missense mutation.

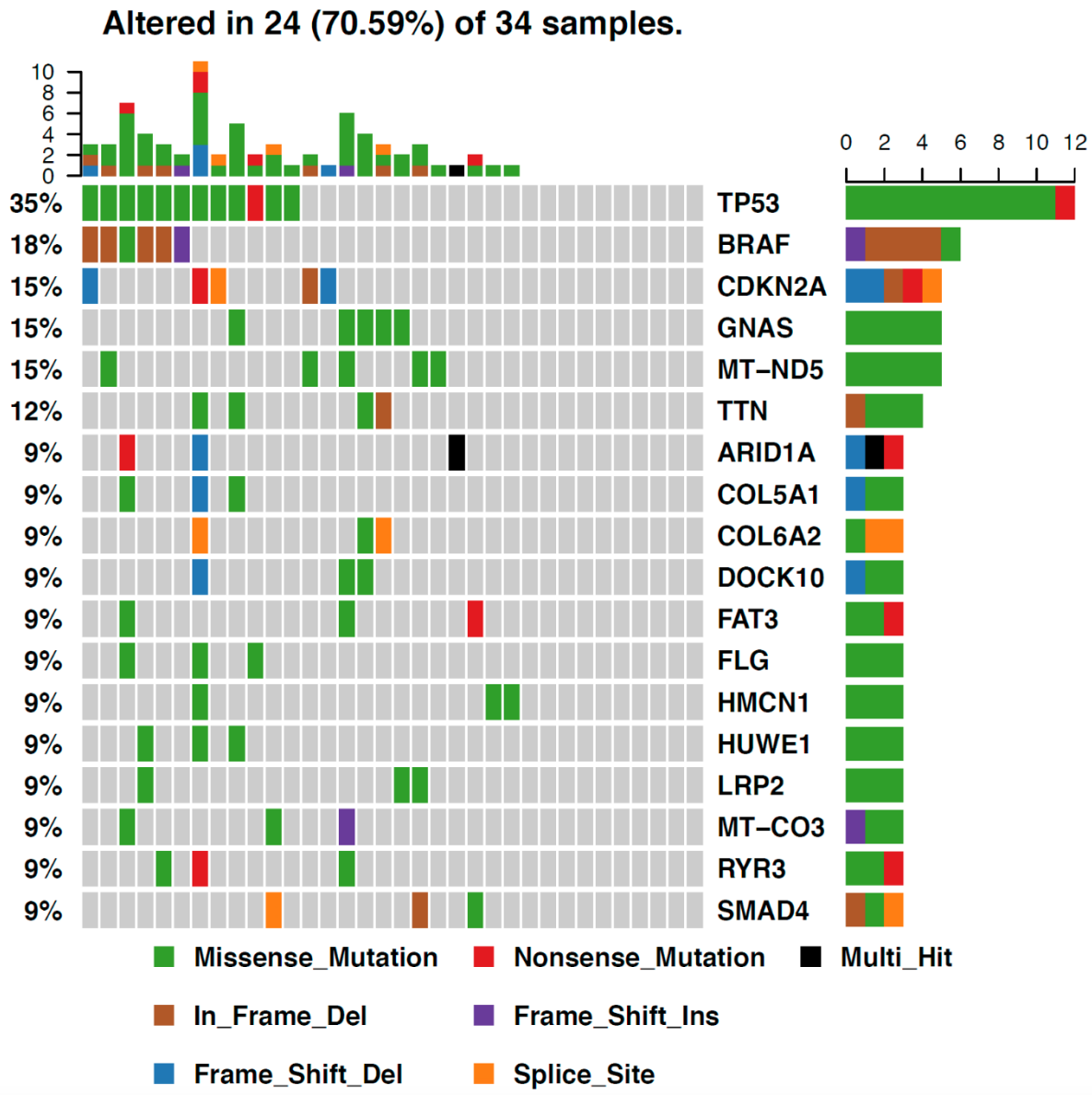


Figure 6. Oncoplot for Australian *KRAS* wild type samples. The figure displays the top genes according to the frequency in Australian *KRAS* wild type cohort and the type of variation in the cohort. The X axis represents the samples and the Y axis shows different genes with hierarchical frequency. The first oncogene is *BRAF* with 18% frequency with mainly in-frame deletion mutation, followed by *GNAS* with 15% frequency with missense mutation.

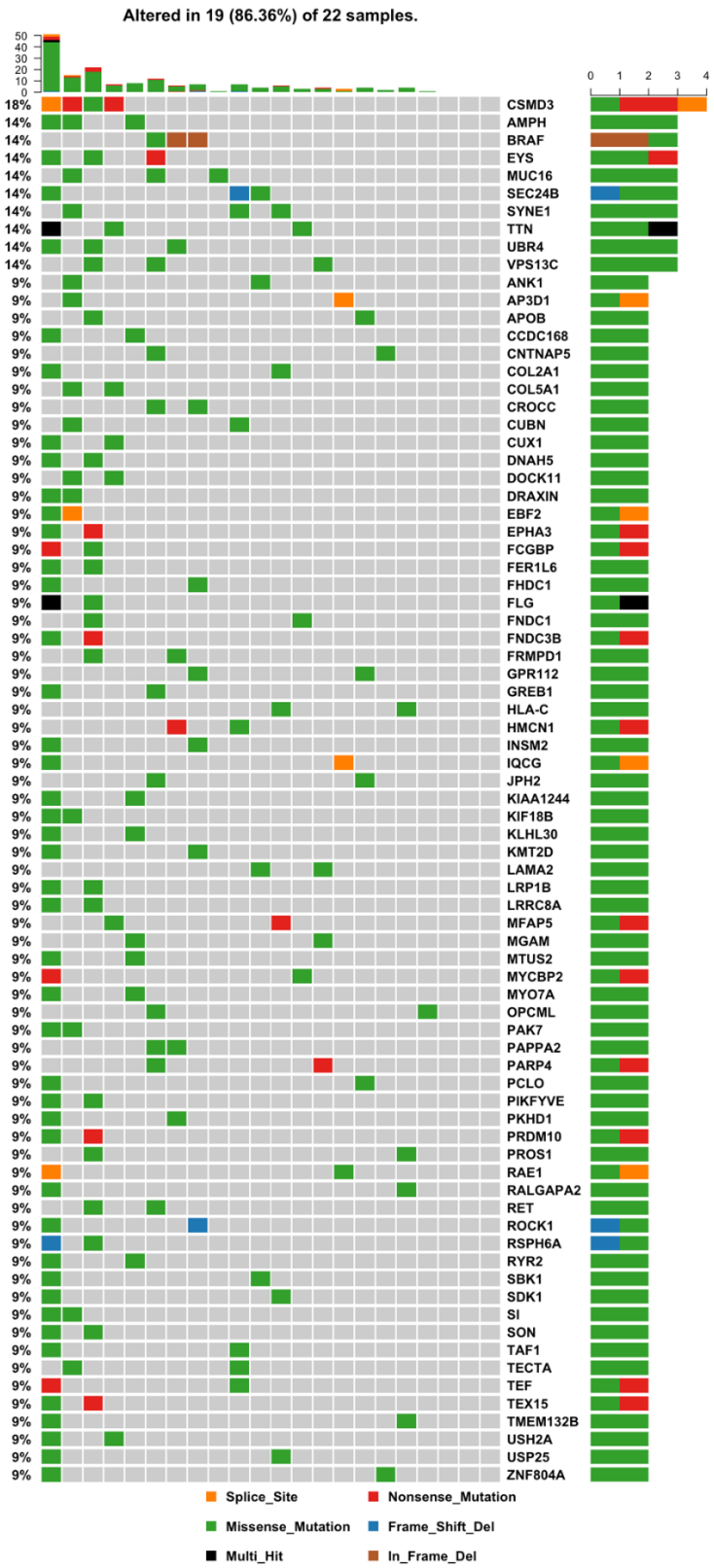


Figure 7. Oncoplot for Canadian *KRAS* wild type samples. The figure displays the top genes according to the frequency in Canadian *KRAS* wild type cohort and the type of variation in the cohort. The X axis represents the samples and the Y axis shows different genes with hierarchical frequency. The first oncogene is *BRAF* with 14% frequency with mainly in-frame deletion mutation, followed by *RET* with 9% frequency with missense mutation.

The second step in SNV analysis was conducted by using CGI. To implement this step, 70 samples in the combined Australian, Canadian and US *KRAS* wild type cohort were analysed for somatic mutation (single nucleotide variants, short insertion and deletions). The input was run on CGI (http://www.cancergenomeinterpreter.org). In addition, the input of the Australian and Canadian *KRAS* mutant of the ICGC panel was run on CGI.

CGI analyses protein-affecting mutations that exists in the Catalogue of Cancer Genes. In other words, it weights based on prior knowledge of the genes. CGI predicts driver mutations on the contribution of every gene and its region in specific cancer types. The prediction is obtained from the analysis of a large cohort of tumours. In addition, when there is no available information for specific cancer types, CGI uses Pan-Cancer information on the corresponding gene [116]. That is to say, CGI works on the individual sample and takes each sample one at a time and then for every event within the patient, it screens the entire set of variants and identifies the genes that are likely to be drivers or passengers based on type of mutation and prior knowledge and application of the Catalogue of Validated Oncogenic Mutations. CGI also assesses the tumorigenicity of variants of unknown significance via OncodriveMUT [116].

The output results of CGI in the two cohorts were verified for separation of driver mutations. In total, 227 variants in the *KRAS* wild type group with average of 5 driver mutations per tumour which ranges from 1 to 60, median of 2. These expectations are the same as previous reports of papers which use 4-5 driver mutations per tumour [143, 144]. In the *KRAS* mutant cohort, 2067 variants were identified as drivers with an average of 4 mutations per tumour which ranges from 1 to 44 with a median of 3. Also, in the *KRAS* mutant cohort, these figures are the same as previous reports of papers which use 4-5 driver mutations per tumours. [143, 144]

To visualise the data, the driver genes with frequency greater than five somatic events in coding regions were plotted using the 'ggplot' function. Figure 8 shows the output analysis of the two groups of *KRAS* wild type and KRAS mutant. In the *KRAS* mutant group *KRAS* is the main oncogene, while in the *KRAS* wild type group *GNAS* is the main and first oncogene with 17% frequency.

TP53, CDKN2A, SMAD4 and *ARID1A* are shared between the *KRAS* wild type and *KRAS* mutant groups. *RNF43* and *MLL3* are specific to the *KRAS* mutant group. *MLL*, *GNAS, ACVR1B* and *ARID2* are unique to the *KRAS* wild type group. The χ-squared test for genes with frequency greater than five somatic events in coding regions was done and the p-value was significant. ($\chi^2 = 143.43$, df = 10, p-value < 2.2e-16).

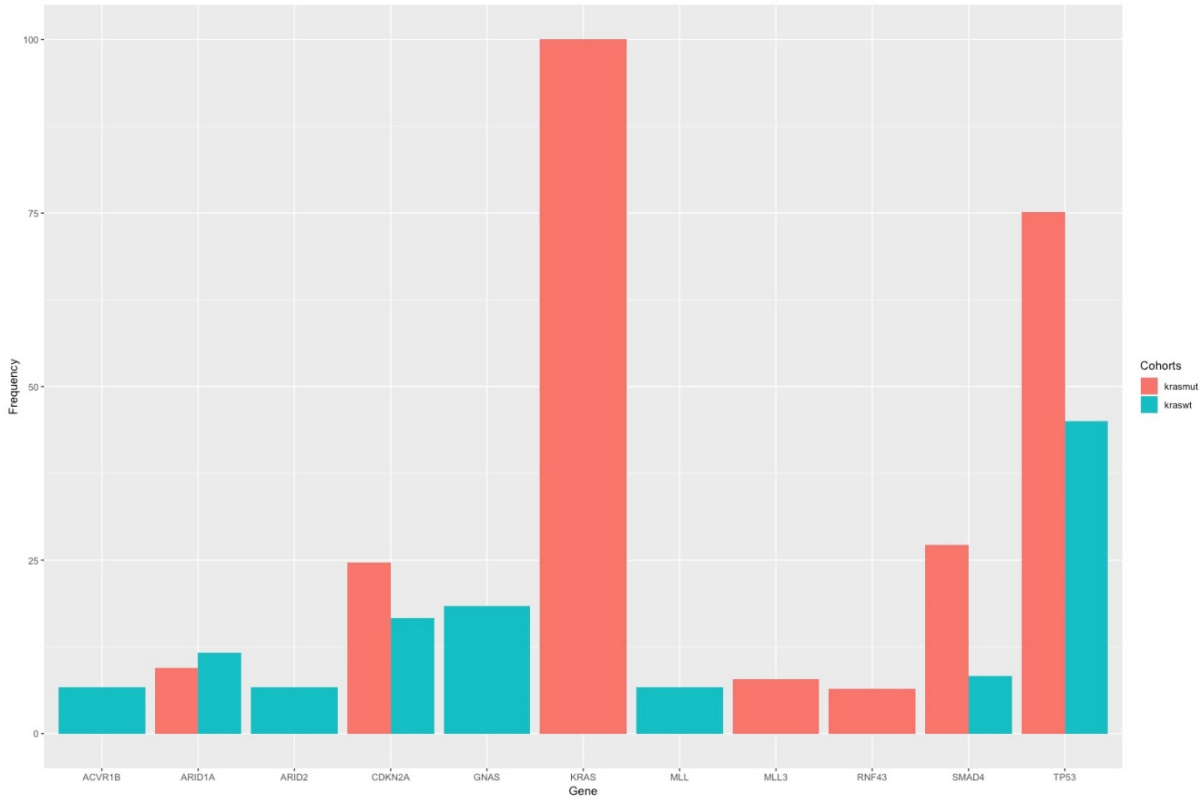


Figure 8. The driver genes in *KRAS* wild type and *KRAS* mutant plotted. The genes with frequency > 5 somatic events in coding regions are shown. The first oncogene in the *KRAS* mutant cohort is *KRAS* and the first oncogene in the *KRAS* wild type cohort is *GNAS* with 18% frequency. ($\chi^2 = 143.43$, df = 10, p-value < 2.2e-16).

4.3 Identifying Genes Driving Tumorigenesis from CNVs

CGI can also predict driver mutations from copy number alterations [116]. A copy number tsv file was extracted from the ICGC database for *KRAS* mutant and *KRAS* wild type samples and coordinated genes were found using bedtools. After preparing the input format for CGI, the input was run on CGI. The driver mutations from the result of the

CGI analysis were plotted using Oncoprint. In the analysis of the copy number variation, homozygous deletion and copy number amplification were conducted separately for each set.

Copy number changes were filtered for homozygous deep deletion (copy number = 0) and amplifications (copy number >= 5). These thresholds match the ICGC standard thresholds for copy number analysis [145] and used for both cohorts of *KRAS* mutant and *KRAS* wild type.

In the *KRAS* wild type cohort, for amplification with CNV greater than or equal to 5, in total 42, 692 genes were called as drivers and passengers and 62 of them were identified as predicted drivers. In addition, in the KRAS wild type cohort, for homozygous deletion, 4592 genes were called as drivers and passengers, and 30 of them were identified as predicted loss of function.

The list of driver genes with CNV greater than or equal to 5 in the *KRAS* wild type cohort were plotted using Oncoprint (Figure 9). Copy number amplifications in a component of the MAP kinase pathway in CNV greater than or equal to 5, contain the list of genes, with high frequency in *EGFR* (12%), *ERBB2* (6%), *AKT2* (16%), *RICTOR* (12%), *FGFR1* (10%) , *FGF3* (9%) and *FGF4* (9%). In addition, amplifications in other oncogenes were observed with high frequency in the *KRAS* wild type group, including *MYC* (23%), *CDK6* (16%) and *MDM4* (11%). On the other hand, the frequency of these genes in the *KRA*S mutant cohort was less than that of the *KRAS* wild type. It seems *KRAS* wild type has more copy number > =5 than KRAS mutant with copy number >= 5.

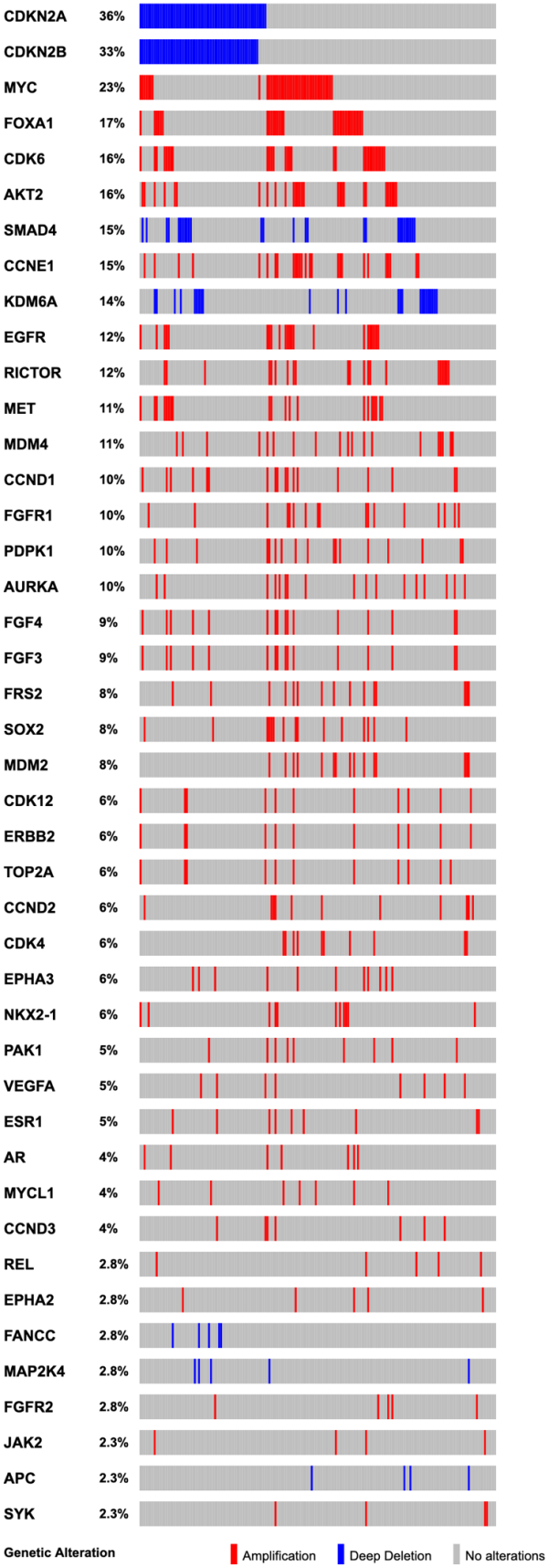


Figure 9. Oncoprint plot for drivers in CNV amplification and deletions among *KRAS* wild type cohort with copy number >= 5. The X axis represents the samples and the Y axis demonstrates the hierarchical frequency of each gene. Deep deletions (CNV <1, blue) or amplifications (CNV >= 5, Red).

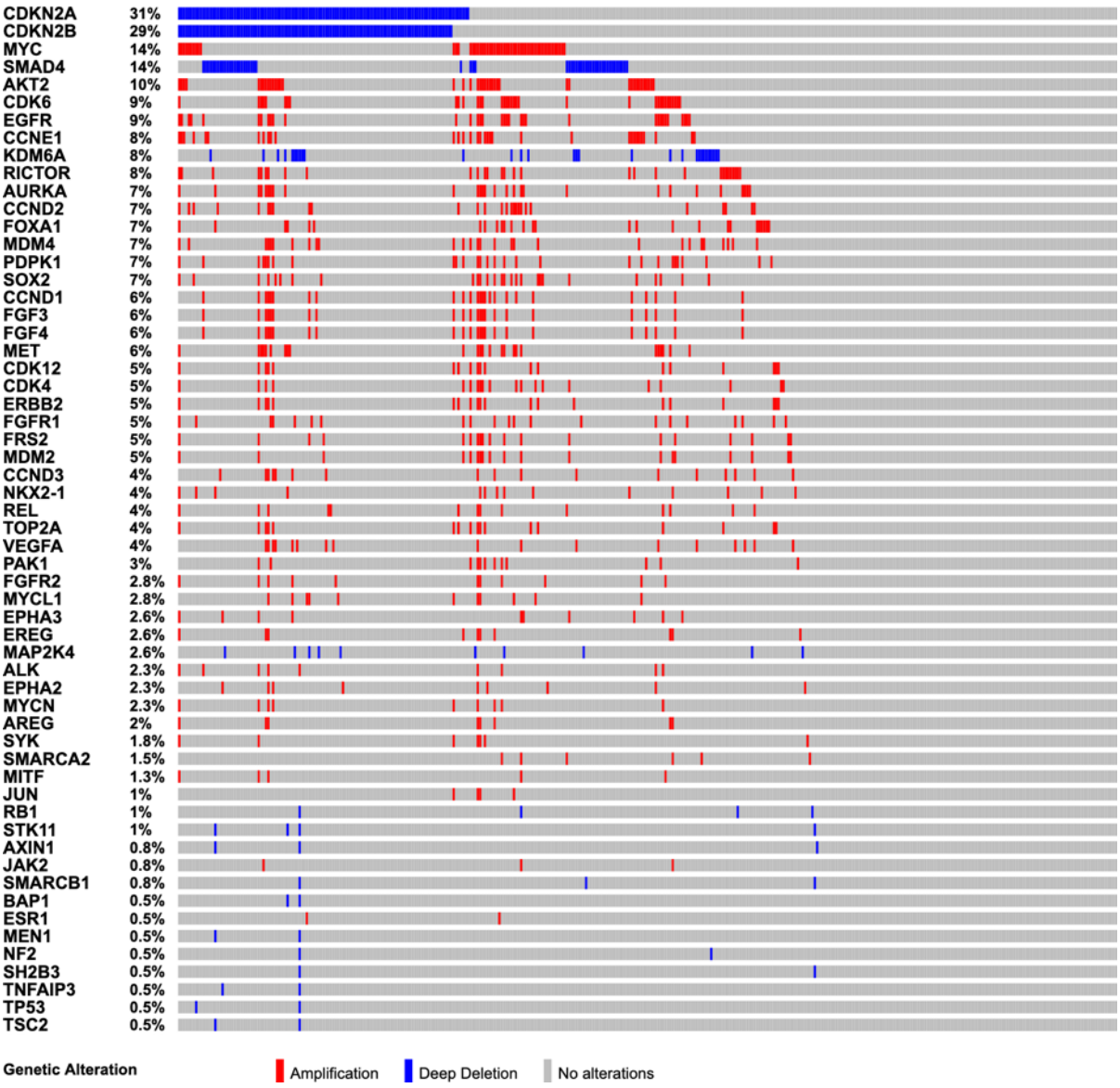


Figure 10. Oncoprint plot for drivers in CNV amplification >= 5 and deletions among *KRAS* mutant cohort. The X axis represents the samples and the Y axis demonstrates the hierarchical frequency of each gene. Deep deletions (CNV <1, blue) or amplifications (CNV>= 5, Red).

4.4 Fusions

In order to find the fusions, one of the tools used was Fusion Hub. Fusion Hub is a web platform tool that has a fusion database from a collection of 10 public domain databases and 14 datasets from literature. In total 24 datasets are available in the analysis by Fusion Hub. The results of Fusion Hub predictions come from the reports of gene fusions from the prior literature [120].

The second tool that was used to predict fusions was Oncofuse. Oncofuse is a Bayesian classifier that assigns functional prediction scores to fusion sequences that were called by other programs. Oncofuse uses 24 features for classification of driver mutations. This classification is based on previous data about known oncogenic fusions hallmarks [126] to discover novel fusion genes to be considered as potential drivers [122]. Additionally, the results of the fusion analysis are prospective candidate fusion drivers that need further validation.

4.4.1 Fusions in *KRAS* Wild Type

For finding the fusion datasets in the *KRAS* wild type cohort, the chromosomal positions corresponding to inversion and translocation were mapped to coordinated genes using bedtools. The output of the subsets was run through bedtools (https://bedtools.readthedocs.io/en/latest/). The coordinated genes for each sample were paired together. After preparing the paired inputs, the paired genes were compared to the Fusion Hub database and Oncofuse. The output of each database was tabulated.

4.4.1.1 Fusion Hub

After merging all the fusion events from inversion and translocation breakpoints that intersect two genes and creating one file, the result was compared with the Fusion Hub database to identify known fusions reported in previous published papers and similar results were extracted in the format of a tab delimited file. Table 3 shows all the fusion events which were extracted from this comparison that have one component in the MAP kinase pathway. In total 36 samples have fusions and 5 samples have *BRAF* fusions. Figures 11a, 11b and 11c represent three fusion events from MAP kinase gene sets. Table

3 lists all of the fusions in the *KRAS* wild type cohort which were in the Fusion hub database and have one component from the MAP Kinase signalling pathway.

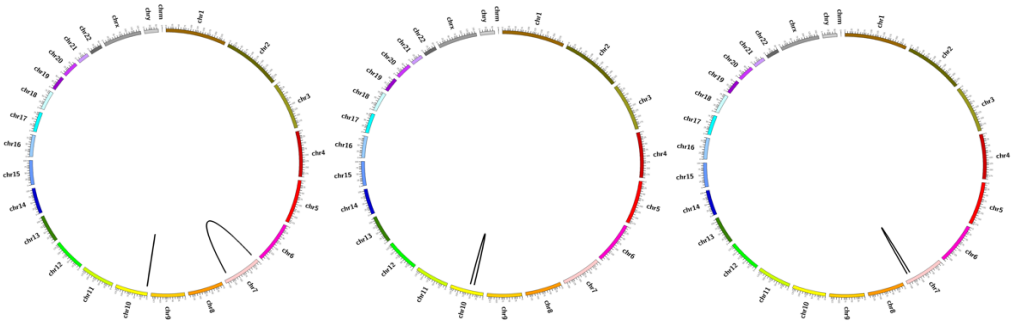


Figure 11.a Figure 11.b. Figure 11.c

Figure 11: Three selected fusions from *KRAS* wild type cohort. 11.a: Circus plot of all fusion events in one of the samples (APGI-1971). The fusions include SDK1_BRAF and NET1_GD12. 11.b: The RET_CCDC6 fusion in one of the samples (APGI_3060).

