Investigation into Clinical Relevance of Filtered Out Novel Mutations in the Thermo Fisher Oncomine Focus Assay

A Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at George Mason University

by

Luke R Grissom
Bachelor of Science
Old Dominion University, 2007

Co-Directors: Dr. Donald Seto, Professor
George Mason University
Dr. Amy Smith, Postdoctoral Associate
Virginia Polytechnic Institute and State University

Spring Semester 2021
George Mason University
Fairfax, VA

DEDICATION

To my family who has supported me through this program and my dog Ruger always by my side.

ACKNOWLEDGEMENTS

I would like to thank the many friends, relatives, and supporters who have made this happen. Dr. Smith, and the other members of my committee were of invaluable help.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Oncomine Focus Assay.....OFA
Next Generation Sequencing .....NGS
National Center for Biotechnology Information.....NCBI
Single Nucleotide Polymorphisms..... SNP
Copy Number Variants .....CNV
Multi-nucleotide Variants ..... MNV
Variants of Unknown Significance.....VUS
Formalin Fixed Paraffin Embedded..... FFPE
Reverse Transcription ..... RT
Torrent Variant Caller.....TVC
College of American Pathologists .....CAP
Combined Annotation Dependent Depletion..... CADD
Deleterious Annotation of Genetic Variants using Neural Networks.....DANN
Functional Analysis through Hidden Markov Models.....FATHMM
Genome Wide Annotation of Variants ..... GWAVA

ABSTRACT

INVESTIGATION INTO CLINICAL RELEVANCE OF FILTERED OUT NOVEL MUTATIONS IN THE THERMO FISHER ONCOMINE FOCUS ASSAY

Luke R Grissom, M.S.

George Mason University, 2021

Thesis Director: Dr. Donald Seto and Dr. Amy Smith

The Oncomine Focus Assay (OFA) is an assay developed by Thermo Fisher and is run on their Ion Torrent S5 next generation sequencing (NGS) platform. The purpose of this assay is to detect known clinically relevant "hotspot" driver mutations for various cancer types and report those mutations, along with potential therapies and clinical trials, to the oncologist for the patient being tested. Thermo Fisher's reporter software automatically filters unknown novel mutations out, leaving only known hotspot driver mutations within the reports sent to the oncologists. The potential clinical relevance of these novel mutations is unknown. RStudio was used to compile and filter the exported sequencing data from 25 months of NGS runs from 1336 patient samples, and the PredictSNP2 bioinformatic tool was used to filter out predicted neutral and synonymous variants. Then NCBI and Varsome were utilized to search for relevant clinical data for the unique novel variants. In total, 374 novel SNPs were found to be potentially pathogenic, but only a few

had documented evidence to support pathogenicity. Twelve unique variants in 15 samples were identified that have evidence of pathogenicity with references in NCBI. The remaining novel SNPs predicted to be pathogenic but do not have any documented evidence could be a potential source for future research of clinically relevant novel mutations in oncology.

CHAPTER ONE: INTRODUCTION

The molecular diagnostics lab at Sentara Norfolk General Hospital runs a targeted oncology hotspot mutation assay called Oncomine Focus Assay (OFA). This assay was developed by Thermo Fisher and is run on their Ion Torrent S5 NGS (Next Generation Sequencing) platform. The purpose of this assay is to detect known clinically relevant "hotspot" driver mutations for various cancer types and report those mutations, along with potential therapies and clinical trials, to the oncologist for the patient being tested. Among many of the known driver mutations the OFA assay detects, there are many novel mutations, mostly SNPs (Single Nucleotide Polymorphisms), that are detected in the covered targeted regions of the patient's genome. Thermo Fisher's reporter software automatically filters these unknown novel mutations out, leaving only known hotspot driver mutations within the reports sent to the oncologists. The potential clinical relevance of these novel mutations is unknown, and they are not included in the patient's final report.