Table 3. List of candidate fusions which were shared between the fusion set of the *KRAS* wildtype cohort and Fusion Hub that have one component in MAP kinase signalling pathway. In total 36 samples have fusions and 5 samples have candidate *BRAF* fusions.

	Chr_from	Chr_from_bkpt	Chr_to	Chr_to_bkpt	Gene_ from	Gene_ to	Donor_id
1	12	2736670	12	2737246	CACNA1C	CACNA1C	APGI_2219
2	12	2240982	12	2257996	CACNA1C	CACNA1C	APGI_2715
3	7	140499490	7	140542751	BRAF	BRAF	APGI_2973
4	20	43622137	20	45202143	STK4	SLC13A3	APGI_2991
5	1	26223517	17	46048428	STMN1	CDK5RAP3	APGI_3333
6	2	102379295	2	102463199	MAP4K4	MAP4K4	APGI_3796
7	7	4282203	7	140497729	SDK1	BRAF	APGI_1971
8	4	86419604	4	87159277	ARHGAP24	MAPK10	APGI_2938
9	7	127349939	7	140489991	SND1	BRAF	APGI_3060
10	4	72426155	15	79312585	SLC4A4	RASGRF1	APGI_3205
11	7	138253575	7	140512331	TRIM24	BRAF	DO51491
12	7	138239528	7	140498312	TRIM24	BRAF	DO51503
13	10	43610446	10	61595080	RET	CCDC6	APGI_2153
14	6	117646939	4	72432608	ROS1	SLC4A4	DO49442
15	8	32574483	8	32574640	NRG1	NRG1	DO49442

4.4.1.2 Using Oncofuse for Identification of Novel Oncogenes

One of the tools which demonstrates known and novel oncogenes is Oncofuse. Oncofuse uses a Bayesian classifier to assign a functional prediction score to fusion sequences. This score is based on p-value for driver probability of each fusion set. In addition, to assure the fusions are in-frame, 'FPG_frame difference = 0' was selected in filtering the output results. The cut off point for the driver probability is driver probability greater than or equal to 0.9 [122]. Table 4 depicts the candidate oncogenic fusions in the Australian and Canadian sets.

In order to compare the results from analysis using Fusion Hub and Oncofuse in *KRAS* wild type VENN diagram package was used [146]. The comparison showed 6 fusions which were shared from the predictions of Fusion Hub and oncofuse. The VENN diagram in Figure 12 shows the number of shared fusions between Fusion Hub and VENN diagram.

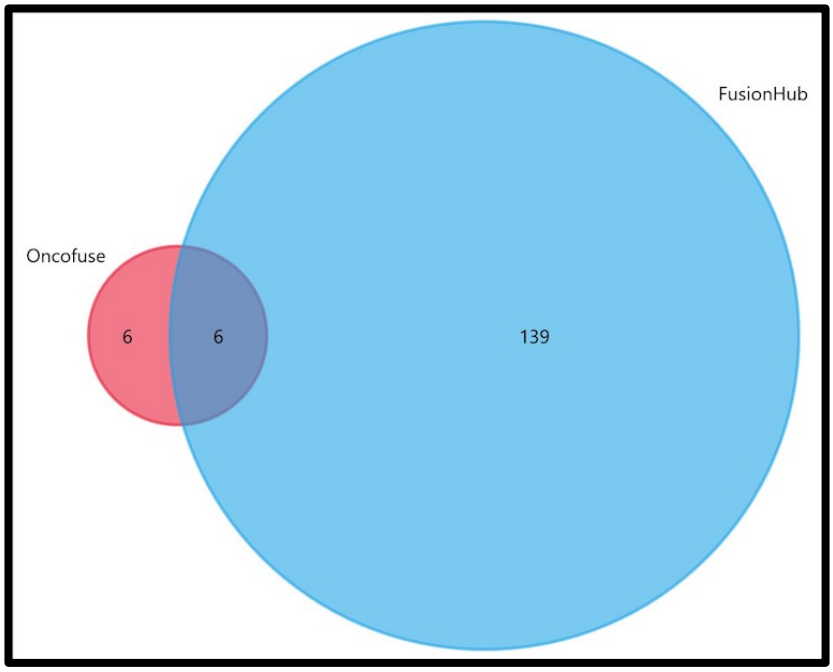


Figure 12. The VENN for comparing results of fusions from Fusion Hub and Oncofuse in *KRAS* wild type cohort. The diagram shows 6 fusions were shared from the analysis of Fusion Hub and Oncofuse.

Table 4. The top candidate fusions in *KRAS* wild type cohort. The result had two sets of filtrations with driver probability >= 0.9 and FPG_frame difference = 0.

	Chr_from	Chr_from_bkpt	Chr_to	Chr_to_bkpt	Gene_From	Gene_To	FRAME_Difference	P_VAL_CORR	DRIVER_PROB	Donor_id
1	chr1	6118471	chr1	8609598	KCNAB2	RERE	0	0.85738504	0.92985032	DO51503
2	chr10	24882761	chr10	25547421	ARHGAP21	GPR158	0	0.7864192	0.95710441	DO224633
3	chr2	169736765	chr18	22763194	SPC25	ZNF521	0	1	0.93325655	DO49430
4	chr5	107679624	chr1	8623425	FBXL17	RERE	0	0.28002954	0.98727139	DO51503
5	chr7	138239528	chr7	140498312	TRIM24	BRAF	0	0.0037432	0.99996597	DO51503
6	chr7	138253575	chr7	140512331	TRIM24	BRAF	0	0.0037432	0.99996597	DO51491
7	chr10	917506	chr21	35158441	LARP4B	ITSN1	0	0.03076422	0.996733	APGI_2997
8	chr11	107928438	chr11	108094514	CUL5	ATM	0	0.58195671	0.90214887	APGI_3796
9	chr11	70217682	chr11	70256964	PPFIA1	CTTN	0	0.01989817	0.99859128	APGI_2991
10	chr18	30665082	chr18	29643337	CCDC178	RNF125	0	0.00390858	0.99982705	APGI_2997
11	chr3	47844481	chr3	47807028	DHX30	SMARCC1	0	0.13597193	0.98074734	APGI_2938
12	chr6	44398009	chr6	138555785	CDC5L	KIAA1244	0	0.0958889	0.98727139	APGI_2997
13	chr7	127349939	chr7	140489991	SND1	BRAF	0	0.000787	0.9999791	APGI_3060
14	chr7	4282203	chr7	140497729	SDK1	BRAF	0	0.000715	0.99997469	APGI_1971
15	chr9	36886523	chr9	33159738	PAX5	B4GALT1	0	0.000202	0.99999821	APGI_3205

Considering Oncofuse, the *KRAS* wild type cohort has 4 candidate fusions that have one component from the MAP kinase signalling pathway**.** Table 5 demonstrates all MAP kinase fusions.

Table 5. The output result of candidate fusions from Oncofuse in *KRAS* wild type cohort that has one gene partner from MAP kinase signalling pathway

Chr_from	Chr_from_bkpt	Chr_to	Chr_to_bkpt	X5_FPG_Gene_ Name	X3_FPG_Gne_ Name	FRAME_ Difference	P_VAL_CORR	DRIVER_PROB	Donor_id
chr7	127349939	Chr7	140489991	SND1	BRAF	0	0.000787	0.9999791	AGPI_3060
chr7	138239528	chr7	140498312	TRIM24	BRAF	0	0.0016334	0.99996597	DO51503
chr7	138253575	chr7	140512331	TRIM24	BRAF	0	0.0016334	0.99996597	DO51491
chr7	4282203	chr7	140497729	SDK1	BRAF	0	0.0016197	0.99997469	APGI-1971

4.4.2 Fusions in *KRAS* Mutant Cohort

4.4.2.1 Fusion Hub

From comparison between the *KRAS* mutant fusion and the Fusion hub database, 2045 fusions were found in 320 samples with an average of 1 fusion per tumour, ranges from 1 to 8 and median of 1. These expectations have similarities with previous reports. Gao et al. suggested in different cancer types the overall number of fusions per sample reported from 0 to 60, with a median of 1 fusion per sample [147]. While the average number of fusions per sample in the analysis of the ICGC data by Fonseca et al. varied from 1 to 14. Fonseca et al. reported that the average number of fusions is different based on histological types and average number of structural variants. The average number of fusions per sample varies from 14 in boneleimyocarcinomas to 1 in 10 cancer types [148]. In the *KRAS* mutant cohort, from 2045 candidate fusions, 56 candidates had one component from the MAP kinase signalling pathway. Table 6 demonstrates all candidate MAP kinase fusions.

Table 6. Candidate Fusions in *KRAS* mutant cohort that have one component from MAP kinase signalling pathway from comparison with Fusion Hub database.

Donor_id	Chr_from	Chr_from_bkpt	Gene_from	Gene_to	Chr_to	Chr_bkpt
DO32904	chr2	128058792	MAP3K2	MAP3K2	chr2	128058996
DO32904	chr3	54393578	CACNA2D3	CACNA2D3	chr3	54393885
DO32904	chr8	99769724	STK3	STK3	chr8	99772583
DO32851	chr3	168972662	MECOM	MECOM	chr3	168972832
DO32851	chr3	30685412	TGFBR2	TGFBR2	chr3	30687398
DO32851	chr3	169142265	MECOM	MECOM	chr3	169108678
DO32851	chr12	118776327	TAOK3	TAOK3	chr12	118757298
DO32851	chr8	38291838	FGFR1	NRXN3	chr14	79807915
DO34905	chr8	99475029	STK3	STK3	chr8	99475146
DO49175	chr12	2612566	CACNA1C	CACNA1C	chr12	2659013
DO34128	chr4	102255133	PPP3CA	PPP3CA	chr4	102255213
DO32928	chr14	50667106	SOS2	NEMF	chr14	50209824
DO33208	chr3	192015689	FGF12	MB21D2	chr3	192620943
DO33400	chr8	99871361	STK3	STK3	chr8	99820743
DO49138	chr19	40738366	AKT2	AKT2	chr19	40740190
DO33984	chr1	68202525	GNG12	GNG12	chr1	68205264
DO32893	chr3	54278887	CACNA2D3	CACNA2D3	chr3	54246607
DO49080	chr22	37632987	RAC2	GGA1	chr22	38005373
DO49080	chr22	21280473	CRKL	SF3A1	chr22	30743648
DO49127	chr7	82010023	CACNA2D1	CACNA2D1	chr7	82010093
DO33042	chr7	55227656	EGFR	EGFR	chr7	55227706
DO33688	chr5	179681482	MAPK9	RASGEF1C	chr5	179611720
DO34368	chr4	102090471	PPP3CA	PPP3CA	chr4	102090603
DO34368	chr1	243782745	AKT3	AKT3	chr1	243783756
DO33056	chr3	54589293	CACNA2D3	CACNA2D3	chr3	54584725
DO34736	chr8	99934020	STK3	STK3	chr8	99936131
DO34817	chr1	206872695	MAPKAPK2	CR1	chr1	207695080
DO32900	chr17	29595042	NF1	NF1	chr17	29572996
DO32819	chr8	99491627	STK3	STK3	chr8	99491743
DO34312	chr17	73329804	GRB2	GRB2	chr17	73339671
DO34680	chr7	81616531	CACNA2D1	CACNA2D1	chr7	81616626
DO49472	chr8	128844461	MYC	PVT1	chr8	128844221
DO32904	chr3	168994277	MECOM	MECOM	chr3	169027777
DO49184	chr19	40786388	AKT2	SNRPA	chr19	41260241
DO49137	chr3	168986256	MECOM	MECOM	chr3	168988949
DO32904	chr3	54393578	CACNA2D3	CACNA2D3	chr3	54393885
DO33432	chr6	137032520	MAP3K5	MAP3K5	chr6	137001570
DO32851	chr3	169142265	MECOM	MECOM	chr3	169108678
DO32851	chr3	30685412	TGFBR2	TGFBR2	chr3	30687398
DO32851	chr3	168972662	MECOM	MECOM	chr3	168972832
DO32851	chr12	118776327	TAOK3	TAOK3	chr12	118757298
DO32851	chr8	38291838	FGFR1	NRXN3	chr14	79807915
DO49158	chr19	40782253	AKT2	AKT2	chr19	40782317
DO49158	chr7	82042920	CACNA2D1	CACNA2D1	chr7	82039592
DO49158	chr12	2341241	CACNA1C	CACNA1C	chr12	2338571
DO33552	chr3	58118415	FLNB	FLNB	chr3	58017021
DO51528	chr17	29661775	NF1	NF1	chr17	29663088
DO32851	chr11	75204325	GDPD5	ARRB1	chr11	74976473
DO33042	chr19	38884754	SPRED3	RASGRP4	chr19	38907100
DO34680	chr19	15793857	CYP4F12	MAP2K4	chr17	11970527
DO49141	chr6	136808360	MAP7	MAP3K5	chr6	136962796
DO49079	chr8	73742392	KCNB2	STK3	chr8	99644553
DO49158	chr16	30011850	INO80E	TAOK2	chr16	29990547
DO227742	chr4	91566337	CCSER1	MAPK10	chr4	87221198
DO227570	chr4	54393276	LNX1	PDGFRA	chr4	54395308

4.4.2.2 Oncofuse

The results from Oncofuse showed 2,242 fusions. Of these, 112 candidates had driver probability greater than or equal to 0.9 and FPG_frame difference= 0. In these 112 candidate fusions, 9 candidate fusions had one component from the MAP Kinase signalling pathway. Table 7 demonstrates all candidate MAP kinase fusions.

Table 7. Fusions in *KRAS* mutant cohort that have one component from MAP Kinase signalling pathway.

Chr_from	Chr_from_bkpt	Chr_to	Chr_to_bkpt	X5_FPG_Gene_Name	X3_FPG_Gene_Name	FPG_Frame_Difference	P_VAL_CORR	DRIVER_PROB	Icgc_donor_id
chr17	26506531	chr17	42293258	NLK	UBTF	0	0.51961389	0.9431981	DO224776
chr22	21280473	chr22	30743648	CRKL	SF3A1	0	0.37253629	0.96618135	DO49080
chr3	50656852	chr11	70260373	MAPKAPK3	CTTN	0	0.00606215	0.99996254	DO49150
chr6	166882675	chr19	4196358	RPS6KA2	ANKRD24	0	0.60700563	0.92787481	DO227558
chr8	99686817	chr17	73391492	STK3	GRB2	0	0.01446457	0.99988313	DO49463
chr1	206729613	chr1	206870719	RASSF5	MAPKAPK2	0	0.19539332	0.98829869	DO231251
chr11	33359929	chr11	36529457	HIPK3	TRAF6	0	0.46332912	0.95199347	DO33049
chr18	48585148	chr22	22139449	SMAD4	MAPK1	0	0.08886513	0.99632544	DO33528
chr19	15793857	chr17	11970527	CYP4F12	MAP2K4	0	0.00455176	0.99999135	DO34680

In order to compare the results from analysis using Fusion Hub and Oncofuse in the *KRAS* mutant cohort, VENN diagram package was used [146]. The comparison showed 22 fusions which were shared from the predictions of Fusion Hub and oncofuse. The VENN diagram in Figure 13 shows the number of shared fusions between Fusion Hub and VENN diagram.

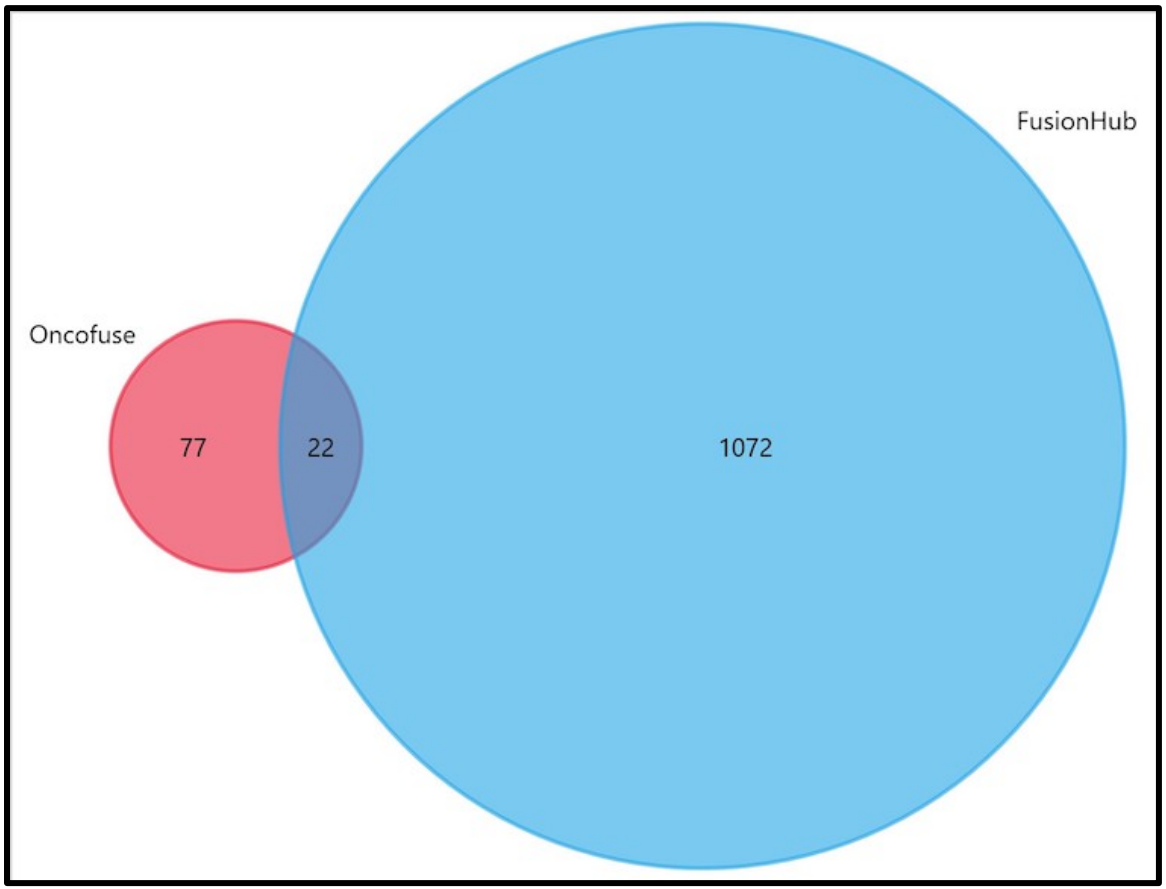


Figure 13. The VENN for comparing results of fusions from Fusion Hub and Oncofuse in *KRAS* mutant cohort. The diagram shows 6 fusions were shared from the analysis of Fusion Hub and Oncofuse.

4.5 RNA Sequencing

Following the second aim of the study, in order to characterise which underlying pathways drive the *KRAS* wild type and *KRAS* mutant cohorts, the RNA sequencing data for the two cohorts of KRAS wild type and KRAS mutant of the Australian and Canadian groups were sourced from ICGC. The Australian *KRAS* mutant cohort has 80 samples. Additionally, the Canadian *KRAS* mutant cohort has 205 samples. The *KRAS* wild type cohort includes 8 samples within the Australian cohort and 19 samples within the Canadian cohort.

The Australian RNA-seq data were adjusted for library size using the count per million (CPM) method. The RLE plot of the adjusted data set (figure 14 and figure 15) shows several outlier samples. The outlier samples were removed for further analysis.

After annotation of the read counts, the RLE plot for each cohort was created. Figures 14 and 15 show the RLE plots of the Australian (n = 88) and Canadian (n=224) cohorts. Based on this analysis the data showed deviation from the zero lines. Outliers of the RLE plots were removed to reduce unwanted variation.

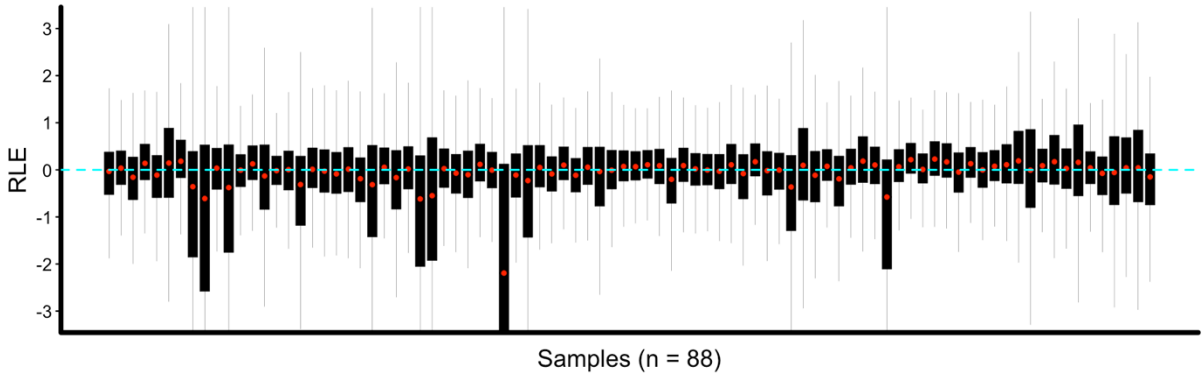


Figure 14. RLE plot for the whole set of the Australian dataset.

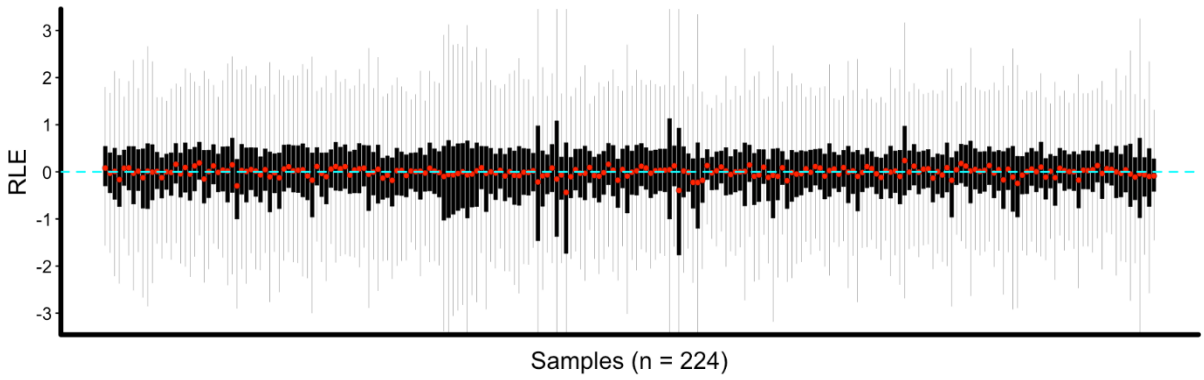


Figure 15. RLE plot for the whole set of the Canadian dataset.

In addition, to show variation in the datasets, the CPM values of the data were used for PCA plotting. Figures 16 and 17 show the PCA plots with application of mutation status.

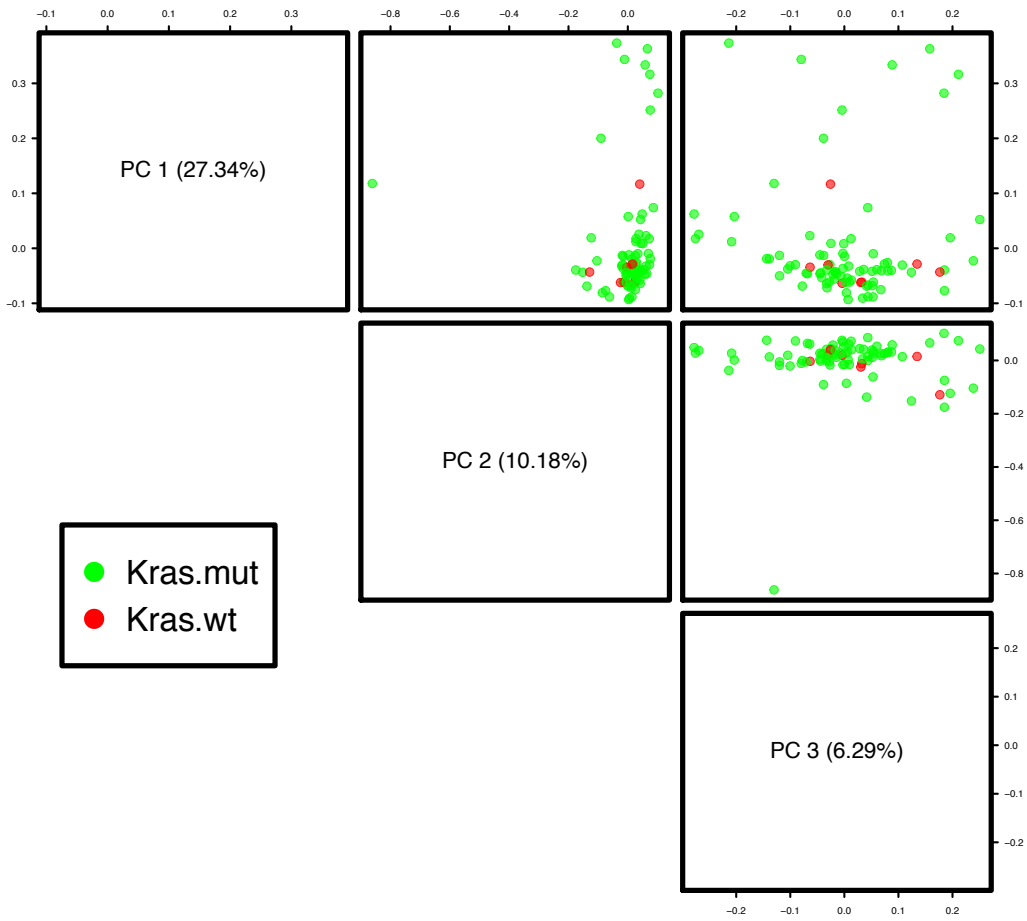


Figure 16. PCA plot of the Australian cohort. It shows the mutational status of the *KRAS* wild type and *KRAS* mutant groups. The first three principal components show 27.34 % of variation in the Australian cohort.

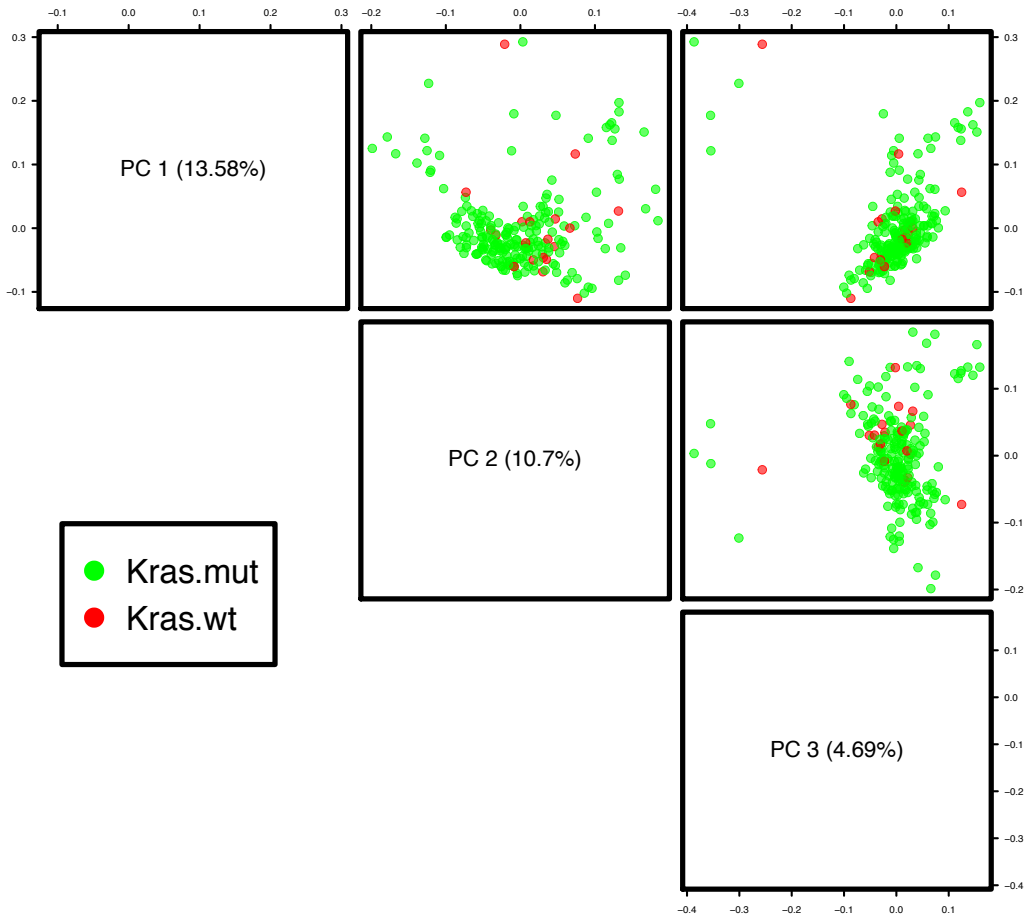


Figure 17. PCA plot of the Canadian cohort. It shows the mutational status of the *KRAS* wild type and *KRAS* mutant groups. The first three principal components show 13.58 % of variation in the dataset.

Differential expression analysis of the Australian and Canadian *KRAS* mutant versus *KRAS* wild-type cohorts was completed using the edgeR package. For detection of expressed genes, the cut-off points of log fold change > 1.5 [137, 138] and less than < -1.5, and FDR of < 0.01 [139, 140], were considered. Volcano plots of the Australian and Canadian *KRAS* mutant versus *KRAS* wild-type cohorts were depicted in figures 18 and 19 to visualise the results of differential expression analysis. In addition, the histograms of p-values of the Australian and Canadian cohorts were depicted in figure 20 and 21 to show the number of differentially expressed genes in each cohort and pattern of p-value in each cohort which showed the p-value distribution of differentially expressed genes.

Using the edgeR package, in the Australian cohort 173 genes were found to be up-regulated and 12 genes were found to be down-regulated. In the Canadian cohort, 140

genes were found to be up-regulated, and 20 genes were found to be down-regulated. Supplementary Tables 1, 2, 3 and 4 show the list of up and down-regulated genes in the Australian and Canadian *KRAS* wild type cohorts with their respective Log fold change, LogCPM, P-value and FDR. However, plotting the list of differentially expressed genes did not show any clustering between *KRAS* wild type and *KRAS* mutant. Figure 22, 23, 24 and 25 display heatmaps of up and down-regulated genes in the Australian and Canadian cohorts.

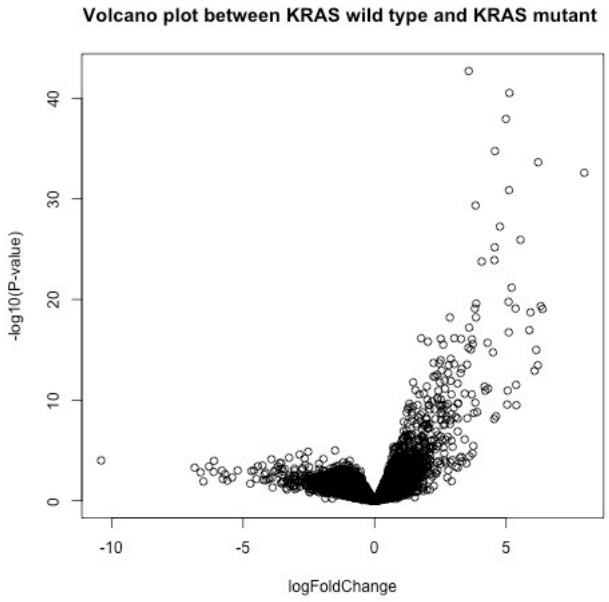


Figure 18. Volcano plot between *KRAS* wild type and *KRAS* mutant of the Australian cohort. The plot is useful for visualisation of the results of differential expression analysis. For detection of expressed genes, the cut-off points of log fold change > 1.5 and less than < -1.5, and FDR of <0.01 were considered. Genes that are highly dysregulated are to the far left and right sides.

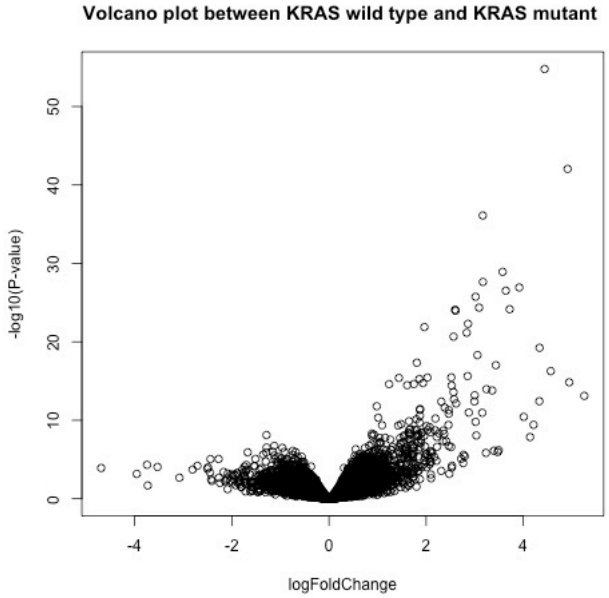


Figure 19. Volcano plot between *KRAS* wild type and *KRAS* mutant of the Canadian cohort. The plot is useful for visualisation of the results of differential expression analysis. For detection of expressed genes, the cut-off points of log fold change > 1.5 and less than < -1.5, and FDR of <0.01 were considered. Genes that are highly dysregulated are to the far left and right sides.

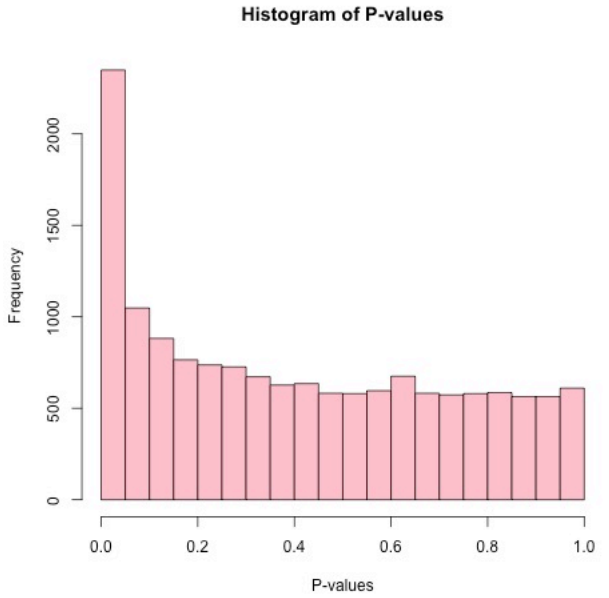


Figure 20. Histogram of p-values of the Australian cohort. The figure showed the p-value distribution of differentially expressed genes in Australian cohort.

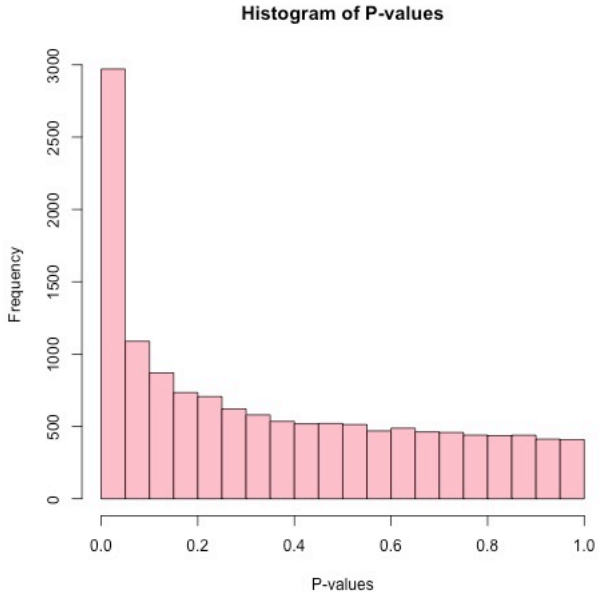


Figure 21. Histogram of p-values of the Canadian cohort. The figure shows the p-value distribution of differentially expressed genes in Canadian cohort.

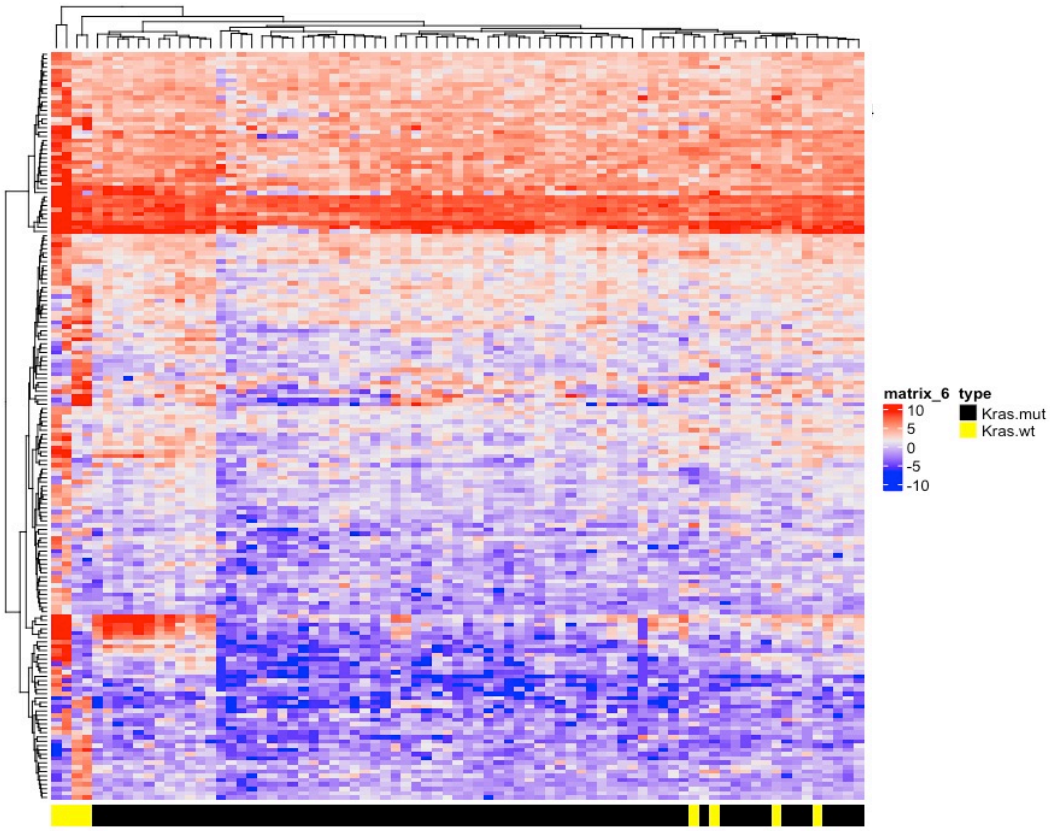


Figure 22. Heatmap of all up-regulated genes in Australian cohort. The heatmap depicts shared patterns of differentially expressed genes between the *KRAS* wild type and *KRAS* mutant in Australian cohort.

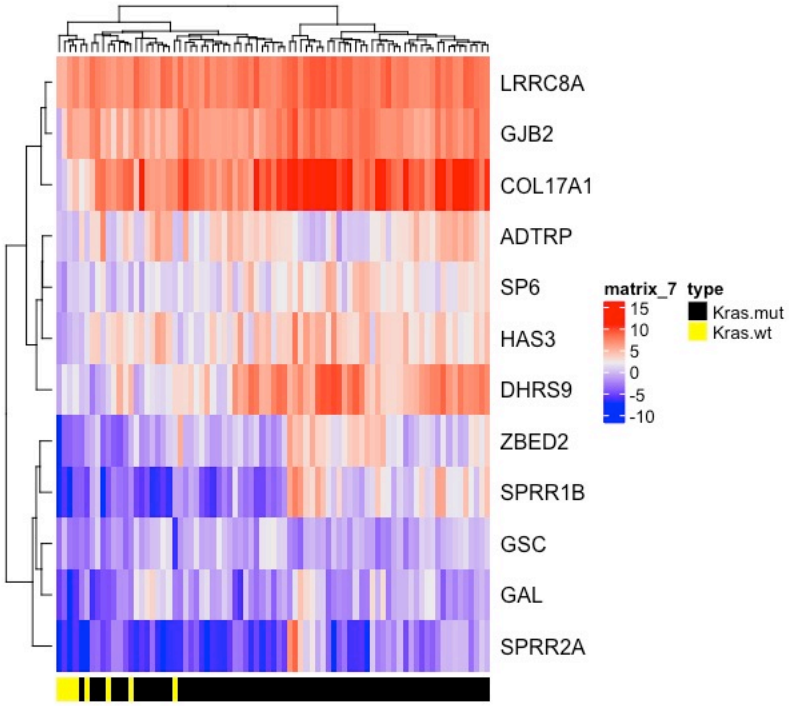


Figure 23. Heatmap of downregulated genes in Australian cohort. The heatmap depicts shared patterns of differentially expressed genes between the *KRAS* wild type and *KRAS* mutant in Australian cohort.

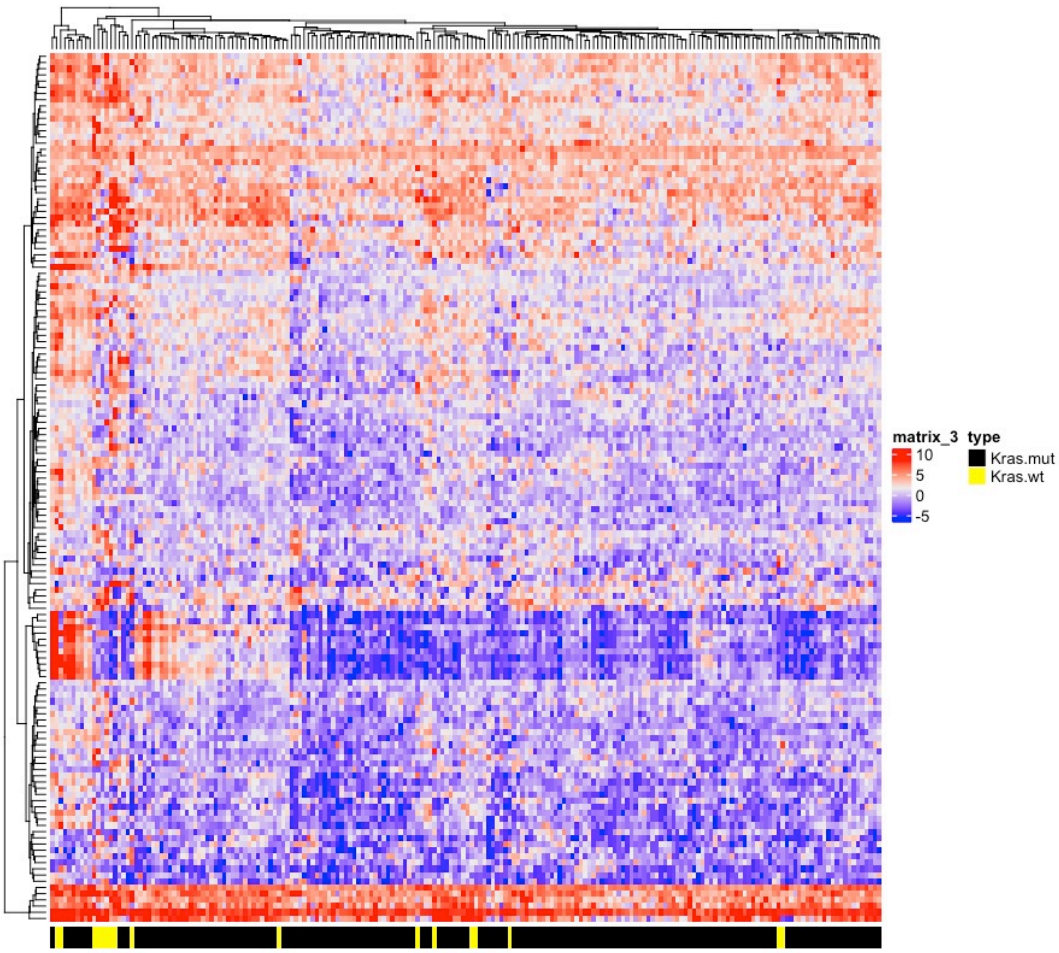


Figure 24. Heatmap of all up-regulated genes in Canadian cohort. The heatmap depicts shared patterns of differentially expressed genes between the *KRAS* wild type and *KRAS* mutant in Canadian cohort.

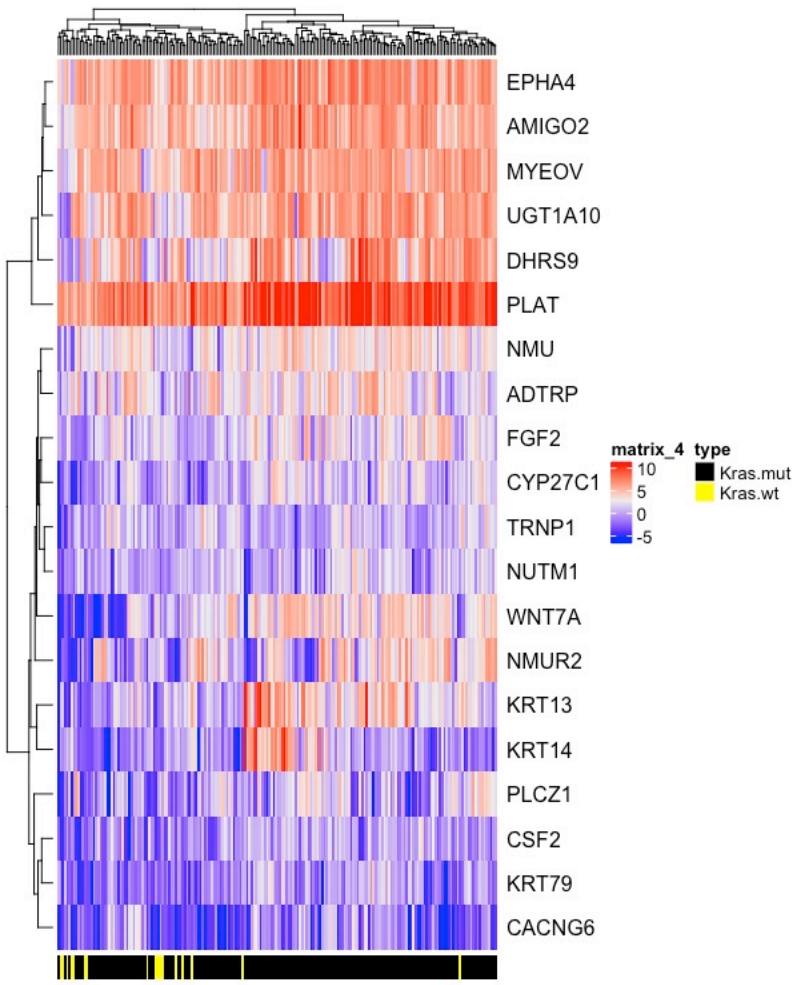


Figure 25. Heatmap of down-regulated genes in Canadian cohort. The heatmap depicts shared patterns of differentially expressed genes between the *KRAS* wild type and *KRAS* mutant in Canadian cohort.

4.6 Pathway Analysis

Tables 8, 9, 10 and 11 show the lists of top up- and down-regulated genes in the Australian and Canadian cohorts, based on KEGG pathway analysis. The results of pathway analysis demonstrate that Calcium signalling ($p = 0.001$) and CAMP ($p = 0.04$) were up-regulated in the *KRAS* wild type cohort of the Australian data set, while Calcium signalling ($p = 0.001$), PI3K-Akt signalling ($p = 0.005$) and Wnt-signalling ($p = 0.03$) were up-regulated in the *KRAS* wild type cohort of the Canadian data set. Additionally, the MAP kinase signalling pathway ($p = 0.04$) was down-regulated in the *KRAS* wild type cohort of the Canadian data set.

Table 8. The 17 significantly up-regulated pathways based on KEGG pathway analysis in the Australian Cohort. The Calcium signalling (p=0.001) with 7 genes including *CACNA1B, CAMK2B, ERBB4, ITPR2, P2RX1, PDE1C, SLC8A3* and CAMP signalling pathways (p = 0.04) in *CAMK2B, FFAR2, GABBR2, RAPGEF4, VIPR2* were up-regulated in the *KRAS* wild type cohort.

Pathway ID	Pathway Name	Number of Genes	Number of DE genes	P_value of DE
path:hsa04974	Protein digestion and absorption	81	7	4.25E-05
path:hsa04740	Olfactory transduction	35	5	4.88E-05
path:hsa04972	Pancreatic secretion	84	6	0.00043206
path:hsa04020	Calcium signalling pathway	149	7	0.00177643
path:hsa04721	Synaptic vesicle cycle	61	4	0.00547726
path:hsa04727	GABAergic synapse	65	4	0.00685669
path:hsa05032	Morphine addiction	68	4	0.00802991
path:hsa04724	Glutamatergic synapse	91	4	0.02148852
path:hsa04966	Collecting duct acid secretion	20	2	0.02220565
path:hsa04725	Cholinergic synapse	93	4	0.02306418
path:hsa04924	Renin secretion	60	3	0.03250357
path:hsa04728	Dopaminergic synapse	109	4	0.03817901
path:hsa05323	Rheumatoid arthritis	68	3	0.04454267
path:hsa00512	Mucin type O-glycan biosynthesis	30	2	0.04715623
path:hsa04024	cAMP signalling pathway	169	5	0.04728544
path:hsa00260	Glycine, serine and threonine metabolism	32	2	0.05297096
path:hsa04911	Insulin secretion	76	3	0.05848555

Table 9. The 4-top down-regulated pathways based on KEGG pathway analysis in the Australian cohort. There were no important down-regulated pathways in the *KRAS* wild type cohort.

Pathway ID	Pathway Name	Number of Genes	Number of DE genes	P_value of DE
path:hsa00830	Retinol metabolism	48	1	0.03799636
path:hsa04974	Protein digestion and absorption	81	1	0.06334615
path:hsa04080	Neuroactive ligand-receptor interaction	170	1	0.12869243
path:hsa01100	Metabolic pathways	1095	1	0.60003566

Table 10 . The 24 significant up-regulated pathways in the Canadian cohort. The Calcium signalling (p = 0.001), PI3K-Akt signalling (p = 0.005) in 9 genes including *CCND2, CCNE1, COL9A3, ERBB4, KIT, LAMA1, NTRK2, RELN, VTN* and WNT-signalling (p = 0.03) in *CCND2, DKK4, NKD1* were up- regulated in the *KRAS* wild type cohort. The 7 genes that were up-regulated in Calcium signalling pathway include *CACNA1B, EDNRB, ERBB4, ITPR2, P2RX1, PDE1C, TACR1*.

Pathway ID	Pathway Name	Number of Genes	Number of DE genes	P_value of DE
path:hsa04972	Pancreatic secretion	76	11	2.88E-10
path:hsa04974	Protein digestion and absorption	73	9	5.40E-08
path:hsa04020	Calcium signalling pathway	128	7	0.00034598
path:hsa04512	ECM-receptor interaction	70	5	0.00074905
path:hsa04080	Neuroactive ligand-receptor interaction	122	6	0.00159632
path:hsa04151	PI3K-Akt signalling pathway	268	9	0.00177241
path:hsa02010	ABC transporters	35	3	0.00543682
path:hsa04614	Renin-angiotensin system	16	2	0.01140724
path:hsa00053	Ascorbate and aldarate metabolism	19	2	0.01593271
path:hsa00561	Glycerolipid metabolism	53	3	0.01702846
path:hsa04924	Renin secretion	56	3	0.01970888
path:hsa05165	Human papillomavirus infection	266	7	0.01980947
path:hsa04721	Synaptic vesicle cycle	57	3	0.02065111
path:hsa04360	Axon guidance	156	5	0.02225853
path:hsa01523	Antifolate resistance	24	2	0.02487373
path:hsa04975	Fat digestion and absorption	28	2	0.033174
path:hsa04510	Focal adhesion	176	5	0.03493359
path:hsa04310	Wnt signalling pathway	125	4	0.03969836
path:hsa00260	Glycine, serine and threonine metabolism	31	2	0.04000738
path:hsa04726	Serotonergic synapse	76	3	0.04316669
path:hsa00380	Tryptophan metabolism	33	2	0.04483194
path:hsa00140	Steroid hormone biosynthesis	34	2	0.04732094
path:hsa04725	Cholinergic synapse	82	3	0.05205479
path:hsa04724	Glutamatergic synapse	82	3	0.05205479

Table 11 . The 14 significantly down-regulated pathways based on KEGG pathway analysis in the Canadian Cohort. The MAP-Kinase signalling pathway (p = 0.04) was down-regulated in the *KRAS* wild type cohort in two genes including *FGF2* and *CACNG6*.