Unlike previous assays such as sanger sequencing or single gene qPCR analysis that had limited coverage for targets of newer treatment options, Next Generation Sequencing has accelerated the usage of personalized medicine in modern healthcare due to its lower cost per sample, faster turnaround and superior sequencing read lengths. From disease genotyping to human genome sequencing, the specific features of many

different conditions can be thoroughly assessed and treatment plans tailored specifically to the patient [2]. Cancer treatment is no exception to this advancement. NGS has quickly progressed research into new mutations of interest, the development of targeted therapies and new clinical trials for cancer treatment [5]. Unlike previous single gene assays testing for only mutations in genes such as EGFR, ALK or ROS, multiplexed NGS assays offer a more robust profile that includes many different oncogenes [18]. The detection of hotspot driver mutations in various cancer types can lead to a variety of novel treatments, therapies and clinical trials that have better prognostic outcomes than traditional chemotherapy alone [1]. By detecting these various hotspot mutations, many new targets are available for pharmacogenomics that develop therapies based on the genetic makeup of a person [4]. Identification of these mutations predict both response to targeted therapies and overall prognosis [5]. In one such study on the use of targeted therapies, 64% of lung adenocarcinoma patients had actionable driver mutations with targeted therapies. The median survival was 3.5 years for patients that received targeted therapy versus a median survival of 2.4 years for those who did not receive targeted therapies [18].

Oncomine Focus Assay (OFA) is a targeted multi-biomarker assay that detects single nucleotide variants (SNV), insertion/deletions, copy number variants (CNV) and gene fusions from DNA and RNA in a single workflow [6]. Fifty-two total genes are covered in the assay, with 35 that are DNA hotspot genes, 19 copy number variants and 23 fusion drivers (See Table 1 for a list of the genes covered). The assay is validated at a 98.7% sensitivity and 99.6% specificity [6]. The results of the assay identify current

actionable genetic variants and some known potential future variant targets for personalized cancer therapies [2]. In most cases, only mutations known to be hotspot variants are passed on from the variant caller modules to the reporter software, as designated by Thermo Fisher. There are a few mutations, however, that are labeled novel by the variant caller software, but still filtered into the reporter software. These variants are not labeled as hotspot mutations and instead listed separately, at the end of the report as variants of unknown significance (VUS).

The objective of this study is to discover how many novel mutations filtered out of any reporting had potential pathogenicity. Further, this study aims to determine if there is a relationship between these potentially pathogenic novel variants and other driver hotspot mutations as well as the primary cancer source of the samples analyzed. Even if there is not current clinical significance to these potentially pathogenic novel mutations, their inclusion as variants of unknown significance in the reports sent to the oncologist could be valuable in the future as newer targeted therapies and trials are discovered.

Table 1: List of Genes covered in Oncomine Focus Assay

DNA Hotspot Genes	Copy Number Variants	Fusion Drivers
AKT1, ALK, AR, BRAF, CDK4,	ALK, AR, BRAF, CCND1, CDK4,	ABL1, AKT3, ALK, AXL, BRAF,
CTNNB1, DDR2, EGFR, ERBB2,	CDK6, EGFR, ERBB2, FGFR1,	EGFR, ERBB2, ERG, ETV1,ETV4,
ERBB3, ERBB4, ESR1, FGFR2,	FGFR2, FGFR3, FGFR4, KIT,	ETV5, FGFR1, FGFR2, FGFR3,
FGFR3, GNA11, GNAQ, HRAS,	RAS, MET, MYC, MYCN,	MET, NTRK1, NTRK2, NTRK3,
IDH1, IDH2, JAK1, JAK2, JAK3,	PDGFRA, PIK3CA	PDGFRA, PPARG, RAF1, RET,
KIT, KRAS, MAP2K1, MAP2K2,		ROS1
MET, MTOR, NRAS, PDGFRA,		
PIK3CA, RAF1, RET, ROS1, SMO		

CHAPTER TWO: METHODS

The samples in this study all originate from de-identified patient tumor biopsies preserved in formalin fixed paraffin embedded (FFPE) tissue blocks, confirmed to be malignant by a pathologist. These tumor biopsies primary sources include breast, lung, colon, pancreas, melanoma, glioblastoma and others. The tissue must contain at least 10% tumor cells, as measured by the pathologist. Thinly sliced sections of the FFPE tissue are extracted using the Promega Maxwell RSC FFPE Plus DNA and Promega Maxwell RSC RNA FFPE kits. Both the DNA and RNA are quantitated using the Thermo Fisher Qubit Fluorometer. The minimum concentration of DNA and RNA for the assay is 0.83 ng/uL and 1ng/uL respectively. Any samples not meeting these thresholds are recut and re-extracted. The extracted samples are then amplified using the OFA assay reverse transcription (RT) and/or polymerase chain reaction (PCR) reagents.