Pathway ID	Pathway Name	Number of Genes	Number of DE genes	P_value of DE
path:hsa00830	Retinol metabolism	45	3	3.67E-05
path:hsa04915	Estrogen signalling pathway	98	2	0.00891132
path:hsa04550	Signalling pathways regulating pluripotency of stem cells	106	2	0.01036098
path:hsa05224	Breast cancer	113	2	0.01170966
path:hsa05226	Gastric cancer	117	2	0.01251335
path:hsa04080	Neuroactive ligand-receptor interaction	122	2	0.01355119
path:hsa05167	Kaposi sarcoma-associated herpesvirus infection	141	2	0.01782277
path:hsa05202	Transcriptional misregulation in cancer	145	2	0.01878637
path:hsa05205	Proteoglycans in cancer	171	2	0.02556908
path:hsa00053	Ascorbate and aldarate metabolism	19	1	0.02749265
path:hsa00040	Pentose and glucuronate interconversions	25	1	0.03602429
path:hsa00860	Porphyrin and chlorophyll metabolism	32	1	0.04588802
path:hsa04010	MAPK signalling pathway	242	2	0.04822895
path:hsa00140	Steroid hormone biosynthesis	34	1	0.04868856

4.7 Summary of Main Findings in the MAPK Pathway

In the *KRAS* wild type cohort, there were samples that most frequently mutated in the MAP kinase pathway component. In the SNV analysis, the genes which were identified as a good candidate in the absence of *RAS* mutation with quite high frequency are *GNAS BRAF* and *RET.* Other low frequency mutated genes include *NRAS* and *MAP2K1.* Recurrent copy number gains (CNV >4) in the component of MAPK pathway were observed in *AKT2*, *EGFR*, *ERBB2*, *RICTOR* , *FGFR1*, *FGF3* and *FGF4* in the *KRAS* wild type cohort. Other important findings are candidate fusion sets. The MAPK pathway candidate fusion sets are *RET-CCDC6; ROS1-SLC4A4; BRAF-SND1; BRAF-SDK1; TRIM24-BRAF; STK4-SLC13A3; ARHGAP24-MAPk10; BRAF-BRAF; STMN1-CDK5RAP3,* and*; SLC4A4-RASGRF1*, which may empower the importance of high frequency of fusions in the MAP kinase signalling pathway.

In the RNA sequencing analysis, RNA expression and Pathway analysis in the whole *KRAS* wild type cohort showed no direct over expression of the MAPK signalling pathway. This finding highlights the results of the genomic analysis section that substitution of one oncogene with another oncogene in the same pathway will not change the expression profile of the pathway. Pathway analysis results for the Australian and Canadian wild type cohorts suggest that Calcium signalling is enriched. This supports the identification of the *GNAS* mutation in the genomic data analysis.

Chapter 5

5 Discussion

KRAS wild type analysis has been surveyed in some previous studies. However, most of these studies have a small sample size and the identified oncogenes were not consistent in all of the studies. In other words, previous studies have been underpowered due to their small size. This study is the first with sufficient numbers to robustly identify drivers in *KRAS* wildtype patients. The present thesis provides the genomic and transcriptomic analysis of *KRAS* wild type pancreatic cancer. The data provide a new hypothesis for drivers of *KRAS* wild type pancreatic cancer. The genomic analysis includes SNV, CNV and fusion analysis.

With advances in cancer genomic studies, whole genome and whole exome sequencing of large samples are the main methods of detecting cancer related genetic abnormalities. These cohorts generate large amounts of data in the form of somatic variants which include single-nucleotide variants (SNV) and small insertions/deletions (indels) [149, 150]. However, visualising the complex data has a significant role in genomic research, and it is difficult to generate publication-based plots. One of the tools that address this complexity is Maftools. Maftools was used in the SNV analysis to visualise somatic driver genes with corresponding statistical frequencies [112].

In the SNV analysis step, in the absence of mutation in *KRAS* in the *KRAS* wild type cohort, the most prominent alternative candidate oncogene is *BRAF*, followed by *GNAS* and *RET* which were identified by Maftools [112] and are candidates for involvement in tumourigenesis. The relationship between *KRAS* wild type and *GNAS* hot spot mutation was reported by Raphael et al. (n = 3/10) [91], while the frequency in our cohort was 17% (n = 12/70). Additionally, in the study by Witkiewicz et al., in two out of 5 cases of *KRAS* wild type Pancreatic cancer samples, the identified oncogene was *GNAS* [93].

Other oncogenes that were identified in the SNV analysis step by Maftools were *BRAF* and *RET* which are components of the MAPK signalling pathway. Pihlak et al. reported that *KRAS* wild type tumours seemed to have variations in other components of the *RAS* pathway. This indicates that *KRAS* wild type PDAC, which accounts for 5-7% of tumours, has a driver mutation in the same pathway [151]. Also, Oikonomou et al. demonstrated that *BRAF*V600E mutation has 138 times more oncogenic activity than *BRAF* wild type

[152]. In addition, it is a more powerful oncogene in comparison with *KRAS*G12V [153]. Mutations in the *BRAF* oncogene exist in more than 30% of *KRAS* wild type PDAC samples [86]. The findings of the present research, however, are not supported by the study of Kanda et al., who reported that mutation in *BRAF* happens in a small group of *KRAS* wild type pancreatic cancers [154]. Genetic analysis in autochthonous models demonstrates that the *BRAF* oncogene is adequate to drive PDAC development [59]. The presence of *KRAS* wild type PDAC samples indicates that mutation in *KRAS* is not mandatory for the development of Pancreatic cancer [155]. Recently Grinshpun et al. reported three cases of *KRAS* wild type having mutations in *RET* and *BRAF*. While somatic mutations of *RET* happen in chronic myelomonocytic leukemia, thyroid and lung cancer [92], it is also one of the identified oncogenes in the SNV and fusion analysis of the *KRAS* wild type cohort.

One of the tools that was used in the present research is CGI. CGI webtool reviews somatic point mutations, small insertions/deletions, copy number alterations and/or gene fusions of an individual and predicts those likely to be drivers versus passengers. It also summarises any evidence of the mutations influencing drug response. CGI identifies all known and likely genomic alterations including the analysis of variants of unknown significance. CGI predicts driver mutations on the contribution of every gene and its region in specific cancer types. The prediction is obtained from the analysis of a large cohort of tumours. In addition, when there is no available information for specific cancer types, CGI uses Pan-Cancer information on the corresponding gene [116]. That is to say, CGI works on the individual sample and takes each sample one at a time and then for every event within the patient, it screens the entire set of variants and identifies the genes that are likely to be drivers based on the type of mutation and prior knowledge. Moreover, the identification of known driver alterations and the biomarkers of drug response are applied by making a match between the tumour type in which they have been described and one of the samples under analysis. CGI interprets mutations from any sequencing platform including exome, whole genome, gene panel sequencing, and VCF [116].

CGI is one of the tools that is used in the analysis of SNV to identify driver mutations from passenger mutations. It brings opportunities that cannot be identified in Maftools analysis such as driver or passenger mutation status or drug response and biomarker information. It also gives detailed explanations about the position of mutations in the

genome, transcriptome and proteins as well as the ExAC database allele frequencies of each variant. It prepares additional information about the identified drivers and specifies whether they are located in TIER 1 or TIER 2. The details about how oncodriveMUT in CGI classifies identified drivers into TIER1 or TIER2 explained in method section [116].

The next section in the genomic analysis is the analysis of copy number variation. A copy number variation (CNV) is a variation in the genome with a size from 50 bp up to the whole chromosome [156]. A CNV is one of the structural variation types due to unbalanced changes in a large proportion of DNA. Copy number gains or losses are structural genomic duplications or deletions [157]. In the copy number variation analysis, copy number >= 5 was picked to minimise the effect of tetraploid samples on the copy number. The results of *KRAS* wild type and *KRAS* mutant have major differences. In addition, high gain amplification in *EGFR* (12%), *ERBB2* (6%), *AKT2* (16%), *RICTOR* (12%), *FGFR1* (10%), *FGF3* (9%), *FGF4* (9%), *MYC* (23%), *CDK6* (16%) and *MDM4* (11%) was seen in the top list of the copy number amplification file. This finding aligns with some of the previous studies. El-Deiry et al. reported a *HER2*/neu amplification in *KRAS* wild type tumours of colorectal cancers [108]. In another study, Ross et al. reported that more than 5% of left colonic tumours with *RAS* wild type have *ERBB2* amplification [109]. Pan-cancer analyses of at least 12 cancer types estimated 14% frequency of *MYC* amplification [158], which is 23% in the *KRAS* wild type cohort. In addition, the frequency of *HER2* amplification in Pancreatic ductal adenocarcinoma is 2% [159], while the frequency of *HER2* in *KRAS* wild type is 6%. These findings are suggestive of suitable driver oncogenes in the absence of *KRAS* for initiation and progression of Pancreatic cancer.

In the fusion analysis two tools were used. Fusion Hub is one of the tools that provide comprehensive information about previously identified fusions in the published databases. It facilitated comparisons between the fusions of the current study and the known fusions. The results of Fusion Hub predictions come from the reports of gene fusions from the previous published literature. This dataset gives a link to the databases, papers, visualisations of the fusions and even information about the fusion gene network. In addition, using Oncofuse gives the opportunity to identify known and novel oncogenes. It also provides information about statistical probability of driver fusions, 'in-frame' status information of breakpoints by score for each fusion by applying Bayesian classifier,

and information about domains of each fusion. It seems that using both of these databases complement each other.

Fusions play an important role in the *KRAS* wild type cohort. There are several candidate fusions identified in this study which had not previously been reported and are novel candidates in Pancreatic cancer, many of which promote the MAPK pathway. These candidate fusions are low frequency and high impact that may drive tumorigenesis. There is a chance of false positives in the prediction of the tools that were applied, as there have been in some other prediction tools [160, 161] , if they are correct, they may drive cancers. Unique candidate oncogenes from the present research which sit in the MAP Kinase pathway include *RET-CCDC6; ROS1-SLC4A4; BRAF-SND1; BRAF-SDK1; TRIM24-BRAF; STK4-SLC13A3; ARHGAP24-MAPK10; BRAF-BRAF; STMN1-CDK5RAP3,* and*; SLC4A4-RASGRF*1.

One of the important functions of cancer genomic research is to find mutations that are druggable. This has implications in personalised medicine. Some of the fusion oncogenes identified in this work are druggable. Among these fusions, some of the attractive fusions are druggable kinase fusions that can be treated with clinically available drugs offering good promise to help patients with this deadly disease [162]. Fusions in *BRAF* are rare and can be targeted [163]. In *SND1-BRAF*, combined treatment with MET inhibitor and RAF inhibitor or MEK inhibitor alone target the phosphorylation of ERK and block cell growth [164]. Oncogenic *BRAF* fusions have been reported in melanoma, glioma, thyroid cancer, pancreatic cancer, non-small cell lung cancer, colorectal cancer and breast cancer [163]. *TRIM24-BRAF* is another druggable fusion which was found in two cases of the *KRAS* wild type cohort. It was reported in melanoma and is sensitive to MEK inhibitors [165]. Interestingly, one of the *KRAS* wild type samples has a fusion in *BRAF-BRAF*, which causes deletion of many regions in *BRAF,* but the kinase domain remains. In addition, many genes are partners of *ALK, ROS*, and *RET* in lung cancer [166]. These fusions are sensitive to tyrosine kinase inhibitors [167]. The *RET* is a therapeutic target in lung and thyroid cancer. Especially, *CCDC6-RET* is frequent in thyroid cancer [168]. This fusion was identified by CGI and Fusion Hub analysis, while being confirmed in the FISH analysis for the corresponding sample. However, this fusion was not identified in the Oncofuse analysis. This is the first report identifying *RET* fusion in Pancreatic cancer. It stablishes that this fusion is also a rare fusion that is found in the Pancreatic cancer

KRAS wild type cohort. Table 12 summarises the candidate oncogenes that are clinically significant and druggable in Pancreatic cancer and other types of cancer.

Table 12. Candidate oncogenes with clinical significance.

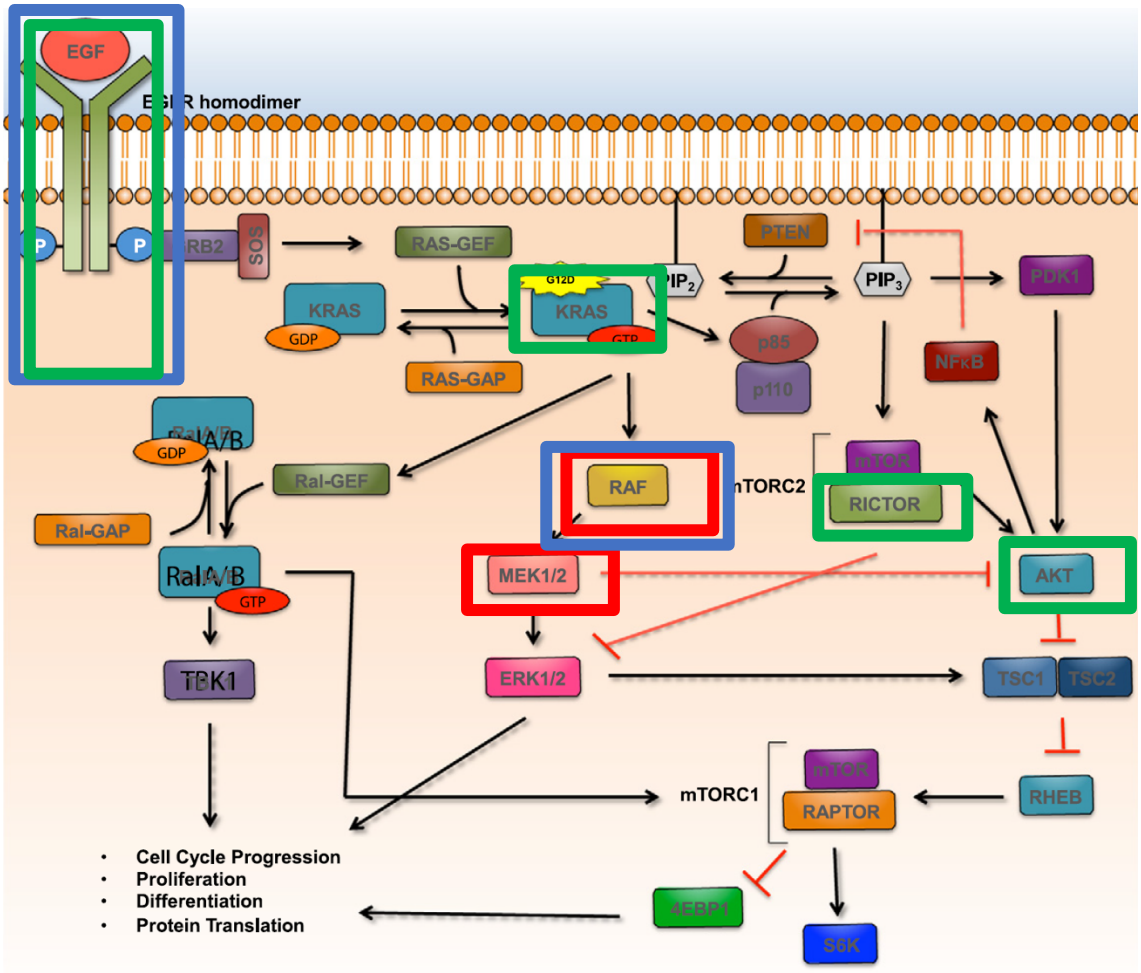
Druggable Targets	Type of Cancer	Sensitivity to	Reference
CCDC6_RET	Thyroid	Combined therapy including TKIs	[169]
CCDC6_RET	Lung	TKIs (Sunitinib &Sorafenib)	[170]
SND1_BRAF	Pancreatic Acinar cell Carcinoma	MEK inhibitor (trametinib)	[171]
SND1_BRAF	Lung	MEK inhibitor (trametinib)	[172]
TRIM24-BRAF	Melanoma	MEK inhibitor (trametinib)	[165]
NRG1 Fusions	NSCLC	anti-HER3 monoclonal antibodies or Tarloxotinib	[173]
Ros1Fusions	NSCLC	Crizotinib	[174]
EGFR amplification	NSCLC	EGFR-TKI	[175]
RICTOR amplification	SCLC	mTORC1/2 inhibitors	[176]
MET amplification	Several neoplasms	MET-TKIs	[177, 178]
ERBB2 amplification	Breast and Gastric	Trastuzumab	[179, 180]
FGF3 and FGF4 amplification	Hepatocellular and Lung Carcinoma	Sorafenib	[181, 182]

In the RNA sequencing analysis, differential expression analysis in the Australian cohort showed 173 genes up-regulated and 12 genes downregulated in the *KRAS* wild type group. Additionally, in the Canadian cohort, 140 genes were up-regulated, and 20 genes were down-regulated in the *KRAS* wild type cohort. Plotting of DE genes in two cohorts did not show any specific clustering between *KRAS* wild type and *KRAS* mutant.
Pathway analysis of differentially expressed genes suggests, in the Australian cohort, *KRAS* wild type was overexpressed in the calcium and CAMP signalling pathway. RNA sequencing analysis of the Canadian cohort with larger sample size showed that calcium signalling, Wnt-signalling and PI3K-Akt signalling were over expressed in the Canadian *KRAS* wild type cohort, while the MAP kinase signalling pathway was under expressed in two genes, *FGF2* and *CACNG6. CACNG6* is one of the calcium signalling genes. Over expression of the PI3K-Akt signalling pathway is consistent with the result of a study by Raphael et al. [84] and a study by Lievre et al. in *RAS* wild type colorectal cancer [183]. In addition, overexpression of the Wnt-signalling and Calcium signalling pathways are correlated together. In other words, two non-canonical Wnt-signalling pathways have been identified. The Wnt/Ca^{2+} signalling pathway regulates calcium release from endoplasmic reticulum for controlling inner cell calcium. Some of the studies reported that cancer cells use non-canonical Wnt signalling to metastasise [184]. Wnt-5a correlates with advanced gastric, melanoma and breast cancers.[185-187]. The result of pathway analysis suggests that Calcium signalling plays a significant role in the *KRAS* wild type cohort.

CaM/Calcium regulates the RAF/MEK/ERK and PI3K/AKT signalling pathways [188, 189]. Calcium and MAPK signalling pathways have overlapped genes. An increase in the concentration of intracellular calcium causes activation of *PKC*, which in turn phosphorylates *RAS* and activates downstream signalling of *MEK* and *ERK* [189]. In addition, CaM activates *PI3Kα* on the membrane of the cell which leads to activation of Akt and downstream components through the mTOR signalling pathway [188]. Joyal et al. suggested that CaM activates *PI3K* directly [190]. CaM is overexpressed in Pancreatic cancer [188], colorectal cancer [191] and lung cancer [192] which is suggestive of the role of calcium in initiation and progression of these ductal cancers [188] .

The current project has three hypotheses. The first hypothesis is that alternative oncogenes drive tumorigenesis in the absence of *KRAS* mutation in the *KRAS* wild type

group. The second hypothesis is that oncogenes are involved in MAPK pathway. The obvious oncogene in the *KRAS* mutant cohort is *KRAS*. Considering all types of mutations which frequently occur in the MAP kinase pathway in the genomic analysis including SNV, CNV and fusion, our data from the genomic step analysis is suggestive of involvement of the MAP kinase pathway in *KRAS* wild type and supportive of the first and second hypotheses. It suggests that MAPK is a dominant regulator of the *KRAS* wild type cohort. In other words, in the SNV analysis, the identified oncogenes were *BRAF, GNAS* and *RET*. In addition, high gain copy number amplification in the *KRAS* wild type cohort such as *MYC* and other genes in the MAP kinase signalling pathway such as *EGFR, ERBB2, AKT2, RICTOR, FGFR1, FGF3, FGF4* are suggestive of good alternative candidates in the absence of *KRAS*. They can cause initiation and progression of Pancreatic cancer in the absence of *KRAS* mutation. Except for *MYC*, the rest of these genes are located in the MAPK pathway, which again highlights the importance of this pathway in the *KRAS* wild type cohort. Rare fusions were identified in the *KRAS* wild type cohort. Some of the identified fusions are clinically validated for drug response. The following fusions were identified in the *KRAS* wild type cohort and have one component from the MAPK pathway: *RET-CCDC6; ROS1-SLC4A4; BRAF-SND1; BRAF-SDK1; TRIM24-BRAF; STK4-SLC13A3; ARHGAP24-MAPk10; BRAF-BRAF; STMN1-CDK5RAP3,* and*; SLC4A4-RASGRF1*. The occurrence of these rare fusions in the *KRAS* wild type cohort is further evidence supporting the first and second hypotheses of the current research. These oncogenic fusions, which are sited in the MAPK pathway in the *KRAS* wild type cohort, are good candidates for initiation and progression of Pancreatic cancer in the absence of *KRAS.*



Mutations: BRAF, NRAS, RET, MAP2K1.
Fusions: RET, ROS1, BRAF, NRG1.
Amplifications: EGFR, FGF3, FGF4, ERBB2, RICTOR, AKT.

Figure 26 . Overview of the MAPK Pathway. Components of MAPK pathway in *KRAS* wild type cohort that have alterations were highlighted [54].

The RNA sequencing analysis in the whole cohort of *KRAS* wild type showed no difference between *KRAS* wild type and *KRAS* mutant. This is because mutations in genes in the MAP kinase pathway other than *KRAS* can have the same signature as *KRAS*. In other words, the same pathway has been broken and one oncogene has been substituted for another oncogene in the same pathway. The lack of differences between *KRAS* mutant and *KRAS* wildtype in the MAPK pathway could suggest that the MAPK pathway is upregulated in both sets, resulting in a lack of difference in the expression.

The third hypothesis was that the oncogenic mutations occur in the MAPK independent pathway. This hypothesis was in line with some of the results from the genomic and transcriptomic analysis. One of the oncogenes which was identified in the genomic analysis at the SNV level using Maftools and CGI analysis was *GNAS*. Moreover, the signalling pathway that was identified in both the Australian and Canadian cohorts was calcium signalling, which is supportive of involvement of *GNAS* at the genomic level. Moreover, Calcium signalling regulates the MAPK signalling pathway through two branches, the RAF/MEK/ERK and PI3K/AKT/mTOR signalling pathways. Additionally, the calcium and MAPK signalling pathways have some overlapped genes. It seems that calcium signalling is still supportive of the role of MAPK in the *KRAS* wild type cohort.

The current study brings together samples from two different consortiums and three different cohorts. The mutation calls that were analysed in three cohorts may have sequencing data of varying quality and coverage from samples that have variation in cellularity. They may have relied on different combinations of variant calling algorithms and also different experimental biases which can potentially cause bias in the results.

5.1 Importance and Conclusions

In the *KRAS* wild type cohort, there were samples that most frequently mutated in the MAP kinase pathway component. Finding alternative oncogenes for *KRAS* potentially have applications for all cohorts of pancreatic cancer. One of the genes which was identified as a good candidate in the absence of *RAS* mutation is *GNAS*. As described in Figure 8, the frequency of this gene is quite high (17%) and it is the first Oncogene identified in the list. *KRAS* wild type has a small sample size and lower power than the *KRAS* mutant cohort in the current dataset. For GNAS to be shown to be the top candidate gene in the small sample size of *KRAS* wild type is a very promising finding. This is consistent with the observations of Raphael et al. and Witkiewicz et al. who reported that GNAS mutations are enriched in *KRAS* wild type samples.

Other important findings are candidate fusion sets which had not previously been reported in *KRAS* wild type cohort Pancreatic cancer. For example, *RET_CCDC6*, which is druggable, was identified as a candidate fusion oncogene. The map kinase pathway

candidate fusion sets are *RET-CCDC6; ROS1-SLC4A4; BRAF-SND1; BRAF-SDK1; TRIM24-BRAF; STK4-SLC13A3; ARHGAP24-MAPk10; BRAF-BRAF; STMN1-CDK5RAP3,* and*; SLC4A4-RASGRF1*, which may empower the importance of high frequency of fusions in the MAP kinase signalling pathway.

There is a chance of false positives in the prediction of the results by the applied tools in fusion analysis, as has been found with some other prediction tools [160, 161] . However, if validation tests confirm the candidate fusion results, the fusion events in the MAP kinase pathway will be promising candidate oncogenic alternatives in the absence of *KRAS* mutation in *KRAS* wild type cohort.

In the RNA sequencing analysis, RNA expression and Pathway analysis in the whole *KRAS* wild type cohort showed no direct over expression of the MAPK signalling pathway. This finding highlights the results of the genomic analysis section that substitution of one oncogene with another oncogene in the same pathway will not change the expression profile of the pathway.

Pathway analysis results for the Australian and Canadian wild type cohorts suggest that Calcium signalling is enriched. This supports the identification of the *GNAS* mutation in the genomic data analysis. In addition, in the Australian *KRAS* wild type cohort, CAMP was enriched. In the Canadian *KRAS* wild type cohort, PI3K-Akt, and Wnt signalling were up-regulated and the MAP kinase signalling pathway was down-regulated.

5.2 Future Perspectives

Several improvements may be undertaken in the future. The first aspect that remains unresolved is to test the significance assessments of observed mutations in SNV, CNV and SV in *KRAS* wild type and compare them with previous literature reviews. In addition, it would be interesting to study the overall frequency of copy number variations between two groups of *KRAS* wild type and *KRAS* mutant. Moreover, the data of the current study bring two consortiums and three different cohorts. These samples may have sequencing data of varying quality and coverage from samples that have variation in cellularity and may have relied on different combinations of variant calling algorithms

which can potentially cause bias. Hence, a uniform re-analysis of these cohorts by uniform pipeline is necessary to provide directly comparable variant data.

The fusion section needs further validation. In other words, the fusions that were found in the genomic analysis need to be validated to see if they are also expressed in the transcriptome. Furthermore, it would be valuable to check the reliability of these candidate fusions, with reference to the confidence of the structural variant calls, and sequencing depth. The candidate fusions should then be confirmed experimentally. There are a number of new tools that are more efficient at detecting a wider range of structural variants such as Manta [193] that can be used.

Another aspect for future work is the removal of the batch effect in the transcriptomic data. The focus of the thesis was more about genomic and transcriptomic analysis but the different roles of non-coding regions and epigenetics on transcriptomic profiles of these groups of patients remain unsolved. The demographic features of the *KRAS* wild type group is another section that should be regarded in the future direction of this project.

Another area which remains is validation of the fusion results by making engineered models of the identified oncogene fusions in *KRAS* wild type Pancreatic cell lines to determine whether or not the phenotype will be the same. In addition, testing can be undertaken to determine whether the identified mutations are druggable in Pancreatic cancer.

Last but not least, the *KRAS* wild type group is a clinically important group of samples for checking different therapy for all groups of pancreatic cancer.

6 References

1. Thomas, D., et al., Truncated O-glycans promote epithelial-to-mesenchymal transition and stemness properties of pancreatic cancer cells. J Cell Mol Med, 2019.
2. Health, A.I.o. and Welfare, Australian cancer incidence and mortality (ACIM) books. Bladder cancer., 2014.
3. Rahib, L., et al., Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. Cancer research, 2014.
4. Mishra, N.K., S. Southekal, and C. Guda, Survival Analysis of Multi-Omics Data Identifies Potential Prognostic Markers of Pancreatic Ductal Adenocarcinoma. Front Genet, 2019. 10: p. 624.
5. Sunami, Y. and J. Kleeff, Immunotherapy of pancreatic cancer. Prog Mol Biol Transl Sci, 2019. 164: p. 189-216.
6. Zeitouni, D., et al., KRAS Mutant Pancreatic Cancer: No Lone Path to an Effective Treatment. Cancers (Basel), 2016. 8(4).
7. Waters, A.M. and C.J. Der, KRAS: The Critical Driver and Therapeutic Target for Pancreatic Cancer. Cold Spring Harb Perspect Med, 2018. 8(9).
8. Bryant, K.L., et al., KRAS: feeding pancreatic cancer proliferation. Trends Biochem Sci, 2014. 39(2): p. 91-100.
9. Aguirre, A.J., Oncogenic NRG1 fusions: A new hope for targeted therapy in pancreatic cancer. Clinical Cancer Research, 2019. 25(15): p. 4589-4591.
10. Heining, C., et al., NRG1 Fusions in KRAS Wild-Type Pancreatic Cancer. Cancer Discov, 2018. 8(9): p. 1087-1095.
11. Becker, A.E., et al., Pancreatic ductal adenocarcinoma: risk factors, screening, and early detection. World journal of gastroenterology: WJG, 2014. 20(32): p. 11182.
12. Collisson, E.A., et al., Molecular subtypes of pancreatic cancer. Nature Reviews Gastroenterology & Hepatology, 2019: p. 1.
13. Lowenfels, A.B. and P. Maisonneuve, Epidemiology and risk factors for pancreatic cancer. Best Pract Res Clin Gastroenterol, 2006. 20(2): p. 197-209.
14. Heining, C., et al., NRG1 fusions in KRAS wild-type pancreatic cancer. Cancer discovery, 2018. 8(9): p. 1087-1095.
15. Vincent, A., et al., Pancreatic cancer. The lancet, 2011. 378(9791): p. 607-620.
16. Capasso, M., et al., Epidemiology and risk factors of pancreatic cancer. Acta Bio Medica Atenei Parmensis, 2018. 89(9-S): p. 141-146.
17. Hidalgo, M., Pancreatic cancer. N Engl J Med, 2010. 362(17): p. 1605-17.
18. Chhoda, A., et al., Current Approaches to Pancreatic Cancer Screening. The American journal of pathology, 2019. 189(1): p. 22-35.
19. Duell, E.J., Epidemiology and potential mechanisms of tobacco smoking and heavy alcohol consumption in pancreatic cancer. Molecular carcinogenesis, 2012. 51(1): p. 40-52.
20. Rahn, S., et al., Diabetes as risk factor for pancreatic cancer: Hyperglycemia promotes epithelial-mesenchymal-transition and stem cell properties in pancreatic ductal epithelial cells. Cancer letters, 2018. 415: p. 129-150.
21. Liao, W.-C., et al., Blood glucose concentration and risk of pancreatic cancer: systematic review and dose-response meta-analysis. Bmj, 2015. 349: p. g7371.

22. Zhang, S., et al., Effects of alcohol drinking and smoking on pancreatic ductal adenocarcinoma mortality: A retrospective cohort study consisting of 1783 patients. Sci Rep, 2017. 7(1): p. 9572.
23. Diaz, K.E. and A.L. Lucas, Familial pancreatic ductal adenocarcinoma. The American journal of pathology, 2019. 189(1): p. 36-43.
24. Wang, Y.-T., et al., Association between alcohol intake and the risk of pancreatic cancer: a dose–response meta-analysis of cohort studies. BMC cancer, 2016. 16(1): p. 212.
25. Shakeri, R., et al., Opium use, cigarette smoking, and alcohol consumption in relation to pancreatic cancer. Medicine, 2016. 95(28).
26. Supek, F. and B. Lehner, Clustered Mutation Signatures Reveal that Error-Prone DNA Repair Targets Mutations to Active Genes. Cell, 2017. 170(3): p. 534-547.e23.
27. Li, D., et al., Pancreatic cancer. Lancet, 2004. 363(9414): p. 1049-57.
28. Wolpin, B.M., et al., ABO blood group and the risk of pancreatic cancer. Journal of the National Cancer Institute, 2009. 101(6): p. 424-431.
29. Permuth-Wey, J. and K.M. Egan, Family history is a significant risk factor for pancreatic cancer: results from a systematic review and meta-analysis. Fam Cancer, 2009. 8(2): p. 109-17.
30. van Lier, M.G., et al., High cancer risk in Peutz-Jeghers syndrome: a systematic review and surveillance recommendations. Am J Gastroenterol, 2010. 105(6): p. 1258-64; author reply 1265.
31. LaRusch, J., et al., Whole exome sequencing identifies multiple, complex etiologies in an idiopathic hereditary pancreatitis kindred. Jop, 2012. 13(3): p. 258-62.
32. Yamamoto, H., et al., Genetic and clinical features of human pancreatic ductal adenocarcinomas with widespread microsatellite instability. Cancer Res, 2001. 61(7): p. 3139-44.
33. Stadler, Z.K., et al., Prevalence of BRCA1 and BRCA2 mutations in Ashkenazi Jewish families with breast and pancreatic cancer. Cancer, 2012. 118(2): p. 493-9.
34. Ren, B., X. Liu, and A.A. Suriawinata, Pancreatic Ductal Adenocarcinoma and Its Precursor Lesions: Histopathology, Cytopathology, and Molecular Pathology. The American journal of pathology, 2019. 189(1): p. 9-21.
35. Bosman, F.T., et al., WHO classification of tumours of the digestive system. 2010: World Health Organization.
36. Makohon-Moore, A. and C.A. Iacobuzio-Donahue, Pancreatic cancer biology and genetics from an evolutionary perspective. Nature reviews Cancer, 2016. 16(9): p. 553.
37. Wood, L.D., M.B. Yurgelun, and M.G. Goggins, Genetics of Familial and Sporadic Pancreatic Cancer. Gastroenterology, 2019.
38. Basturk, O., et al., A revised classification system and recommendations from the Baltimore consensus meeting for neoplastic precursor lesions in the pancreas. The American journal of surgical pathology, 2015. 39(12): p. 1730.
39. Kanda, M., et al., Presence of somatic mutations in most early-stage pancreatic intraepithelial neoplasia. Gastroenterology, 2012. 142(4): p. 730-733. e9.
40. van Heek, N.T., et al., Telomere shortening is nearly universal in pancreatic intraepithelial neoplasia. The American journal of pathology, 2002. 161(5): p. 1541-1547.

41. Hata, T., et al., Genome-Wide Somatic Copy Number Alterations and Mutations in High-Grade Pancreatic Intraepithelial Neoplasia. The American journal of pathology, 2018. 188(7): p. 1723-1733.
42. Wood, L.D., M.B. Yurgelun, and M.G. Goggins, Genetics of Familial and Sporadic Pancreatic Cancer. Gastroenterology, 2019. 156(7): p. 2041-2055.
43. Adsay, N.V., et al., Pathologically and biologically distinct types of epithelium in intraductal papillary mucinous neoplasms: delineation of an "intestinal" pathway of carcinogenesis in the pancreas. The American journal of surgical pathology, 2004. 28(7): p. 839-848.
44. 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.
45. Wu, J., et al., Whole-exome sequencing of neoplastic cysts of the pancreas reveals recurrent mutations in components of ubiquitin-dependent pathways. Proceedings of the National Academy of Sciences, 2011. 108(52): p. 21188-21193.
46. Furukawa, T., et al., Classification of types of intraductal papillary-mucinous neoplasm of the pancreas: a consensus study. Virchows Archiv, 2005. 447(5): p. 794-799.
47. Grant, T.J., K. Hua, and A. Singh, Molecular Pathogenesis of Pancreatic Cancer. Prog Mol Biol Transl Sci, 2016. 144: p. 241-275.
48. Wu, J., et al., Recurrent GNAS mutations define an unexpected pathway for pancreatic cyst development. Sci Transl Med, 2011. 3(92): p. 92ra66.
49. Yamaguchi, H., et al., Somatic mutations in PIK3CA and activation of AKT in intraductal tubulopapillary neoplasms of the pancreas. Am J Surg Pathol, 2011. 35(12): p. 1812-7.
50. Riva, G., et al., Histo-molecular oncogenesis of pancreatic cancer: From precancerous lesions to invasive ductal adenocarcinoma. World J Gastrointest Oncol, 2018. 10(10): p. 317-327.
51. Hezel, A.F., et al., Genetics and biology of pancreatic ductal adenocarcinoma. Genes Dev, 2006. 20(10): p. 1218-49.
52. Xiao, Q., et al., Cancer-Associated Fibroblasts in Pancreatic Cancer Are Reprogrammed by Tumor-Induced Alterations in Genomic DNA Methylation. Cancer Res, 2016. 76(18): p. 5395-404.
53. Sherman, M.H., et al., Stromal cues regulate the pancreatic cancer epigenome and metabolome. Proc Natl Acad Sci U S A, 2017. 114(5): p. 1129-1134.
54. Mann, K.M., et al., KRAS-related proteins in pancreatic cancer. Pharmacol Ther, 2016. 168: p. 29-42.
55. Jones, S., et al., Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science, 2008. 321(5897): p. 1801-6.
56. Hruban, R.H., A. Maitra, and M. Goggins, Update on pancreatic intraepithelial neoplasia. Int J Clin Exp Pathol, 2008. 1(4): p. 306-16.
57. Karnoub, A.E. and R.A. Weinberg, Ras oncogenes: split personalities. Nat Rev Mol Cell Biol, 2008. 9(7): p. 517-31.
58. Calhoun, E.S., et al., BRAF and FBXW7 (CDC4, FBW7, AGO, SEL10) mutations in distinct subsets of pancreatic cancer: potential therapeutic targets. Am J Pathol, 2003. 163(4): p. 1255-60.
59. Collisson, E.A., et al., A central role for RAF-->MEK-->ERK signaling in the genesis of pancreatic ductal adenocarcinoma. Cancer Discov, 2012. 2(8): p. 685-93.

60. Asano, T., et al., The PI 3-kinase/Akt signaling pathway is activated due to aberrant Pten expression and targets transcription factors NF-kappaB and c-Myc in pancreatic cancer cells. Oncogene, 2004. 23(53): p. 8571-80.
61. Altomare, D.A., et al., Frequent activation of AKT2 kinase in human pancreatic carcinomas. J Cell Biochem, 2002. 87(4): p. 470-6.
62. Ferro, E. and L. Trabalzini, RalGDS family members couple Ras to Ral signalling and that's not all. Cell Signal, 2010. 22(12): p. 1804-10.
63. Lim, K.H., et al., Divergent roles for RalA and RalB in malignant growth of human pancreatic carcinoma cells. Curr Biol, 2006. 16(24): p. 2385-94.
64. Olayioye, M.A., et al., The ErbB signaling network: receptor heterodimerization in development and cancer. Embo j, 2000. 19(13): p. 3159-67.
65. Fitzgerald, T.L., et al., Roles of EGFR and KRAS and their downstream signaling pathways in pancreatic cancer and pancreatic cancer stem cells. Advances in biological regulation, 2015. 59: p. 65-81.
66. Larbouret, C., et al., In pancreatic carcinoma, dual EGFR/HER2 targeting with cetuximab/trastuzumab is more effective than treatment with trastuzumab/erlotinib or lapatinib alone: implication of receptors' down-regulation and dimers' disruption. Neoplasia, 2012. 14(2): p. 121-30.
67. Komoto, M., et al., HER2 overexpression correlates with survival after curative resection of pancreatic cancer. Cancer Sci, 2009. 100(7): p. 1243-7.
68. Peverelli, E., et al., cAMP in the pituitary: an old messenger for multiple signals. J Mol Endocrinol, 2014. 52(1): p. R67-77.
69. Patra, K.C., et al., Mutant GNAS drives pancreatic tumourigenesis by inducing PKA-mediated SIK suppression and reprogramming lipid metabolism. Nature cell biology, 2018. 20(7): p. 811.
70. Gomez-Pinilla, P.J., P.J. Camello, and M.J. Pozo, Pancreatic calcium signaling: role in health and disease. Pancreatology, 2009. 9(4): p. 329-333.
71. Tang, H., et al., Genes-environment interactions in obesity- and diabetes-associated pancreatic cancer: a GWAS data analysis. Cancer Epidemiol Biomarkers Prev, 2014. 23(1): p. 98-106.
72. Caldas, C., et al., Frequent somatic mutations and homozygous deletions of the p16 (MTS1) gene in pancreatic adenocarcinoma. Nat Genet, 1994. 8(1): p. 27-32.
73. Gonzalez-Zulueta, M., et al., Methylation of the 5' CpG island of the p16/CDKN2 tumor suppressor gene in normal and transformed human tissues correlates with gene silencing. Cancer Res, 1995. 55(20): p. 4531-5.
74. Wilentz, R.E., et al., Inactivation of the p16 (INK4A) tumor-suppressor gene in pancreatic duct lesions: loss of intranuclear expression. Cancer Res, 1998. 58(20): p. 4740-4.
75. Qiu, W., et al., Disruption of p16 and activation of Kras in pancreas increase ductal adenocarcinoma formation and metastasis in vivo. Oncotarget, 2011. 2(11): p. 862-73.
76. Remmers, N., et al., Molecular pathology of early pancreatic cancer. Cancer Biomark, 2010. 9(1-6): p. 421-40.
77. Hahn, S.A., et al., DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. Science, 1996. 271(5247): p. 350-3.
78. Kojima, K., et al., Inactivation of Smad4 accelerates Kras(G12D)-mediated pancreatic neoplasia. Cancer Res, 2007. 67(17): p. 8121-30.
79. Pelosi, E., G. Castelli, and U. Testa, Pancreatic cancer: molecular characterization, clonal evolution and cancer stem cells. Biomedicines, 2017. 5(4): p. 65.

80. Weinberg, R., The biology of cancer. New York, NY: Garland Science. 2007, Taylor & Francis Group.
81. Morton, J.P., et al., Mutant p53 drives metastasis and overcomes growth arrest/senescence in pancreatic cancer. Proc Natl Acad Sci U S A, 2010. 107(1): p. 246-51.
82. Pishvaian, M.J., et al., Molecular Profiling of Patients with Pancreatic Cancer: Initial Results from the Know Your Tumor Initiative. Clinical Cancer Research, 2018. 24(20): p. 5018-5027.
83. Aguirre, A.J., et al., Real-time genomic characterization of advanced pancreatic cancer to enable precision medicine. Cancer discovery, 2018. 8(9): p. 1096-1111.
84. Raphael, B.J., et al., Integrated genomic characterization of pancreatic ductal adenocarcinoma. Cancer cell, 2017. 32(2): p. 185-203. e13.
85. Barrett, M.T., et al., Clinical study of genomic drivers in pancreatic ductal adenocarcinoma. Br J Cancer, 2017. 117(4): p. 572-582.
86. Witkiewicz, A.K., et al., Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. Nat Commun, 2015. 6: p. 6744.
87. Shimada, Y., et al., An Oncogenic ALK Fusion and an RRAS Mutation in KRAS Mutation-Negative Pancreatic Ductal Adenocarcinoma. Oncologist, 2017. 22(2): p. 158-164.
88. Sausen, M., et al., Clinical implications of genomic alterations in the tumour and circulation of pancreatic cancer patients. Nat Commun, 2015. 6: p. 7686.
89. Heining, C., et al., NRG1 Fusions in KRAS Wild-type Pancreatic Cancer. Cancer discovery, 2018: p. CD-18-0036.
90. Aguirre, A.J., et al., Real-time genomic characterization of advanced pancreatic cancer to enable precision medicine. Cancer discovery, 2018: p. CD-18-0275.
91. Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. Cancer Cell, 2017. 32(2): p. 185-203.e13.
92. Grinshpun, A., et al., Beyond KRAS: Practical Molecular Targets in Pancreatic Adenocarcinoma. Case Rep Oncol, 2019. 12(1): p. 7-13.
93. Witkiewicz, A.K., et al., Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. Nature communications, 2015. 6: p. 6744.
94. Safran, H., et al., Overexpression of the HER-2/neu oncogene in pancreatic adenocarcinoma. American journal of clinical oncology, 2001. 24(5): p. 496-499.
95. Chin, L., J.N. Andersen, and P.A. Futreal, Cancer genomics: from discovery science to personalized medicine. Nat Med, 2011. 17(3): p. 297-303.
96. Kim, S.T., et al., Impact of KRAS mutations on clinical outcomes in pancreatic cancer patients treated with first-line gemcitabine-based chemotherapy. Mol Cancer Ther, 2011. 10(10): p. 1993-9.
97. Windon, A.L., et al., A KRAS wild type mutational status confers a survival advantage in pancreatic ductal adenocarcinoma. Journal of gastrointestinal oncology, 2018. 9(1): p. 1.
98. Schultheis, B., et al., Gemcitabine combined with the monoclonal antibody nimotuzumab is an active first-line regimen in KRAS wildtype patients with locally advanced or metastatic pancreatic cancer: a multicenter, randomized phase IIb study. Ann Oncol, 2017. 28(10): p. 2429-2435.
99. Boeck, S., et al., Gemcitabine plus erlotinib (GE) followed by capecitabine (C) versus capecitabine plus erlotinib (CE) followed by gemcitabine (G) in advanced pancreatic cancer (APC): A randomized, cross-over phase III trial of the

Arbeitsgemeinschaft Internistische Onkologie (AIO). Journal of Clinical Oncology, 2010. 28(18_suppl): p. LBA4011-LBA4011.
100. Guan, J.L., et al., KRAS mutation in patients with lung cancer: a predictor for poor prognosis but not for EGFR-TKIs or chemotherapy. Ann Surg Oncol, 2013. 20(4): p. 1381-8.
101. Shin, D.H., et al., Oncogenic function and clinical implications of SLC3A2-NRG1 fusion in invasive mucinous adenocarcinoma of the lung. Oncotarget, 2016. 7(43): p. 69450-69465.
102. Fernandez-Cuesta, L. and R.K. Thomas, Molecular pathways: targeting NRG1 fusions in lung cancer. Clinical Cancer Research, 2015. 21(9): p. 1989-1994.
103. Shi, E., et al., FGFR1 and NTRK3 actionable alterations in "Wild-Type" gastrointestinal stromal tumors. Journal of translational medicine, 2016. 14(1): p. 339.
104. Le Rolle, A.F., et al., Identification and characterization of RET fusions in advanced colorectal cancer. Oncotarget, 2015. 6(30): p. 28929-37.
105. Nikolaev, S.I., et al., Exome sequencing identifies recurrent somatic MAP2K1 and MAP2K2 mutations in melanoma. Nature genetics, 2012. 44(2): p. 133.
106. Fernandez-Cuesta, L. and R.K. Thomas, Molecular Pathways: Targeting NRG1 Fusions in Lung Cancer. Clin Cancer Res, 2015. 21(9): p. 1989-94.
107. Rau, T.T., et al., Inflammatory response in serrated precursor lesions of the colon classified according to WHO entities, clinical parameters and phenotype-genotype correlation. J Pathol Clin Res, 2016. 2(2): p. 113-24.
108. El-Deiry, W.S., et al., Molecular profiling of 6,892 colorectal cancer samples suggests different possible treatment options specific to metastatic sites. Cancer Biol Ther, 2015. 16(12): p. 1726-37.
109. Ross, J.S., et al., Targeting HER2 in colorectal cancer: The landscape of amplification and short variant mutations in ERBB2 and ERBB3. Cancer, 2018. 124(7): p. 1358-1373.
110. Comprehensive molecular characterization of human colon and rectal cancer. Nature, 2012. 487(7407): p. 330-7.
111. Teng, H., et al., Identification of recurrent and novel mutations by wholegenome sequencing of colorectal tumors from the Han population in Shanghai, eastern China. Mol Med Rep, 2018. 18(6): p. 5361-5370.
112. Mayakonda, A. and H.P. Koeffler, Maftools: Efficient analysis, visualization and summarization of MAF files from large-scale cohort based cancer studies. BioRxiv, 2016: p. 052662.
113. Mayakonda, A. Maftools. https://bioconductor.org/packages/release/bioc/html/maftools.html.
114. Quinlan, A.R. and I.M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics, 2010. 26(6): p. 841-2.
115. Kent, W.J., et al., The human genome browser at UCSC. Genome Res, 2002. 12(6): p. 996-1006.
116. Tamborero, D., et al., Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. Genome Med, 2018. 10(1): p. 25.
117. Tamborero, D., et al., Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. Genome medicine, 2018. 10(1): p. 25.
118. Shih, J.H. and M.P. Fay, Pearson's chi-square test and rank correlation inferences for clustered data. Biometrics, 2017. 73(3): p. 822-834.
119. Kim, H.-Y., Statistical notes for clinical researchers: chi-squared test and Fisher's exact test. Restorative dentistry & endodontics, 2017. 42(2): p. 152-155.

120. Panigrahi, P., A. Jere, and K. Anamika, FusionHub: A unified web platform for annotation and visualization of gene fusion events in human cancer. PloS one, 2018. 13(5): p. e0196588.
121. Djotsa Nono, A.B., K. Chen, and X. Liu, Computational prediction of genetic drivers in cancer. eLS, 2001: p. 1-16.
122. Shugay, M., et al., Oncofuse: a computational framework for the prediction of the oncogenic potential of gene fusions. Bioinformatics, 2013. 29(20): p. 2539-2546.
123. Novo, F.J., I.O. de Mendibil, and J.L. Vizmanos, TICdb: a collection of gene-mapped translocation breakpoints in cancer. BMC Genomics, 2007. 8: p. 33.
124. Huang da, W., B.T. Sherman, and R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc, 2009. 4(1): p. 44-57.
125. Frank, E., et al., Data mining in bioinformatics using Weka. Bioinformatics, 2004. 20(15): p. 2479-81.
126. Shugay, M., et al., Genomic hallmarks of genes involved in chromosomal translocations in hematological cancer. PLoS Comput Biol, 2012. 8(12): p. e1002797.
127. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, 2010. 26(1): p. 139-140.
128. Gentleman, R.C., et al., Bioconductor: open software development for computational biology and bioinformatics. Genome Biol, 2004. 5(10): p. R80.
129. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, 2010. 26(1): p. 139-40.
130. Durinck, S., et al., Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nature protocols, 2009. 4(8): p. 1184.
131. Ringnér, M., What is principal component analysis? Nature biotechnology, 2008. 26(3): p. 303.
132. Karamizadeh, S., et al., An overview of principal component analysis. Journal of Signal and Information Processing, 2013. 4(03): p. 173.
133. Gandolfo, L.C. and T.P. Speed, RLE plots: Visualizing unwanted variation in high dimensional data. PloS one, 2018. 13(2): p. e0191629.
134. Risso, D. EDASeq.https://www.bioconductor.org/packages/release/bioc/html/EDASeq.html.
135. Khatri, P., M. Sirota, and A.J. Butte, Ten years of pathway analysis: current approaches and outstanding challenges. PLoS Comput Biol, 2012. 8(2): p. e1002375.
136. Carlson, M., KEGG. db: A set of annotation maps for KEGG. R package version 3.1. 2. 2016.
137. Cardoso, T.F., et al., RNA-seq based detection of differentially expressed genes in the skeletal muscle of Duroc pigs with distinct lipid profiles. Scientific reports, 2017. 7: p. 40005.
138. Dergunova, L.V., et al., Genome-wide transcriptome analysis using RNA-Seq reveals a large number of differentially expressed genes in a transient MCAO rat model. BMC genomics, 2018. 19(1): p. 655.
139. Li, M., Q. Sun, and X. Wang, Transcriptional landscape of human cancers. Oncotarget, 2017. 8(21): p. 34534.