After PCR, the amplified samples are then processed for NGS with the Thermo Fisher Ion AmpliSeq Library kit that is designed to prepare amplicon libraries for sequencing. This includes adding a nucleotide barcode to each sample so libraries can be combined onto one NGS chip for sequencing. The NGS semiconductor chips used are the Thermo Fisher ion 520 or ion 530 chip which only differ on the amount of sequencing reads they can support. The ion 520 chip can be used for up to 8 samples (DNA and RNA) plus controls while the ion 530 chip can be used for up to 18 patients (DNA and

RNA) plus controls. The chips are prepared and loaded with the sample libraries using the Thermo Fisher Ion Chef instrument. Afterwards the chips are sequenced on the Thermo Fisher Ion Torrent S5 NGS platform.

Sequencing data is aligned using the hg19 reference genome and all variants are identified by the Torrent Variant Caller (TVC) in the ion torrent toolbox. All detected variants must have at minimum 5% allelic frequency and 250 read coverage to be reportable per assay and lab protocol. Each NGS chip is summarized into a tab separated excel document. This summary includes all variant info for all samples including controls for the run. Every gene and location analyzed is listed including whether they are present/absent, their quality metrics, coverage, variant reference and mutation call as well as sample information. Every successful chip summary run since the beginning of 2019 through the end of January 2021 was downloaded in csv format. RStudio with R 4.0.4 loaded was used to import and concatenate all excel documents into one data frame. Each value was separated into its own column with labels created where necessary. All non-patient samples such as controls, validation samples and College of American Pathologists (CAP) samples were then filtered out. Once only patient samples were left in the data frames, all absent or no call variants were removed. After, any columns not pertinent to this analysis were removed as well. Any variants that did not meet the quality thresholds set by the assay and lab were removed, which included variants with allelic frequencies under 5% and variants with coverage less than 250 reads.

All variants present in all samples over the 25-month period were separated into novel and hotspot lists. The novel variants were then separated into SNP variants and the

non-SNP variants that included multi nucleotide variants (MNV), insertions and deletions. There were only 97 unique non-SNP variants and most were multi nucleotide variants that typically included SNP variants that were separately called due to variations in the variant caller. The few novel insertion and deletions identified were present in a large percentage of samples suggesting germline origins. Due to these factors and the much higher prevalence of novel SNP variants in the data, SNPs remained the focus of the study. The variant names were organized into a format acceptable for the PredictSNP2 tool called SIMPLE format that listed each variant with chromosome ID, location, reference nucleotide call and nucleotide mutation. An example is Chr1,123456789,A,G. The resulting list of SNP variants were filtered into a new data frame that only listed each unique novel SNP variant from the main sample list. This list of unique variants was exported from RStudio into a text file to upload to PredictSNP2.

PredictSNP2 is a bioinformatics tool that combines several other SNP analysis tools to give predictions on whether a variant may be deleterious or neutral along with confidence percentage score for each tool utilized. These tools include CADD (Combined Annotation Dependent Depletion), DANN (Deleterious Annotation of Genetic Variants using Neural Networks), FATHMM (Functional Analysis through Hidden Markov Models), FunSeq2 (Flexible framework to prioritize regulatory mutations from cancer genome sequencing) and GWAVA (Genome Wide Annotation of Variants) [3]. The primary prediction and confidence score resulted is labeled PredictSNP2, which is a consensus classifier for prediction of the effects of nucleotide variants [3]. The results of

PredictSNP2 also classify SNP variants with their function such as exonic or intronic, as well as regulatory, splicing, synonymous or non-synonymous.

The results of the PredictSNP2 analysis from all of the unique novel SNP variants were output to a vcf file that was then imported back into RStudio. This list was loaded into a new data frame that separated out all classifiers, predictions with confidence scores, as well as any other pertinent information the PredictSNP2 tool provided such as potential clinical significance it may have found from NCBI. Clinical significance was only offered on a small selection of variants. Any variants listed as Benign or Likely Benign were filtered out of the data frame. Next, using the prediction from the PredictSNP2 consensus, any variant predicted as neutral and classified as synonymous or intronic was filtered out of the list. The remaining variants predicted to be neutral by PredictSNP2 consensus were left in for further analysis as a few other tools predicted they could be possibly deleterious and they were classified as non-synonymous SNPs. The resulting list of 492 variants were searched against NCBI and varsome using a standard variant format. An example is Chr1:123456789A>G. Any NCBI information and databases with references to each variant were added to the data frame. Finally, variants labeled benign or likely benign were removed.