140. Gao, X., et al., An integrated RNA-Seq and network study reveals the effect of nicotinamide on adrenal androgen synthesis. Clinical and Experimental Pharmacology and Physiology, 2020.
141. Mayakonda, A., et al., Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome research, 2018. 28(11): p. 1747-1756.
142. Tamborero, D., A. Gonzalez-Perez, and N. Lopez-Bigas, OncodriveCLUST: exploiting the positional clustering of somatic mutations to identify cancer genes. Bioinformatics, 2013. 29(18): p. 2238-2244.
143. Pan-cancer analysis of whole genomes. Nature, 2020. 578(7793): p. 82-93.
144. Young, E.L., et al., Universal Panel Testing of Pancreatic Cancer Cases for Cancer Predisposition. bioRxiv, 2017: p. 195537.
145. Waddell, N., et al., Whole genomes redefine the mutational landscape of pancreatic cancer. Nature, 2015. 518(7540): p. 495-501.
146. Chen, H. and P.C. Boutros, VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC bioinformatics, 2011. 12(1): p. 35.
147. Gao, Q., et al., Driver fusions and their implications in the development and treatment of human cancers. Cell reports, 2018. 23(1): p. 227-238. e3.
148. Fonseca, N.A., et al., Comprehensive genome and transcriptome analysis reveals genetic basis for gene fusions in cancer. bioRxiv, 2017: p. 148684.
149. Mardis, E.R. and R.K. Wilson, Cancer genome sequencing: a review. Human molecular genetics, 2009. 18(R2): p. R163-R168.
150. Vogelstein, B., et al., Cancer genome landscapes. science, 2013. 339(6127): p. 1546-1558.
151. Pihlak, R., et al., Advances in Molecular Profiling and Categorisation of Pancreatic Adenocarcinoma and the Implications for Therapy. Cancers (Basel), 2018. 10(1).
152. Davies, H., et al., Mutations of the BRAF gene in human cancer. Nature, 2002. 417(6892): p. 949-54.
153. Oikonomou, E., et al., BRAF(V600E) efficient transformation and induction of microsatellite instability versus KRAS(G12V) induction of senescence markers in human colon cancer cells. Neoplasia, 2009. 11(11): p. 1116-31.
154. Kanda, M., et al., Presence of somatic mutations in most early-stage pancreatic intraepithelial neoplasia. Gastroenterology, 2012. 142(4): p. 730-733.e9.
155. Ying, H., et al., Genetics and biology of pancreatic ductal adenocarcinoma. Genes Dev, 2016. 30(4): p. 355-85.
156. Trost, B., et al., A comprehensive workflow for read depth-based identification of copy-number variation from whole-genome sequence data. The American Journal of Human Genetics, 2018. 102(1): p. 142-155.
157. Wang, H., D. Nettleton, and K. Ying, Copy number variation detection using next generation sequencing read counts. BMC Bioinformatics, 2014. 15: p. 109.
158. Kalkat, M., et al., MYC Deregulation in Primary Human Cancers. Genes (Basel), 2017. 8(6).
159. Chou, A., et al., Clinical and molecular characterization of HER2 amplified-pancreatic cancer. Genome Med, 2013. 5(8): p. 78.
160. Haas, B.J., et al., Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. Genome biology, 2019. 20(1): p. 213.
161. Carrara, M., et al., State-of-the-art fusion-finder algorithms sensitivity and specificity. BioMed research international, 2013. 2013.

162. Wu, C.-C., et al., FusionPathway: Prediction of pathways and therapeutic targets associated with gene fusions in cancer. PLoS computational biology, 2018. 14(7): p. e1006266.
163. Ross, J.S., et al., The distribution of BRAF gene fusions in solid tumors and response to targeted therapy. International journal of cancer, 2016. 138(4): p. 881-890.
164. Lee, N.V., et al., A novel SND1-BRAF fusion confers resistance to c-Met inhibitor PF-04217903 in GTL16 cells though MAPK activation. PloS one, 2012. 7(6): p. e39653.
165. Hutchinson, K.E., et al., BRAF fusions define a distinct molecular subset of melanomas with potential sensitivity to MEK inhibition. Clin Cancer Res, 2013. 19(24): p. 6696-702.
166. Cerrato, A., R. Visconti, and A. Celetti, The rationale for druggability of CCDC6-tyrosine kinase fusions in lung cancer. Molecular cancer, 2018. 17(1): p. 46.
167. Montor, W.R., A.R.O.S.E. Salas, and F.H.M. de Melo, Receptor tyrosine kinases and downstream pathways as druggable targets for cancer treatment: the current arsenal of inhibitors. Molecular cancer, 2018. 17(1): p. 55.
168. Hu, X., et al., TumorFusions: an integrative resource for reporting cancer-associated transcript fusions in 33 tumor types. bioRxiv, 2017: p. 162180.
169. Cerrato, A., R. Visconti, and A. Celetti, The rationale for druggability of CCDC6-tyrosine kinase fusions in lung cancer. Mol Cancer, 2018. 17(1): p. 46.
170. Wang, R., et al., RET fusions define a unique molecular and clinicopathologic subtype of non-small-cell lung cancer. J Clin Oncol, 2012. 30(35): p. 4352-9.
171. Chmielecki, J., et al., Comprehensive genomic profiling of pancreatic acinar cell carcinomas identifies recurrent RAF fusions and frequent inactivation of DNA repair genes. Cancer Discov, 2014. 4(12): p. 1398-405.
172. Jang, J.S., et al., Common Oncogene Mutations and Novel SND1-BRAF Transcript Fusion in Lung Adenocarcinoma from Never Smokers. Sci Rep, 2015. 5: p. 9755.
173. Malapelle, U., et al., The role of emerging biomarkers in the treatment of non-small cell lung cancer: current challenges and the way forward. Expert Opinion on Investigational Drugs.
174. He, Y., et al., Different Types of ROS1 Fusion Partners Yield Comparable Efficacy to Crizotinib. Oncol Res, 2019. 27(8): p. 901-910.
175. Zhang, X., et al., EGFR gene copy number as a predictive/biomarker for patients with non-small-cell lung cancer receiving tyrosine kinase inhibitor treatment: a systematic review and meta-analysis. J Investig Med, 2017. 65(1): p. 72-81.
176. Sakre, N., et al., RICTOR amplification identifies a subgroup in small cell lung cancer and predicts response to drugs targeting mTOR. Oncotarget, 2017. 8(4): p. 5992.
177. Suryavanshi, M., et al., MET Amplification and Response to MET Inhibitors in Stage IV Lung Adenocarcinoma. Oncol Res Treat, 2017. 40(4): p. 198-202.
178. Kawakami, H., et al., Targeting MET Amplification as a New Oncogenic Driver. Cancers (Basel), 2014. 6(3): p. 1540-52.
179. Lee, J., et al., Detection of ERBB2 (HER2) Gene Amplification Events in Cell-Free DNA and Response to Anti-HER2 Agents in a Large Asian Cancer Patient Cohort. Front Oncol, 2019. 9: p. 212.
180. Fuchs, E.M., et al., High-level ERBB2 gene amplification is associated with a particularly short time-to-metastasis, but results in a high rate of complete

response once trastuzumab-based therapy is offered in the metastatic setting. Int J Cancer, 2014. 135(1): p. 224-31.
181. Kaibori, M., et al., Increased FGF19 copy number is frequently detected in hepatocellular carcinoma with a complete response after sorafenib treatment. Oncotarget, 2016. 7(31): p. 49091-49098.
182. Arao, T., et al., FGF3/FGF4 amplification and multiple lung metastases in responders to sorafenib in hepatocellular carcinoma. Hepatology, 2013. 57(4): p. 1407-15.
183. Lievre, A., et al., Protein biomarkers predictive for response to anti-EGFR treatment in RAS wild-type metastatic colorectal carcinoma. Br J Cancer, 2017. 117(12): p. 1819-1827.
184. Harb, J., P.-J. Lin, and J. Hao, Recent Development of Wnt Signaling Pathway Inhibitors for Cancer Therapeutics. Current oncology reports, 2019. 21(2): p. 12.
185. Luga, V., et al., Exosomes mediate stromal mobilization of autocrine Wnt-PCP signaling in breast cancer cell migration. Cell, 2012. 151(7): p. 1542-56.
186. Kurayoshi, M., et al., Expression of Wnt-5a is correlated with aggressiveness of gastric cancer by stimulating cell migration and invasion. Cancer Res, 2006. 66(21): p. 10439-48.
187. Weeraratna, A.T., et al., Wnt5a signaling directly affects cell motility and invasion of metastatic melanoma. Cancer Cell, 2002. 1(3): p. 279-88.
188. Nussinov, R., et al., The Key Role of Calmodulin in KRAS-Driven Adenocarcinomas. Mol Cancer Res, 2015. 13(9): p. 1265-73.
189. Wen, Y., M.J. Alshikho, and M.R. Herbert, Pathway Network Analyses for Autism Reveal Multisystem Involvement, Major Overlaps with Other Diseases and Convergence upon MAPK and Calcium Signaling. PLoS One, 2016. 11(4): p. e0153329.
190. Joyal, J.L., et al., Calmodulin activates phosphatidylinositol 3-kinase. J Biol Chem, 1997. 272(45): p. 28183-6.
191. Chen, Y., et al., Identification of differently expressed genes in human colorectal adenocarcinoma. World J Gastroenterol, 2006. 12(7): p. 1025-32.
192. Liu, G.X., H.F. Sheng, and S. Wu, A study on the levels of calmodulin and DNA in human lung cancer cells. Br J Cancer, 1996. 73(7): p. 899-901.
193. Chen, X., et al., Manta: rapid detection of structural variants and indels for clinical sequencing applications. bioRxiv, 2015: p. 024232.

Supplementary Tables

Supplementary Table 1. The list of up-regulated genes in *KRAS* wild type of Australian cohort. The genes were ranked on logFC.

Row.numbers	Gene.names	logFC	logCPM	PValue	FDR
1	DLK1	7.97292951	5.55485919	2.54E-33	6.33E-30
2	MUC2	6.38638539	7.59526363	8.84E-20	6.60E-17
3	SERPINI2	6.31642271	3.31337232	4.52E-20	3.97E-17
4	GALNT8	6.2192759	2.2790112	2.28E-34	6.82E-31
5	REG1B	6.20470658	8.22342028	3.45E-14	1.10E-11
6	CPA2	6.1450185	7.73288531	1.02E-15	4.02E-13
7	RBPJL	6.08208775	1.41317857	1.21E-13	3.63E-11
8	LIMS4	5.92872396	1.88558997	1.91E-19	1.36E-16
9	CEL	5.88508195	8.82056839	1.11E-17	6.65E-15
10	BRINP3	5.54537297	0.82529942	1.18E-26	1.75E-23
11	REG3G	5.39000139	4.890646	3.15E-10	5.23E-08
12	CLPS	5.37913239	5.16724909	3.06E-12	7.48E-10
13	HEPACAM2	5.35818161	4.76238599	7.78E-20	6.12E-17
14	KIRREL2	5.20926128	3.0752551	6.61E-22	7.05E-19
15	TEX11	5.12611822	1.46271373	3.01E-41	2.25E-37
16	ERP27	5.11199955	4.37590605	1.36E-31	2.89E-28
17	REG1A	5.09972448	10.2152784	1.85E-17	1.06E-14
18	CDH10	5.09527885	1.3336024	1.69E-20	1.69E-17
19	B4GALNT2	5.0593667	1.19744194	1.14E-11	2.54E-09
20	SPINK4	5.03985458	6.26774831	2.77E-10	4.86E-08
21	KIF19	4.99311853	1.59510227	1.13E-38	5.61E-35
22	PRSS3	4.76419629	8.0626321	5.72E-28	9.48E-25
23	CELA3B	4.6165	5.80825847	4.03E-09	5.42E-07
24	KLHL32	4.57450776	0.80500488	1.77E-35	6.61E-32
25	DCC	4.56694963	0.75865872	6.59E-26	8.95E-23
26	SLC16A12	4.55423401	3.32010969	1.23E-24	1.54E-21
27	CPA1	4.54775704	8.07529079	7.60E-09	9.54E-07
28	TMEM132D	4.50063843	1.25181151	1.80E-15	6.88E-13
29	LRRC26	4.3161645	0.84353105	7.20E-12	1.66E-09
30	TGM3	4.29420669	0.06921096	1.99E-16	9.01E-14
31	GP2	4.20787258	7.63157596	1.16E-11	2.55E-09
32	DPP10	4.1812141	1.36408432	4.55E-12	1.08E-09
33	PTCHD1	4.0733331	2.18150318	1.68E-24	1.93E-21
34	EYA1	3.89757342	0.80107865	1.53E-09	2.27E-07
35	GALNTL6	3.86243943	0.09390789	2.56E-20	2.39E-17
36	BRSK2	3.8544767	1.3537116	5.96E-19	3.93E-16
37	SV2B	3.83689217	3.91175948	4.51E-30	8.42E-27
38	GABBR2	3.82905552	1.33266437	1.79E-10	3.35E-08

39	ERBB4	3.81322771	1.71264633	7.79E-20	6.12E-17
40	SLC38A3	3.77677832	2.01636857	1.92E-09	2.70E-07
41	ATOH1	3.75492827	1.06017987	3.62E-06	0.0002843
42	ITM2A	3.7299675	5.30644852	2.70E-16	1.18E-13
43	PTGDR2	3.70372404	0.3212642	9.15E-17	4.41E-14
44	SLC2A14	3.69431632	1.06713018	2.70E-11	5.45E-09
45	PKHD1L1	3.66061394	1.2914409	9.21E-16	3.72E-13
46	KCNK3	3.59457402	2.58764361	6.27E-18	3.90E-15
47	ITPR2	3.57901038	8.05599966	1.96E-43	2.93E-39
48	PHYHIPL	3.56675648	2.23527312	6.15E-16	2.55E-13
49	SLIT1	3.52411623	0.18683411	6.24E-09	8.03E-07
50	SLC38A11	3.51617703	3.4405296	2.93E-14	9.71E-12
51	PCSK1	3.42757262	4.09284272	2.27E-11	4.76E-09
52	PON1	3.42665527	1.05164323	8.06E-07	7.24E-05
53	SPSB4	3.3116291	0.31944038	7.38E-14	2.30E-11
54	DIRAS3	3.28940279	2.34825957	8.20E-17	4.08E-14
55	COCH	3.26788921	3.14992747	2.07E-13	6.05E-11
56	DLGAP1	3.16920342	1.34669328	2.00E-10	3.68E-08
57	CAMK2B	3.16522888	1.42203312	2.28E-10	4.11E-08
58	DNER	3.16112576	3.06577512	1.29E-07	1.33E-05
59	UGT2B7	3.160867	4.09905357	6.89E-09	8.79E-07
60	FOLH1	3.1558537	3.47849778	2.23E-12	5.56E-10
61	RGS7	3.07461297	1.53308496	4.68E-09	6.18E-07
62	PRKG2	3.03027926	2.65917192	7.03E-17	3.85E-14
63	ELOVL2	3.0208051	1.23787215	2.56E-14	8.68E-12
64	CACNA1B	3.0130181	0.23388267	5.57E-10	8.66E-08
65	CUZD1	2.96486507	4.40928668	4.36E-06	0.00032574
66	GFRA1	2.91983965	3.9328104	2.70E-11	5.45E-09
67	XPNPEP2	2.91950399	0.98828557	1.06E-06	9.21E-05
68	NXPE2	2.91066821	1.01885372	2.62E-09	3.62E-07
69	DIRAS1	2.89523086	0.70754879	1.23E-12	3.21E-10
70	SNORC	2.89517321	2.25093985	7.87E-15	2.94E-12
71	ATP6V1C2	2.88392802	1.64364361	9.76E-09	1.18E-06
72	IGFALS	2.86331667	0.71962124	2.47E-08	2.81E-06
73	HSPA2	2.86324487	4.9660532	6.06E-19	3.93E-16
74	KCTD16	2.81466178	1.44781887	3.35E-14	1.09E-11
75	ZNF385B	2.79649064	0.54435635	2.11E-13	6.06E-11
76	CCL24	2.78647891	3.18044271	4.45E-07	4.23E-05
77	SCG5	2.77746911	5.54356612	2.36E-10	4.19E-08
78	RHBDL3	2.74019039	0.00119327	4.07E-09	5.43E-07
79	NOTUM	2.71983146	1.31137348	7.77E-09	9.68E-07
80	IRX2	2.70826996	0.80269499	8.83E-07	7.85E-05

81	SNAP25	2.69556007	2.59283975	3.51E-10	5.70E-08
82	RSPH4A	2.67815912	-0.1681771	7.24E-12	1.66E-09
83	SERPINA5	2.66309312	5.87075388	2.39E-08	2.76E-06
84	ST18	2.66029401	2.67174052	1.14E-12	3.04E-10
85	COL11A2	2.62041172	1.46461765	3.19E-08	3.50E-06
86	HPN	2.61294772	3.82322392	5.16E-06	0.00037383
87	RHOBTB3	2.60801747	7.38845931	3.04E-16	1.30E-13
88	BCHE	2.59964085	2.06154586	8.38E-10	1.25E-07
89	FFAR2	2.59532088	0.99523113	3.12E-10	5.23E-08
90	NRG2	2.58020239	0.44828162	2.60E-09	3.62E-07
91	HBA1	2.5753316	2.72756479	9.35E-07	8.22E-05
92	CDKL4	2.54767938	-0.2529231	1.15E-13	3.51E-11
93	NR5A2	2.54403159	5.61390431	3.26E-10	5.35E-08
94	DPP6	2.52847542	2.08389374	9.50E-09	1.16E-06
95	INHA	2.52125908	-0.1771415	1.69E-09	2.43E-07
96	RET	2.52052968	1.48088847	8.56E-11	1.66E-08
97	CCDC110	2.51438454	1.07706589	8.07E-17	4.08E-14
98	NEURL1	2.51413802	2.75722692	3.97E-07	3.83E-05
99	GATM	2.51081466	7.32056991	1.07E-14	3.91E-12
100	BAALC	2.48749196	1.40117435	1.75E-12	4.42E-10
101	GRIP2	2.47994009	1.62723203	3.78E-10	6.07E-08
102	CRACR2A	2.45743625	4.47175135	2.13E-14	7.38E-12
103	AKR1C4	2.44683269	1.00561357	1.82E-06	0.00014977
104	FGF12	2.42295659	1.13243084	1.19E-11	2.58E-09
105	BIRC7	2.40315308	1.24103499	4.06E-06	0.00031085
106	ANO3	2.40071328	0.55958916	3.02E-07	2.99E-05
107	SYCP2L	2.39403825	-0.0990201	5.76E-11	1.15E-08
108	FAM222A	2.39019086	1.83984918	7.65E-11	1.50E-08
109	P2RX1	2.38324142	1.78601767	2.93E-13	8.27E-11
110	SLC35F3	2.34303501	1.90097497	1.29E-07	1.33E-05
111	SLC8A3	2.30949221	-0.2418095	7.37E-09	9.33E-07
112	REP15	2.30480773	2.25475881	7.86E-06	0.00055332
113	CCND2	2.29659286	7.47834509	4.52E-13	1.23E-10
114	FLRT2	2.28016039	5.65710416	8.34E-10	1.25E-07
115	SERPINA4	2.26720606	5.0218974	8.09E-07	7.24E-05
116	MAP2K6	2.24186801	4.08435261	1.99E-14	7.08E-12
117	MYO16	2.24061739	0.31471253	4.47E-13	1.23E-10
118	ATP6V1B1	2.21935187	0.59337879	4.30E-06	0.00032302
119	LYPD5	2.19224977	2.550204	1.32E-10	2.50E-08
120	VIPR2	2.15499963	0.83292427	1.09E-06	9.44E-05
121	C2CD4A	2.13439775	4.04996944	3.83E-06	0.00029652
122	SLC30A2	2.13130404	2.1734069	1.06E-06	9.21E-05

123	SPINK1	2.1183044	8.45617437	5.02E-06	0.00036584
124	TMED6	2.11572964	0.15127232	6.98E-06	0.00049664
125	PHEX	2.09015357	1.47307682	8.34E-10	1.25E-07
126	FAM169A	2.08351714	3.46850959	2.40E-08	2.76E-06
127	EVA1A	2.08021042	3.47937426	2.97E-10	5.17E-08
128	AOX1	2.05078258	4.33052467	4.10E-06	0.00031256
129	NCAM1	2.05062941	2.9608763	5.20E-08	5.59E-06
130	CERS4	2.04661759	2.89904459	3.08E-10	5.22E-08
131	ACSL6	2.03367214	3.58891376	7.09E-08	7.57E-06
132	XBP1	2.02697864	8.14953107	1.55E-16	7.23E-14
133	GRB14	2.02187411	1.94798894	5.05E-08	5.47E-06
134	PDE1C	2.00657681	2.25407094	8.86E-11	1.70E-08
135	LGI2	2.00106011	2.72426679	2.60E-08	2.92E-06
136	GNG7	1.9844628	2.19817752	4.43E-12	1.07E-09
137	BNIP3	1.97410814	3.87295443	8.24E-09	1.02E-06
138	ADAMTS6	1.96836725	4.16904036	6.04E-09	7.84E-07
139	GAMT	1.96425685	2.32158455	2.54E-08	2.88E-06
140	CADPS	1.95744252	5.06745814	4.84E-06	0.00035614
141	GDF15	1.9571581	5.64015132	1.00E-05	0.00068187
142	DPEP1	1.95254737	4.14116348	1.12E-05	0.00076014
143	KLHL41	1.94231645	0.42260343	1.94E-06	0.00015865
144	PPP1R3C	1.90292116	4.21115684	3.14E-08	3.48E-06
145	PROX1	1.90131443	4.84962613	2.22E-06	0.00018018
146	MYRIP	1.89534513	1.46682858	1.66E-06	0.00013747
147	EDA	1.88901631	2.24375198	2.73E-08	3.05E-06
148	TOX	1.88533743	3.16030071	2.15E-07	2.20E-05
149	SV2A	1.87358599	3.06682886	2.26E-11	4.76E-09
150	TSPAN7	1.84622828	2.96443376	7.44E-07	6.81E-05
151	SPAG17	1.83970372	2.58679965	4.62E-07	4.34E-05
152	DNAH14	1.83252175	4.67807142	1.61E-09	2.34E-07
153	SNCA	1.80170078	2.05805829	7.87E-08	8.33E-06
154	XRCC4	1.79889093	4.87523766	4.50E-07	4.25E-05
155	CXCL12	1.78788481	5.17210823	1.11E-06	9.50E-05
156	AIFM3	1.77594602	1.37027767	3.77E-07	3.65E-05
157	RBPJ	1.77100517	7.10602578	7.22E-17	3.85E-14
158	HEY1	1.75177057	3.24397101	2.88E-09	3.95E-07
159	RAPGEF4	1.74979682	3.70642432	1.15E-08	1.39E-06
160	SEC11C	1.69347992	4.85619516	2.58E-11	5.35E-09
161	FILIP1	1.68950466	3.53472681	5.98E-06	0.00042923
162	SLC43A1	1.6882698	4.28146018	1.13E-06	9.56E-05
163	STK31	1.67630298	2.73099259	3.61E-06	0.0002843
164	MEST	1.67618944	6.44238188	5.18E-09	6.78E-07

165	CIB2	1.66642573	2.24919015	7.49E-07	6.82E-05
166	EML6	1.65926652	2.73093064	4.27E-06	0.00032208
167	SDK1	1.64451103	4.16687953	9.01E-07	7.96E-05
168	PFKFB4	1.63498702	3.52103601	3.08E-10	5.22E-08
169	BATF2	1.61369295	3.48249238	5.20E-06	0.00037536
170	CYFIP2	1.60109414	4.89636633	4.18E-08	4.56E-06
171	GDPD5	1.58800469	3.8319123	2.36E-08	2.75E-06
172	UBE2QL1	1.56334129	1.2556373	8.59E-06	0.00059831
173	TVP23C-CDRT4	1.55614004	3.1953811	1.01E-11	2.29E-09

Supplementary Table 2. The list of down regulated genes in Australian KRAS wild type cohort. The genes were ranked on Logfold changes.

Row.numbers	Gene.names	logFC	logCPM	PValue	FDR
1	SPRR2A	-10.396969	2.49594148	1.00E-04	0.0050101
2	SPRR1B	-6.1058777	2.95487297	0.00010827	0.00531857
3	DHRS9	-3.9219068	6.38931653	7.49E-05	0.00392588
4	ZBED2	-3.6028495	2.58097498	0.00017518	0.00795148
5	GAL	-3.5491723	0.87894352	0.00018251	0.0082093
6	COL17A1	-3.2512671	9.24043852	5.30E-05	0.00298495
7	HAS3	-2.8504779	4.06883233	2.61E-05	0.00160885
8	ADTRP	-2.6608782	3.65100098	6.33E-05	0.00343725
9	SP6	-2.5280332	2.84224382	1.36E-05	0.00090857
10	GSC	-2.1124313	0.25827155	0.00021378	0.0093616
11	GJB2	-1.945892	6.76443367	7.02E-05	0.00371997
12	LRRC8A	-1.503207	7.49498876	1.00E-05	0.00068187

Supplementary Table 3. The list of up-regulated genes in *KRAS* wild type of Canadian cohort. The genes were ranked on logFC.

Row.numbers	Gene.names	logFC	logCPM	PValue	FDR
1	PNLIP	5.25465079	5.44712968	7.71E-14	2.85E-11
2	AMY2A	4.94564142	3.81906824	1.43E-15	7.24E-13
3	NOTUM	4.91197721	2.60326873	9.06E-43	6.19E-39
4	CTRC	4.56418438	4.15251003	5.40E-17	3.51E-14
5	CELA2B	4.43814517	2.41608513	1.66E-55	2.28E-51
6	CEL	4.33401806	6.75909965	5.74E-20	4.61E-17
7	CELA3B	4.32984368	4.01170009	3.73E-13	1.31E-10
8	CPA1	4.21090219	6.18623523	3.76E-10	8.57E-08
9	CPA2	4.13906112	4.86129338	1.38E-08	2.00E-06
10	CPB1	4.00850047	6.20315205	3.67E-11	9.64E-09
11	SEMA3D	3.91616825	2.16583881	1.17E-27	2.67E-24
12	ITIH2	3.71907259	4.39860292	6.95E-25	9.50E-22
13	PHYHIPL	3.63958019	1.27614919	3.01E-27	5.87E-24
14	SALL1	3.57398765	2.20535015	1.19E-29	4.08E-26
15	REG1B	3.49097851	5.1584209	6.43E-07	6.32E-05
16	CTRB2	3.47293554	5.14841404	1.11E-06	9.84E-05
17	CRLF1	3.43480283	2.81107997	9.47E-18	6.47E-15
18	CTRB1	3.39497613	4.53682805	7.60E-07	7.26E-05
19	SV2B	3.35911418	1.78915606	1.58E-14	6.35E-12
20	SIX2	3.24175787	2.04071266	1.08E-14	4.49E-12
21	CELA3A	3.2306222	5.25313628	1.44E-06	0.00012214
22	SLC16A6	3.16799805	2.27274463	2.29E-28	6.27E-25
23	LAMA1	3.16493146	3.78448092	7.77E-37	3.54E-33
24	RIMS1	3.15463941	1.3486124	1.10E-11	3.06E-09
25	MBOAT4	3.08898939	1.64523644	4.34E-25	6.59E-22
26	PDE1C	3.05543289	2.27377182	4.91E-19	3.73E-16
27	SIM1	3.03239869	1.95577698	8.67E-09	1.38E-06
28	IGFBP1	3.01730931	3.8371931	1.58E-10	3.78E-08
29	ALK	3.01684918	1.69533431	1.78E-26	3.05E-23
30	CUZD1	2.99850295	3.97119161	6.51E-14	2.47E-11
31	GLDC	2.99430229	1.91216646	4.01E-13	1.37E-10
32	IL33	2.87719169	3.13823473	1.02E-11	2.91E-09
33	MYO16	2.86116779	2.10503675	5.06E-23	5.32E-20
34	ERP27	2.85345492	2.38563854	2.32E-16	1.44E-13
35	DPEP1	2.83432582	3.00779694	6.92E-22	6.31E-19
36	DPP10	2.79400057	1.09644971	3.94E-06	0.00028572
37	FBN3	2.76800618	0.84286068	2.55E-06	0.00019608
38	OTC	2.72121855	1.77874847	8.13E-06	0.00049174

39	CDX2	2.61620057	2.95526737	7.17E-13	2.33E-10
40	GRID1	2.6051769	3.01273152	1.07E-24	1.22E-21
41	ELN	2.59542126	6.2990268	8.09E-25	1.00E-21
42	AGT	2.57392026	5.44963946	2.08E-13	7.49E-11
43	ZNF521	2.5637066	4.30166687	2.18E-21	1.86E-18
44	CLDN3	2.56363035	4.00549198	2.60E-14	1.01E-11
45	GALNT16	2.52556097	1.86317253	3.68E-15	1.57E-12
46	NRXN3	2.51804792	2.55765041	3.38E-16	2.01E-13
47	EPHA7	2.51533964	2.89838993	2.54E-07	2.82E-05
48	AKR1C4	2.46747434	1.98619393	5.01E-12	1.46E-09
49	COL9A3	2.46545587	4.50992033	1.45E-11	3.96E-09
50	DLGAP1	2.46399396	1.82629998	1.85E-07	2.12E-05
51	NELL2	2.45827535	2.92567541	3.91E-10	8.77E-08
52	VTN	2.41137071	3.41515842	4.51E-08	6.05E-06
53	ITIH5	2.38179358	4.59777045	2.44E-12	7.58E-10
54	FSTL4	2.36980579	1.5481735	5.41E-09	9.13E-07
55	KLK1	2.34980623	3.4949693	2.08E-09	3.94E-07
56	UNC93A	2.34824018	0.90590439	4.60E-08	6.10E-06
57	ST6GALNAC5	2.31241853	1.10263985	4.21E-13	1.40E-10
58	CACNA1B	2.2621993	1.68709542	4.60E-09	8.17E-07
59	CRTAC1	2.22363385	1.28899638	2.42E-07	2.73E-05
60	CTNND2	2.21554497	3.07279114	1.28E-08	1.89E-06
61	SLC38A3	2.21151686	1.35430161	1.44E-05	0.00078253
62	MSI1	2.19050874	1.48508313	5.90E-11	1.49E-08
63	SLC34A2	2.18521691	6.13453426	1.92E-09	3.69E-07
64	REG1A	2.16749388	7.66497282	4.59E-06	0.00032001
65	KIF1A	2.12043864	4.49173039	1.57E-06	0.00013152
66	DKK4	2.11330878	1.42788899	1.10E-06	9.84E-05
67	SLC16A12	2.09836608	1.57099761	1.84E-05	0.00094035
68	VSNL1	2.09318919	1.7070402	3.27E-09	6.12E-07
69	GAL3ST2	2.08480805	1.04111555	8.56E-07	8.02E-05
70	PROX1	2.07367951	4.27526511	7.55E-10	1.59E-07
71	KIRREL2	2.04925934	2.55389063	8.74E-06	0.00051974
72	YBX2	2.03987849	1.46576247	3.04E-06	0.00022865
73	CCND2	2.03031881	4.47846877	1.78E-09	3.52E-07
74	GGH	2.02620916	3.58363907	3.55E-16	2.02E-13
75	SP5	2.01006677	1.89613314	5.47E-10	1.17E-07
76	EPHA6	2.0100619	0.86388307	1.94E-06	0.00015681
77	UNC5A	1.99040278	2.11449047	2.12E-06	0.00016886
78	ERBB4	1.97670488	1.25036114	2.07E-07	2.36E-05
79	SLC30A2	1.9736862	1.70100628	4.76E-06	0.00032697
80	RELN	1.96451684	4.64705671	1.27E-05	0.00070525

81	IRF2BP1	1.96328788	4.71975691	1.31E-22	1.28E-19
82	TMPRSS6	1.95644076	2.56896184	1.01E-08	1.55E-06
83	TACR1	1.93477368	1.0671359	1.81E-15	8.83E-13
84	P2RX1	1.91142021	1.56460835	3.72E-09	6.87E-07
85	FBN2	1.90416551	3.38036239	4.12E-09	7.41E-07
86	BCL11A	1.89193129	2.20175655	7.54E-08	9.45E-06
87	DRP2	1.89053959	1.56411162	2.75E-08	3.76E-06
88	ABCC2	1.87462331	4.16618364	2.87E-07	3.11E-05
89	CERS4	1.87382181	2.28884696	3.61E-12	1.10E-09
90	GRB14	1.87124952	3.30144685	4.74E-12	1.41E-09
91	DCDC2	1.8712402	4.36789109	2.62E-11	7.02E-09
92	TMEM150C	1.866797	1.12911193	1.98E-10	4.67E-08
93	HAAO	1.86076555	2.89170818	5.40E-16	2.84E-13
94	FREM1	1.8445209	3.26861227	4.98E-09	8.61E-07
95	TRPM8	1.83848489	1.47619546	4.30E-06	0.00030801
96	KCNQ3	1.83816694	3.48244726	5.81E-08	7.50E-06
97	GRIP2	1.83252039	2.42116046	6.16E-07	6.11E-05
98	CYP2C8	1.82885054	2.32964408	1.08E-08	1.62E-06
99	LAMP5	1.82398585	3.12347913	6.73E-09	1.11E-06
100	DTNA	1.80919606	3.4129013	1.84E-09	3.60E-07
101	ALDH1B1	1.80827288	4.31118391	4.53E-18	3.26E-15
102	HES6	1.80601708	2.51136744	1.04E-08	1.58E-06
103	RUBCNL	1.80010606	2.24444573	7.55E-11	1.88E-08
104	ERVMER34-1	1.77303537	1.9094895	1.69E-09	3.40E-07
105	INHBB	1.74723304	2.32336642	1.73E-06	0.0001439
106	NUAK2	1.74712983	2.98695139	2.39E-15	1.13E-12
107	EDNRB	1.74643354	2.5574163	3.32E-07	3.49E-05
108	FGF14	1.73406463	1.40053666	2.66E-07	2.93E-05
109	DEFB1	1.73188928	4.40188998	1.39E-07	1.65E-05
110	IGFN1	1.72828279	2.5684379	4.93E-06	0.00033207
111	SNAP25	1.71584603	2.50255529	2.94E-06	0.00022196
112	NR5A2	1.68802677	4.80915542	2.39E-06	0.00018556
113	AKR1C1	1.68717647	5.58024907	1.40E-07	1.65E-05
114	GAMT	1.68086684	2.343805	6.14E-08	7.84E-06
115	MEX3A	1.67829881	1.57444316	1.34E-09	2.77E-07
116	SLC7A2	1.66828148	4.54140758	4.33E-06	0.00030841
117	CLDN15	1.66739795	3.90593594	6.24E-09	1.04E-06
118	CBX2	1.66275413	1.92778232	2.50E-08	3.47E-06
119	CDKN1C	1.65139191	3.73801204	1.62E-09	3.31E-07
120	NKD1	1.64810737	3.56527965	1.31E-10	3.19E-08
121	SCML2	1.64705362	0.6545814	7.17E-06	0.00045191
122	CCDC110	1.64351517	0.85453321	4.95E-09	8.61E-07

123	LGI2	1.63721455	0.92194442	1.13E-05	0.00064155
124	MYRIP	1.63661926	1.92423251	3.81E-09	6.95E-07
125	KIT	1.62770746	3.98715943	2.51E-08	3.47E-06
126	NTRK2	1.61706649	3.25223252	6.80E-06	0.00043246
127	ITPR2	1.60933487	7.57996673	3.36E-15	1.48E-12
128	ZIK1	1.5977576	1.45971788	2.35E-06	0.00018324
129	LARGE2	1.58559332	3.92752796	5.56E-07	5.67E-05
130	LRRN4	1.5837274	1.34306249	9.96E-07	9.07E-05
131	RGN	1.57879838	1.43375005	1.56E-06	0.00013152
132	CALCRL	1.57060797	3.45596718	3.07E-07	3.28E-05
133	CCNE1	1.55133392	3.26181941	1.15E-08	1.71E-06
134	AIF1L	1.54974625	2.03237698	6.86E-07	6.65E-05
135	ABCB1	1.54294153	4.7317752	5.05E-06	0.00033845
136	HUNK	1.53987141	2.42486271	8.82E-09	1.39E-06
137	GPC2	1.52774569	0.89275685	5.49E-08	7.21E-06
138	WNK2	1.52196005	6.21937037	5.81E-06	0.00037634
139	SLIT3	1.51451397	4.20732236	2.12E-06	0.00016886
140	BRSK2	1.51091806	2.39968165	4.38E-07	4.53E-05

Supplementary Table 4. The list of down-regulated genes in KRAS wild type of Canadian cohort. The genes were ranked on logFC.

Gene.names	Row.names	logFC	logCPM	PValue	FDR
1	KRT14	-4.6895851	5.359329	0.00012063	0.0040122
2	KRT13	-3.7449794	5.58301662	4.75E-05	0.00198614
3	CACNG6	-3.5295136	0.46696682	8.94E-05	0.00318402
4	NMUR2	-2.8073291	3.74497093	0.0001908	0.00554951
5	PLCZ1	-2.7081174	2.08931457	6.24E-05	0.00240354
6	KRT79	-2.5041279	1.03184228	0.00015699	0.00485519
7	WNT7A	-2.4913995	3.5087808	9.92E-05	0.00345777
8	CSF2	-2.4399484	0.96481129	8.97E-06	0.00053095
9	CYP27C1	-2.2718315	2.04491206	8.55E-06	0.00051274
10	DHRS9	-2.1684764	5.91745925	0.00038645	0.00921956
11	ADTRP	-1.7851427	3.60925077	0.00021521	0.00604077
12	NUTM1	-1.739303	1.59913335	8.19E-05	0.00300046
13	FGF2	-1.7153308	2.94951324	0.00034284	0.00847483
14	PLAT	-1.7100554	8.71532379	0.0002499	0.00670387
15	UGT1A10	-1.7005134	5.74825512	4.93E-05	0.00202387
16	TRNP1	-1.6908537	1.71239975	0.00019245	0.00557035
17	MYEOV	-1.6736654	5.61987319	1.22E-06	0.00010589
18	NMU	-1.544519	3.47973544	7.25E-05	0.00268737
19	AMIGO2	-1.5280546	5.80236203	6.48E-05	0.00247208
20	EPHA4	-1.5137668	6.17966775	8.39E-06	0.00050525

Appendices

Appendix-1

Sample IDs for *KRAS* wild type cohort.

Donor ID	APGI_ID	Icgc_Donor_id	Project_Code
ICGC_0053	APGI_2158	DO32976	PACA-AU
ICGC_0063	APGI_2219	DO33016	PACA-AU
ICGC_0075	APGI_1971	DO33077	PACA-AU
ICGC_0095	APGI_2171	DO33272	PACA-AU
ICGC_0112	APGI_2267	DO33240	PACA-AU
ICGC_0138	APGI_2590	DO33424	PACA-AU
ICGC_0150	APGI_2715	DO33488	PACA-AU
ICGC_0167	APGI_2779	DO34184	PACA-AU
ICGC_0187	APGI_2824	DO34256	PACA-AU
ICGC_0203	APGI_2935	DO34344	PACA-AU
ICGC_0209	APGI_2274	DO33640	PACA-AU
ICGC_0214	APGI_2938	DO34432	PACA-AU
ICGC_0229	APGI_2941	DO34496	PACA-AU
ICGC_0230	APGI_2973	DO34504	PACA-AU
ICGC_0241	APGI_2629	DO33680	PACA-AU
ICGC_0257	APGI_2657	DO33808	PACA-AU
ICGC_0259	APGI_2659	DO33824	PACA-AU
ICGC_0286	APGI_2708	DO34096	PACA-AU
ICGC_0308	APGI_3096	DO34752	PACA-AU
ICGC_0338	APGI_3060	DO34961	PACA-AU
ICGC_0347	APGI_3090	DO35017	PACA-AU
ICGC_0352	APGI_3299	DO35057	PACA-AU
ICGC_0391	APGI_2153	DO49076	PACA-AU
ICGC_0395	APGI_2991	DO49087	PACA-AU
ICGC_0411	APGI_3371	DO49124	PACA-AU
ICGC_0414	APGI_3510	DO49132	PACA-AU
ICGC_0486	APGI_2997	DO49090	PACA-AU
ICGC_0529	APGI_3796	DO49201	PACA-AU
ICGC_0094	APGI_2165	DO33568	PACA-AU
ICGC_0399	APGI_3066	DO49097	PACA-AU
ICGC_0180	APGI_2799	DO34216	PACA-AU
ICGC_0548	APGI_3205	NOT IN DCC	PACA-AU
GARV_0664	APGI_3327	NOT IN DCC	PACA-AU
GARV_0674	APGI_3333	NOT IN DCC	PACA-AU
PCSI_0339		DO224633	PACA-CA
PCSI_0591		DO224719	PACA-CA
PCSI_0626		DO224782	PACA-CA

PCSI_0627		DO227661	PACA-CA
PCSI_0024		DO35138	PACA-CA
PCSI_0090		DO35242	PACA-CA
PCSI_0120		DO35266	PACA-CA
PCSI_0108		DO35330	PACA-CA
PCSI_0305		DO49430	PACA-CA
PCSI_0347		DO49442	PACA-CA
PCSI_0015		DO51476	PACA-CA
PCSI_0328		DO51479	PACA-CA
PCSI_0145		DO51489	PACA-CA
PCSI_0458		DO51491	PACA-CA
PCSI_0326		DO51503	PACA-CA
PCSI_0248		DO51508	PACA-CA
PCSI_0330		DO51517	PACA-CA
PCSI_0286		DO51521	PACA-CA
PCSI_0608		DO224750	PACA-CA
PCSI_0164		DO49418	PACA-CA
PCSI_0572		DO51487	PACA-CA
PCSI_0506		DO51502	PACA-CA
TCGA-3A-A9J0		DO218962	PAAD-US
TCGA-3A-A9I5		DO218951	PAAD-US
TCGA-LB-A7SX		DO49409	PAAD-US
TCGA-XD-AAUH		DO51721	PAAD-US
TCGA-IB-AAUT		DO51710	PAAD-US
TCGA-US-A77E		DO49416	PAAD-US
TCGA-2J-AABA		DO51659	PAAD-US
TCGA-US-A776		DO50270	PAAD-US
TCGA-LB-A8F3		DO50262	PAAD-US
TCGA-IB-A5SQ		DO46653	PAAD-US
TCGA-IB-7888		DO46641	PAAD-US
TCGA-IB-7891		DO46623	PAAD-US
TCGA-F2-A44H		DO46619	PAAD-US
TCGA-HZ-8317		DO32802	PAAD-US

Appendix-2

Sample IDs for *KRAS* mutant cohort.

Donor ID	Project Code	KRAS Status
DO32829	PACA-AU	MUTANT
DO32875	PACA-AU	MUTANT
DO32878	PACA-AU	MUTANT
DO32887	PACA-AU	MUTANT
DO32900	PACA-AU	MUTANT
DO32936	PACA-AU	MUTANT
DO33091	PACA-AU	MUTANT
DO33128	PACA-AU	MUTANT
DO33168	PACA-AU	MUTANT
DO33256	PACA-AU	MUTANT
DO33336	PACA-AU	MUTANT
DO33344	PACA-AU	MUTANT
DO33368	PACA-AU	MUTANT
DO33376	PACA-AU	MUTANT
DO33392	PACA-AU	MUTANT
DO33400	PACA-AU	MUTANT
DO33408	PACA-AU	MUTANT
DO33472	PACA-AU	MUTANT
DO33480	PACA-AU	MUTANT
DO33512	PACA-AU	MUTANT
DO33528	PACA-AU	MUTANT
DO33544	PACA-AU	MUTANT
DO33552	PACA-AU	MUTANT
DO33600	PACA-AU	MUTANT
DO33632	PACA-AU	MUTANT
DO33656	PACA-AU	MUTANT
DO33984	PACA-AU	MUTANT
DO34240	PACA-AU	MUTANT
DO34264	PACA-AU	MUTANT
DO34288	PACA-AU	MUTANT
DO34312	PACA-AU	MUTANT
DO34336	PACA-AU	MUTANT
DO34368	PACA-AU	MUTANT
DO34376	PACA-AU	MUTANT
DO34448	PACA-AU	MUTANT
DO34600	PACA-AU	MUTANT
DO34608	PACA-AU	MUTANT

DO34616	PACA-AU	MUTANT
DO34640	PACA-AU	MUTANT
DO34656	PACA-AU	MUTANT
DO34680	PACA-AU	MUTANT
DO34696	PACA-AU	MUTANT
DO34720	PACA-AU	MUTANT
DO34728	PACA-AU	MUTANT
DO34736	PACA-AU	MUTANT
DO34785	PACA-AU	MUTANT
DO34793	PACA-AU	MUTANT
DO34801	PACA-AU	MUTANT
DO34809	PACA-AU	MUTANT
DO34817	PACA-AU	MUTANT
DO34849	PACA-AU	MUTANT
DO34905	PACA-AU	MUTANT
DO34945	PACA-AU	MUTANT
DO49078	PACA-AU	MUTANT
DO49079	PACA-AU	MUTANT
DO49080	PACA-AU	MUTANT
DO49105	PACA-AU	MUTANT
DO49113	PACA-AU	MUTANT
DO49127	PACA-AU	MUTANT
DO49129	PACA-AU	MUTANT
DO49130	PACA-AU	MUTANT
DO49133	PACA-AU	MUTANT
DO49135	PACA-AU	MUTANT
DO49137	PACA-AU	MUTANT
DO49138	PACA-AU	MUTANT
DO49166	PACA-AU	MUTANT
DO49168	PACA-AU	MUTANT
DO49172	PACA-AU	MUTANT
DO49175	PACA-AU	MUTANT
DO49181	PACA-AU	MUTANT
DO49183	PACA-AU	MUTANT
DO49184	PACA-AU	MUTANT
DO49185	PACA-AU	MUTANT
DO49193	PACA-AU	MUTANT
DO49198	PACA-AU	MUTANT
DO49199	PACA-AU	MUTANT
DO49204	PACA-AU	MUTANT
DO221539	PACA-CA	MUTANT
DO221540	PACA-CA	MUTANT

DO221541	PACA-CA	MUTANT
DO221542	PACA-CA	MUTANT
DO221543	PACA-CA	MUTANT
DO221545	PACA-CA	MUTANT
DO221546	PACA-CA	MUTANT
DO224575	PACA-CA	MUTANT
DO224642	PACA-CA	MUTANT
DO224648	PACA-CA	MUTANT
DO224656	PACA-CA	MUTANT
DO224688	PACA-CA	MUTANT
DO224698	PACA-CA	MUTANT
DO224705	PACA-CA	MUTANT
DO224712	PACA-CA	MUTANT
DO224714	PACA-CA	MUTANT
DO224724	PACA-CA	MUTANT
DO224734	PACA-CA	MUTANT
DO224740	PACA-CA	MUTANT
DO224745	PACA-CA	MUTANT
DO224758	PACA-CA	MUTANT
DO224764	PACA-CA	MUTANT
DO224767	PACA-CA	MUTANT
DO224770	PACA-CA	MUTANT
DO224776	PACA-CA	MUTANT
DO224779	PACA-CA	MUTANT
DO224784	PACA-CA	MUTANT
DO227544	PACA-CA	MUTANT
DO227558	PACA-CA	MUTANT
DO227564	PACA-CA	MUTANT
DO227570	PACA-CA	MUTANT
DO227581	PACA-CA	MUTANT
DO227596	PACA-CA	MUTANT
DO227604	PACA-CA	MUTANT
DO227625	PACA-CA	MUTANT
DO227633	PACA-CA	MUTANT
DO227636	PACA-CA	MUTANT
DO227648	PACA-CA	MUTANT
DO227652	PACA-CA	MUTANT
DO227671	PACA-CA	MUTANT
DO227684	PACA-CA	MUTANT
DO227687	PACA-CA	MUTANT
DO227695	PACA-CA	MUTANT
DO227704	PACA-CA	MUTANT

DO227721	PACA-CA	MUTANT
DO227727	PACA-CA	MUTANT
DO227736	PACA-CA	MUTANT
DO227742	PACA-CA	MUTANT
DO230463	PACA-CA	MUTANT
DO230464	PACA-CA	MUTANT
DO231247	PACA-CA	MUTANT
DO231251	PACA-CA	MUTANT
DO231256	PACA-CA	MUTANT
DO231264	PACA-CA	MUTANT
DO231278	PACA-CA	MUTANT
DO231284	PACA-CA	MUTANT
DO35082	PACA-CA	MUTANT
DO35083	PACA-CA	MUTANT
DO35085	PACA-CA	MUTANT
DO35098	PACA-CA	MUTANT
DO35116	PACA-CA	MUTANT
DO35126	PACA-CA	MUTANT
DO35128	PACA-CA	MUTANT
DO35132	PACA-CA	MUTANT
DO35140	PACA-CA	MUTANT
DO35144	PACA-CA	MUTANT
DO35148	PACA-CA	MUTANT
DO35152	PACA-CA	MUTANT
DO35184	PACA-CA	MUTANT
DO35198	PACA-CA	MUTANT
DO35210	PACA-CA	MUTANT
DO35226	PACA-CA	MUTANT
DO35228	PACA-CA	MUTANT
DO35230	PACA-CA	MUTANT
DO35236	PACA-CA	MUTANT
DO35258	PACA-CA	MUTANT
DO35290	PACA-CA	MUTANT
DO35305	PACA-CA	MUTANT
DO35350	PACA-CA	MUTANT
DO35360	PACA-CA	MUTANT
DO35365	PACA-CA	MUTANT
DO35406	PACA-CA	MUTANT
DO35424	PACA-CA	MUTANT
DO35454	PACA-CA	MUTANT
DO35496	PACA-CA	MUTANT
DO49419	PACA-CA	MUTANT

DO49420	PACA-CA	MUTANT
DO49421	PACA-CA	MUTANT
DO49424	PACA-CA	MUTANT
DO49427	PACA-CA	MUTANT
DO49433	PACA-CA	MUTANT
DO49436	PACA-CA	MUTANT
DO49439	PACA-CA	MUTANT
DO49448	PACA-CA	MUTANT
DO49451	PACA-CA	MUTANT
DO49454	PACA-CA	MUTANT
DO49457	PACA-CA	MUTANT
DO49460	PACA-CA	MUTANT
DO49463	PACA-CA	MUTANT
DO49469	PACA-CA	MUTANT
DO49472	PACA-CA	MUTANT
DO49475	PACA-CA	MUTANT
DO49481	PACA-CA	MUTANT
DO51464	PACA-CA	MUTANT
DO51465	PACA-CA	MUTANT
DO51466	PACA-CA	MUTANT
DO51467	PACA-CA	MUTANT
DO51468	PACA-CA	MUTANT
DO51469	PACA-CA	MUTANT
DO51470	PACA-CA	MUTANT
DO51472	PACA-CA	MUTANT
DO51473	PACA-CA	MUTANT
DO51474	PACA-CA	MUTANT
DO51475	PACA-CA	MUTANT
DO51478	PACA-CA	MUTANT
DO51480	PACA-CA	MUTANT
DO51481	PACA-CA	MUTANT
DO51482	PACA-CA	MUTANT
DO51483	PACA-CA	MUTANT
DO51484	PACA-CA	MUTANT
DO51485	PACA-CA	MUTANT
DO51490	PACA-CA	MUTANT
DO51492	PACA-CA	MUTANT
DO51493	PACA-CA	MUTANT
DO51494	PACA-CA	MUTANT
DO51495	PACA-CA	MUTANT
DO51496	PACA-CA	MUTANT
DO51497	PACA-CA	MUTANT

DO51498	PACA-CA	MUTANT
DO51500	PACA-CA	MUTANT
DO51501	PACA-CA	MUTANT
DO51504	PACA-CA	MUTANT
DO51506	PACA-CA	MUTANT
DO51509	PACA-CA	MUTANT
DO51510	PACA-CA	MUTANT
DO51511	PACA-CA	MUTANT
DO51512	PACA-CA	MUTANT
DO51513	PACA-CA	MUTANT
DO51514	PACA-CA	MUTANT
DO51515	PACA-CA	MUTANT
DO51518	PACA-CA	MUTANT
DO51519	PACA-CA	MUTANT
DO51520	PACA-CA	MUTANT
DO51522	PACA-CA	MUTANT
DO51523	PACA-CA	MUTANT
DO51524	PACA-CA	MUTANT
DO51525	PACA-CA	MUTANT
DO51526	PACA-CA	MUTANT
DO51527	PACA-CA	MUTANT
DO51528	PACA-CA	MUTANT
DO51529	PACA-CA	MUTANT
DO51530	PACA-CA	MUTANT
DO51531	PACA-CA	MUTANT
DO51532	PACA-CA	MUTANT
DO51533	PACA-CA	MUTANT
DO51534	PACA-CA	MUTANT
DO51535	PACA-CA	MUTANT
DO51536	PACA-CA	MUTANT
DO51537	PACA-CA	MUTANT
DO51538	PACA-CA	MUTANT
DO51540	PACA-CA	MUTANT
DO51541	PACA-CA	MUTANT
DO51542	PACA-CA	MUTANT
DO51543	PACA-CA	MUTANT
DO51545	PACA-CA	MUTANT
DO51548	PACA-CA	MUTANT
DO51549	PACA-CA	MUTANT
DO32817	PACA-AU	MUTANT
DO32819	PACA-AU	MUTANT
DO32821	PACA-AU	MUTANT

DO32825	PACA-AU	MUTANT
DO32831	PACA-AU	MUTANT
DO32833	PACA-AU	MUTANT
DO32835	PACA-AU	MUTANT
DO32837	PACA-AU	MUTANT
DO32851	PACA-AU	MUTANT
DO32893	PACA-AU	MUTANT
DO32896	PACA-AU	MUTANT
DO32904	PACA-AU	MUTANT
DO32908	PACA-AU	MUTANT
DO32912	PACA-AU	MUTANT
DO32916	PACA-AU	MUTANT
DO32928	PACA-AU	MUTANT
DO32960	PACA-AU	MUTANT
DO32968	PACA-AU	MUTANT
DO32972	PACA-AU	MUTANT
DO32980	PACA-AU	MUTANT
DO32984	PACA-AU	MUTANT
DO33000	PACA-AU	MUTANT
DO33008	PACA-AU	MUTANT
DO33012	PACA-AU	MUTANT
DO33028	PACA-AU	MUTANT
DO33037	PACA-AU	MUTANT
DO33042	PACA-AU	MUTANT
DO33049	PACA-AU	MUTANT
DO33056	PACA-AU	MUTANT
DO33063	PACA-AU	MUTANT
DO33152	PACA-AU	MUTANT
DO33160	PACA-AU	MUTANT
DO33184	PACA-AU	MUTANT
DO33200	PACA-AU	MUTANT
DO33208	PACA-AU	MUTANT
DO33248	PACA-AU	MUTANT
DO33264	PACA-AU	MUTANT
DO33288	PACA-AU	MUTANT
DO33312	PACA-AU	MUTANT
DO33432	PACA-AU	MUTANT
DO33496	PACA-AU	MUTANT
DO33560	PACA-AU	MUTANT
DO33664	PACA-AU	MUTANT
DO33688	PACA-AU	MUTANT
DO33792	PACA-AU	MUTANT

DO33960	PACA-AU	MUTANT
DO34072	PACA-AU	MUTANT
DO34088	PACA-AU	MUTANT
DO34120	PACA-AU	MUTANT
DO34128	PACA-AU	MUTANT
DO34584	PACA-AU	MUTANT
DO34921	PACA-AU	MUTANT
DO35009	PACA-AU	MUTANT
DO49141	PACA-AU	MUTANT
DO49142	PACA-AU	MUTANT
DO49149	PACA-AU	MUTANT
DO49150	PACA-AU	MUTANT
DO49151	PACA-AU	MUTANT
DO49158	PACA-AU	MUTANT
DO49161	PACA-AU	MUTANT
DO49195	PACA-AU	MUTANT
DO221547	PACA-CA	MUTANT
DO224596	PACA-CA	MUTANT
DO227482	PACA-CA	MUTANT
DO227485	PACA-CA	MUTANT
DO227554	PACA-CA	MUTANT
DO35112	PACA-CA	MUTANT
DO35118	PACA-CA	MUTANT
DO35122	PACA-CA	MUTANT
DO35216	PACA-CA	MUTANT
DO35376	PACA-CA	MUTANT
DO49422	PACA-CA	MUTANT
DO49445	PACA-CA	MUTANT
DO49466	PACA-CA	MUTANT
DO49484	PACA-CA	MUTANT
DO51499	PACA-CA	MUTANT
DO51544	PACA-CA	MUTANT
DO51546	PACA-CA	MUTANT
DO35114	PACA-CA	MUTANT
DO35186	PACA-CA	MUTANT
DO35200	PACA-CA	MUTANT
DO51477	PACA-CA	MUTANT
DO51486	PACA-CA	MUTANT
DO51488	PACA-CA	MUTANT
DO32843	PACA-AU	MUTANT
DO32853	PACA-AU	MUTANT
DO32860	PACA-AU	MUTANT

DO32863	PACA-AU	MUTANT
DO32924	PACA-AU	MUTANT
DO32992	PACA-AU	MUTANT
DO33024	PACA-AU	MUTANT
DO33105	PACA-AU	MUTANT
DO33112	PACA-AU	MUTANT
DO33176	PACA-AU	MUTANT
DO33624	PACA-AU	MUTANT
DO33704	PACA-AU	MUTANT
DO33712	PACA-AU	MUTANT
DO33776	PACA-AU	MUTANT
DO33800	PACA-AU	MUTANT
DO34032	PACA-AU	MUTANT
DO34064	PACA-AU	MUTANT
DO34112	PACA-AU	MUTANT
DO34136	PACA-AU	MUTANT
DO34144	PACA-AU	MUTANT
DO34464	PACA-AU	MUTANT
DO34624	PACA-AU	MUTANT
DO34664	PACA-AU	MUTANT
DO34672	PACA-AU	MUTANT
DO35073	PACA-AU	MUTANT
DO46673	PACA-AU	MUTANT
DO46676	PACA-AU	MUTANT
DO46688	PACA-AU	MUTANT
DO46694	PACA-AU	MUTANT
DO46709	PACA-AU	MUTANT
DO46712	PACA-AU	MUTANT
DO46724	PACA-AU	MUTANT
DO49118	PACA-AU	MUTANT
DO49159	PACA-AU	MUTANT
DO35086	PACA-CA	MUTANT
DO35090	PACA-CA	MUTANT
DO35094	PACA-CA	MUTANT
DO35104	PACA-CA	MUTANT
DO35120	PACA-CA	MUTANT
DO35134	PACA-CA	MUTANT
DO35146	PACA-CA	MUTANT
DO35156	PACA-CA	MUTANT
DO35158	PACA-CA	MUTANT
DO35166	PACA-CA	MUTANT
DO35172	PACA-CA	MUTANT

DO35174	PACA-CA	MUTANT
DO35176	PACA-CA	MUTANT
DO35178	PACA-CA	MUTANT
DO35180	PACA-CA	MUTANT
DO35188	PACA-CA	MUTANT
DO35192	PACA-CA	MUTANT
DO35196	PACA-CA	MUTANT
DO35204	PACA-CA	MUTANT
DO35206	PACA-CA	MUTANT
DO35208	PACA-CA	MUTANT
DO35212	PACA-CA	MUTANT
DO35218	PACA-CA	MUTANT
DO35234	PACA-CA	MUTANT
DO35239	PACA-CA	MUTANT
DO35245	PACA-CA	MUTANT
DO35254	PACA-CA	MUTANT
DO35315	PACA-CA	MUTANT
DO35325	PACA-CA	MUTANT
DO35345	PACA-CA	MUTANT
DO35388	PACA-CA	MUTANT
DO35394	PACA-CA	MUTANT
DO35460	PACA-CA	MUTANT
DO35478	PACA-CA	MUTANT
DO35484	PACA-CA	MUTANT
DO35502	PACA-CA	MUTANT
DO32827	PACA-AU	MUTANT
DO32839	PACA-AU	MUTANT
DO32845	PACA-AU	MUTANT
DO32849	PACA-AU	MUTANT
DO32855	PACA-AU	MUTANT
DO32857	PACA-AU	MUTANT
DO32866	PACA-AU	MUTANT
DO32869	PACA-AU	MUTANT
DO32872	PACA-AU	MUTANT
DO32881	PACA-AU	MUTANT
DO32884	PACA-AU	MUTANT
DO32890	PACA-AU	MUTANT
DO32920	PACA-AU	MUTANT
DO32932	PACA-AU	MUTANT
DO32940	PACA-AU	MUTANT
DO32948	PACA-AU	MUTANT
DO32952	PACA-AU	MUTANT

DO32956	PACA-AU	MUTANT
DO32964	PACA-AU	MUTANT
DO32988	PACA-AU	MUTANT
DO32996	PACA-AU	MUTANT
DO33004	PACA-AU	MUTANT
DO33084	PACA-AU	MUTANT
DO33098	PACA-AU	MUTANT
DO33120	PACA-AU	MUTANT
DO33136	PACA-AU	MUTANT
DO33144	PACA-AU	MUTANT
DO33192	PACA-AU	MUTANT
DO33216	PACA-AU	MUTANT
DO33224	PACA-AU	MUTANT
DO33232	PACA-AU	MUTANT
DO33280	PACA-AU	MUTANT
DO33320	PACA-AU	MUTANT
DO33328	PACA-AU	MUTANT
DO33360	PACA-AU	MUTANT
DO33464	PACA-AU	MUTANT
DO33520	PACA-AU	MUTANT
DO33536	PACA-AU	MUTANT
DO33576	PACA-AU	MUTANT
DO33584	PACA-AU	MUTANT
DO33592	PACA-AU	MUTANT
DO33608	PACA-AU	MUTANT
DO33616	PACA-AU	MUTANT
DO33672	PACA-AU	MUTANT
DO33696	PACA-AU	MUTANT
DO33720	PACA-AU	MUTANT
DO33728	PACA-AU	MUTANT
DO33760	PACA-AU	MUTANT
DO33768	PACA-AU	MUTANT
DO33784	PACA-AU	MUTANT
DO33816	PACA-AU	MUTANT
DO33832	PACA-AU	MUTANT
DO33856	PACA-AU	MUTANT
DO33864	PACA-AU	MUTANT
DO33880	PACA-AU	MUTANT
DO33888	PACA-AU	MUTANT
DO33904	PACA-AU	MUTANT
DO33912	PACA-AU	MUTANT
DO33920	PACA-AU	MUTANT

DO33968	PACA-AU	MUTANT
DO33976	PACA-AU	MUTANT
DO33992	PACA-AU	MUTANT
DO34000	PACA-AU	MUTANT
DO34008	PACA-AU	MUTANT
DO34016	PACA-AU	MUTANT
DO34040	PACA-AU	MUTANT
DO34048	PACA-AU	MUTANT
DO34160	PACA-AU	MUTANT
DO34192	PACA-AU	MUTANT
DO34200	PACA-AU	MUTANT
DO34208	PACA-AU	MUTANT
DO34232	PACA-AU	MUTANT
DO34272	PACA-AU	MUTANT
DO34280	PACA-AU	MUTANT
DO34320	PACA-AU	MUTANT
DO34328	PACA-AU	MUTANT
DO34392	PACA-AU	MUTANT
DO34456	PACA-AU	MUTANT
DO34480	PACA-AU	MUTANT
DO34520	PACA-AU	MUTANT
DO34544	PACA-AU	MUTANT
DO34576	PACA-AU	MUTANT
DO34632	PACA-AU	MUTANT
DO34648	PACA-AU	MUTANT
DO34688	PACA-AU	MUTANT
DO34744	PACA-AU	MUTANT
DO34760	PACA-AU	MUTANT
DO34769	PACA-AU	MUTANT
DO34825	PACA-AU	MUTANT
DO34833	PACA-AU	MUTANT
DO34841	PACA-AU	MUTANT
DO34865	PACA-AU	MUTANT
DO34873	PACA-AU	MUTANT
DO34881	PACA-AU	MUTANT
DO34889	PACA-AU	MUTANT
DO34897	PACA-AU	MUTANT
DO34913	PACA-AU	MUTANT
DO34929	PACA-AU	MUTANT
DO34969	PACA-AU	MUTANT
DO34977	PACA-AU	MUTANT
DO34985	PACA-AU	MUTANT

DO35001	PACA-AU	MUTANT
DO35033	PACA-AU	MUTANT
DO35041	PACA-AU	MUTANT
DO35049	PACA-AU	MUTANT
DO46679	PACA-AU	MUTANT
DO46682	PACA-AU	MUTANT
DO46685	PACA-AU	MUTANT
DO46697	PACA-AU	MUTANT
DO46700	PACA-AU	MUTANT
DO46703	PACA-AU	MUTANT
DO46706	PACA-AU	MUTANT
DO46715	PACA-AU	MUTANT
DO46718	PACA-AU	MUTANT
DO46721	PACA-AU	MUTANT
DO46727	PACA-AU	MUTANT
DO46730	PACA-AU	MUTANT
DO46733	PACA-AU	MUTANT
DO46736	PACA-AU	MUTANT
DO46739	PACA-AU	MUTANT
DO46741	PACA-AU	MUTANT
DO49075	PACA-AU	MUTANT
DO49089	PACA-AU	MUTANT
DO49094	PACA-AU	MUTANT
DO49096	PACA-AU	MUTANT
DO49099	PACA-AU	MUTANT
DO49100	PACA-AU	MUTANT
DO49101	PACA-AU	MUTANT
DO49109	PACA-AU	MUTANT
DO49111	PACA-AU	MUTANT
DO49115	PACA-AU	MUTANT
DO49116	PACA-AU	MUTANT
DO49121	PACA-AU	MUTANT
DO49123	PACA-AU	MUTANT
DO49125	PACA-AU	MUTANT
DO49126	PACA-AU	MUTANT
DO49128	PACA-AU	MUTANT
DO49134	PACA-AU	MUTANT
DO49136	PACA-AU	MUTANT
DO49140	PACA-AU	MUTANT
DO49143	PACA-AU	MUTANT
DO49144	PACA-AU	MUTANT
DO49146	PACA-AU	MUTANT

DO49147	PACA-AU	MUTANT
DO49148	PACA-AU	MUTANT
DO49152	PACA-AU	MUTANT
DO49154	PACA-AU	MUTANT
DO49156	PACA-AU	MUTANT
DO49157	PACA-AU	MUTANT
DO49160	PACA-AU	MUTANT
DO49162	PACA-AU	MUTANT
DO49163	PACA-AU	MUTANT
DO49165	PACA-AU	MUTANT
DO49176	PACA-AU	MUTANT
DO49182	PACA-AU	MUTANT
DO49188	PACA-AU	MUTANT
DO49189	PACA-AU	MUTANT
DO49190	PACA-AU	MUTANT
DO49194	PACA-AU	MUTANT
DO49197	PACA-AU	MUTANT
DO49200	PACA-AU	MUTANT
DO49202	PACA-AU	MUTANT
DO49203	PACA-AU	MUTANT
DO35084	PACA-CA	MUTANT
DO35102	PACA-CA	MUTANT
DO35106	PACA-CA	MUTANT
DO35110	PACA-CA	MUTANT
DO35150	PACA-CA	MUTANT
DO35154	PACA-CA	MUTANT
DO35162	PACA-CA	MUTANT
DO35270	PACA-CA	MUTANT
DO35275	PACA-CA	MUTANT
DO35100	PACA-CA	MUTANT

Appendix-3

The command line used for bedtools analysis.

Bedtools intersect -a input -b reference genome table browser.bed -wo > output.txt

Appendix- 4

The command line that was used for oncofuse.

Java -Xmx1G -jar Oncofuse.jar Input.txt coord - outcoord.txt

Appendix- 5

The ID samples of *KRAS* wild type in RNA sequencing analysis section of Australian cohort.

Number	Samples_Wildype_Au
1	DO49201
2	DO49090
3	DO49087
4	DO49076
5	DO34504
6	DO34432
7	DO33488
8	DO34961

Appendix- 6

The ID samples of *KRAS* mutant in RNA sequencing analysis section of Australian cohort.

Number	Samples Mutant Au
1	DO49204
2	DO49198
3	DO49199
4	DO49193
5	DO49178
6	DO49185
7	DO49183
8	DO49184
9	DO49181
10	DO49168
11	DO49166
12	DO49175
13	DO49172
14	DO49138
15	DO49137
16	DO49135
17	DO49133
18	DO49129
19	DO49127
20	DO49130
21	DO49113
22	DO49105
23	DO49079
24	DO49078
25	DO49080
26	DO49074
27	DO33256
28	DO34448
29	DO33128
30	DO33472
31	DO33480
32	DO34785
33	DO34793
34	DO34728
35	DO34720
36	DO33408
37	DO34736
38	DO33400
39	DO33392
40	DO33376
41	DO34680
42	DO33368
43	DO34696
44	DO33336
45	DO33344
46	DO34640
47	DO34656
48	DO34608
49	DO34600
50	DO34616
51	DO34368

52	DO34376
53	DO34336
54	DO33091
55	DO34312
56	DO34264
57	DO34240
58	DO34288
59	DO33984
60	DO32936
61	DO32900
62	DO32887
63	DO32860
64	DO32863
65	DO32875
66	DO32878
67	DO32829
68	DO34945
69	DO33600
70	DO34905
71	DO33656
72	DO33632
73	DO34809
74	DO34801
75	DO34817
76	DO33552
77	DO33544
78	DO34849
79	DO33512
80	DO33528

Appendix- 7

The ID samples of *KRAS* wild type in RNA sequencing analysis section of Canadian cohort.

Number	Samples_Wildtype_Ca
1	DO35138_SP125729_SA412846
2	DO51521_SP113890_SA558706
3	DO51517_SP133674_SA533612
4	DO51502_SP125714_SA533642
5	DO51503_SP125738_SA533736
6	DO51491_SP197529_SA600942
7	DO51487_SP125723_SA533639
8	DO51489_SP118083_SA533782
9	DO51476_SP125756_SA533615
10	DO51479_SP125719_SA533649
11	DO49442_SP117980_SA520314
12	DO49430_SP117002_SA520290
13	DO49418_SP118048_SA520258
14	DO227661_SP192369_SA594596
15	DO35330_SP117076_SA533803
16	DO224633_SP197362_SA600909
17	DO224782_SP133956_SA569484
18	DO224719_SP133832_SA569361
19	DO224750_SP133884_SA569407

Appendix- 8

The ID samples of *KRAS* mutant in RNA sequencing analysis section of Canadian cohort.

Number	Samples_Mutant_Ca
1	DO35083_SP125746_SA533585
2	DO35082_SP117323_SA533741
3	DO35085_SP125782_SA533810
4	DO35236_SP117568_SA533744
5	DO35230_SP125753_SA558648
6	DO35242_SP196657_SA601310
7	DO35210_SP196645_SA601301
8	DO35228_SP125781_SA520181
9	DO35226_SP125799_SA533683
10	DO35222_SP125772_SA533718
11	DO35290_SP125699_SA533747
12	DO35258_SP117216_SA520169
13	DO35116_SP125771_SA533781
14	DO35128_SP117908_SA520214
15	DO35126_SP117463_SA520193
16	DO35198_SP125776_SA520229
17	DO35184_SP125808_SA558662
18	DO35152_SP125718_SA533785
19	DO35136_SP125724_SA533709
20	DO35132_SP196616_SA413109
21	DO35148_SP196669_SA601313
22	DO35144_SP117878_SA520283
23	DO35140_SP125687_SA533694
24	DO35098_SP196624_SA412917
25	DO35098_SP116962_SA520264
26	DO230464_SP197578_SA600953
27	DO230464_SP197581_SA600954
28	DO230463_SP197409_SA600919
29	DO230463_SP197430_SA600922
30	DO230463_SP197443_SA600924
31	DO230463_SP197448_SA600925
32	DO230463_SP197454_SA600926
33	DO230463_SP197461_SA600928
34	DO230463_SP197470_SA600930
35	DO230463_SP197476_SA600931
36	DO51548_SP125809_SA533704
37	DO51549_SP125766_SA533695

38	DO51540_SP117722_SA533660
39	DO51541_SP125737_SA533723
40	DO51542_SP197519_SA600940
41	DO51543_SP125780_SA533777
42	DO51545_SP197535_SA600943
43	DO51535_SP113875_SA558701
44	DO51536_SP125762_SA533805
45	DO51537_SP197509_SA600938
46	DO51538_SP125711_SA533706
47	DO51530_SP125698_SA533703
48	DO51531_SP125741_SA533674
49	DO51532_SP125802_SA533772
50	DO51533_SP125811_SA533653
51	DO51534_SP113900_SA558718
52	DO51524_SP125725_SA533734
53	DO51525_SP125793_SA533780
54	DO51526_SP113829_SA600908
55	DO51527_SP125763_SA533684
56	DO51528_SP125686_SA533813
57	DO51529_SP125768_SA533659
58	DO51520_SP125689_SA533688
59	DO51522_SP125785_SA533668
60	DO51523_SP125730_SA533729
61	DO51513_SP197521_SA600941
62	DO51514_SP117949_SA533646
63	DO51515_SP125755_SA533690
64	DO51518_SP197398_SA600916
65	DO51519_SP125795_SA533749
66	DO51510_SP125706_SA533652
67	DO51511_SP197515_SA600939
68	DO51512_SP125761_SA533753
69	DO51504_SP125783_SA533698
70	DO51505_SP117037_SA533665
71	DO51506_SP125750_SA533692
72	DO51507_SP117760_SA533730
73	DO51509_SP113796_SA558703
74	DO51500_SP125752_SA533664
75	DO51501_SP125731_SA533724
76	DO51494_SP113865_SA558705
77	DO51495_SP117656_SA533745
78	DO51496_SP197382_SA600913
79	DO51497_SP125712_SA533764

80	DO51498_SP197505_SA600937
81	DO51490_SP125694_SA533686
82	DO51492_SP117351_SA533757
83	DO51493_SP113837_SA558693
84	DO51483_SP113873_SA558695
85	DO51484_SP125710_SA533743
86	DO51485_SP117476_SA533796
87	DO51480_SP113743_SA558696
88	DO51481_SP125760_SA533663
89	DO51482_SP113884_SA600944
90	DO51472_SP125685_SA533750
91	DO51473_SP117215_SA533580
92	DO51474_SP125770_SA533721
93	DO51475_SP125797_SA533733
94	DO51478_SP125693_SA533643
95	DO51470_SP125717_SA533625
96	DO51469_SP117340_SA533786
97	DO51464_SP133632_SA533783
98	DO51465_SP125716_SA533682
99	DO51466_SP125758_SA533661
100	DO51467_SP125696_SA533676
101	DO51468_SP113727_SA558697
102	DO49478_SP108828_SA558763
103	DO49475_SP197401_SA600917
104	DO49481_SP197405_SA600918
105	DO49469_SP197387_SA600914
106	DO49463_SP117290_SA520318
107	DO49472_SP197392_SA600915
108	DO49457_SP197376_SA600912
109	DO49454_SP117523_SA533752
110	DO49451_SP197372_SA600911
111	DO49460_SP133717_SA569290
112	DO49460_SP192350_SA594419
113	DO49448_SP125739_SA533814
114	DO49439_SP116985_SA520312
115	DO49436_SP117911_SA520306
116	DO49433_SP108735_SA558720
117	DO49427_SP117690_SA520282
118	DO49424_SP197357_SA600907
119	DO49420_SP117261_SA520270
120	DO49421_SP108706_SA558700
121	DO49419_SP125764_SA520262

122	DO227604_SP197538_SA600945
123	DO227625_SP192358_SA594488
124	DO227596_SP192353_SA594448
125	DO227531_SP192391_SA594735
126	DO227531_SP197681_SA600970
127	DO227531_SP192396_SA594754
128	DO227544_SP192398_SA594764
129	DO227544_SP192399_SA594767
130	DO227544_SP197688_SA600971
131	DO227558_SP192404_SA594817
132	DO227558_SP197700_SA600972
133	DO227564_SP197712_SA600973
134	DO227564_SP192410_SA594843
135	DO227570_SP192412_SA594845
136	DO227570_SP192413_SA594858
137	DO227570_SP197716_SA600974
138	DO227581_SP192351_SA594438
139	DO227704_SP192379_SA594665
140	DO227704_SP197585_SA600955
141	DO227721_SP192381_SA594682
142	DO227727_SP192383_SA594700
143	DO227736_SP192385_SA594716
144	DO227742_SP192388_SA594726
145	DO227633_SP197544_SA600946
146	DO227636_SP192361_SA594516
147	DO227636_SP197547_SA600947
148	DO227648_SP192365_SA594567
149	DO227652_SP192367_SA594580
150	DO227671_SP192371_SA594609
151	DO227687_SP192375_SA594645
152	DO227684_SP192373_SA594633
153	DO227695_SP192377_SA594653
154	DO35454_SP125757_SA558649
155	DO35442_SP125732_SA520221
156	DO35424_SP196608_SA412965
157	DO35424_SP117165_SA520292
158	DO35496_SP125695_SA533647
159	DO35406_SP196620_SA412929
160	DO35406_SP125704_SA520300
161	DO35350_SP117643_SA520304
162	DO35365_SP133678_SA569289
163	DO35360_SP125742_SA533725

164	DO35305_SP117440_SA520203
165	DO224575_SP133555_SA569178
166	DO224575_SP191680_SA594270
167	DO224642_SP197366_SA600910
168	DO224648_SP133708_SA569280
169	DO224656_SP133731_SA569303
170	DO224688_SP133764_SA569320
171	DO224698_SP133784_SA569333
172	DO224767_SP133925_SA569447
173	DO224764_SP133914_SA569434
174	DO224770_SP133937_SA569459
175	DO224776_SP133946_SA569467
176	DO224776_SP197568_SA600950
177	DO224779_SP133951_SA569474
178	DO224784_SP133962_SA569491
179	DO224705_SP133802_SA569344
180	DO224712_SP133812_SA569351
181	DO224714_SP133823_SA569357
182	DO224714_SP192363_SA594539
183	DO224724_SP133848_SA569370
184	DO224724_SP192364_SA594549
185	DO224734_SP133858_SA569379
186	DO224745_SP133874_SA569402
187	DO224745_SP197552_SA600948
188	DO224740_SP133866_SA569389
189	DO224752_SP133895_SA569418
190	DO224758_SP133908_SA569428
191	DO224758_SP197558_SA600949
192	DO221539_SP125703_SA558678
193	DO221543_SP125722_SA558715
194	DO221544_SP125804_SA558659
195	DO221541_SP125713_SA558699
196	DO221542_SP125769_SA558653
197	DO221540_SP125807_SA558660
198	DO221545_SP125778_SA558656
199	DO221546_SP125784_SA558658
200	DO231247_SP197497_SA600935
201	DO231256_SP197606_SA600959
202	DO231251_SP197595_SA600957
203	DO231264_SP197631_SA600963
204	DO231278_SP197654_SA600966
205	DO231284_SP197672_SA600968