CHAPTER THREE: RESULTS

There were 1336 samples from all of the NGS runs included in this study. From these samples, 51179 variants were present in the analysis. Of these variants, 35044 were from patient samples that passed quality metrics and coverage thresholds. There were 34209 of these variants listed as novel and 835 identified as hotspot variants. There were 32354 of these novel variants classified as SNPs. As expected, many variants were repeated in multiple samples, and 1241 of those variants were found to be unique. These 1241 unique variants were exported from RStudio and uploaded to PredictSNP2.

PredictSNP2 results

PredictSNP2 provided multiple translational predictions for each SNP as well as classified them as intronic, exonic, splicing, UTR3 or UTR5. Exonic SNPs were then further labeled as synonymous SNP, non-synonymous SNP, or stopgain. The PredictSNP2 results were imported back into RStudio and then filtered to remove SNPs that were predicted to be neutral by the consensus classifier and labeled as synonymous. Some remaining synonymous labeled SNPs were left in since they had alternative predictions from other tools from PredictSNP2 that determined they were deleterious. Also, some SNPs predicted as neutral by the consensus classifier but labeled as nonsynonymous were left in the data frame as well, since some of the other prediction

tools from Predict SNP2 predicted they may be deleterious. In total 492 unique novel SNPs remained to be further analyzed for pathogenicity.

NCBI and Varsome results

Novel SNPs were searched against both the NCBI database as well as Varsome for any clinical information and history of pathogenicity. In total 31 unique novel mutations had published evidence of pathogenicity in clinical applications. After correlating this with reported results from Oncomine Reporter from Thermo Fisher, 19 of these mutations were listed in the reports as Variants of Unknown Significance (VUS). This means Thermo Fisher must already have identified these non-hotspot variants as potentially relevant, but they do not have enough evidence to label them a hotspot mutation as of yet. These 19 mutations were removed from this analysis, leaving 12 novel pathogenic mutations completely filtered out of Thermo Fisher's reporting. These 12 mutations were found in 15 total samples in the dataset. See Table 2 for mutation and PredictSNP2 details of these 15 samples and the 12 novel pathogenic mutations found. Only 2 mutations were found in more than one sample: Chr5:112151184A>G (rs1064793022), an intronic mutation of the APC gene found in 3 Colon Adenocarcinoma samples and ChrX:66941751C>G (rs137852591), an exonic non-synonymous SNV of the AR gene found in 2 Breast Cancer samples.

Table 2: Novel SNP mutations that have evidence of Pathogenicity from NCBI and PredictSNP2
For PredictSNP2 scores (d = deleterious, n = neutral, ?=Uncertain)
Prediction Tool Names (CADD = Combined Annotation Dependent Depletion, DANN = Deleterious Annotation of Genetic Variants using Neural Networks, FATHMM = Functional Analysis through Hidden Markov Models, FunSeq2 = Flexible framework to prioritize regulatory mutations from cancer genome sequencing, and GWAVA (Genome Wide Annotation of Variants)

Sample Name	Source	Mutation	Gene ID	Functional	Exon	PSNP	CADD	DANN	FATTHM	FunSeq2	GWAVA
1	Colon	chr1,11174459,C,T	MTOR	Exonic	Non synonymous SNV	d-87%	d-84%	d-71%	d-80%	d-62%	?-48%
2	Lung	chr10,43615610,C,T	RET	Exonic	Stopgain	d-74%	d-70%	d-71%	d-65%	d-77%	n-71%
3	Lung	chr15,66727451,A,C	MAP2K1	Exonic	Non synonymous SNV	d-87%	d-71%	d-60%	d-83%	d-62%	d-52%
4	Lung	chr15,66729175,G,A	MAP2K1	Exonic	Non synonymous SNV	d-87%	d-84%	d-71%	d-83%	d-62%	?-48%
5	Colon	chr17,41203103,C,A	BRCA1	Exonic	Non synonymous SNV	d-82%	d-80%	d-69%	n-63%	d-62%	d-52%
6	Breast	chr4,55593604,G,A	KIT	Exonic	Stopgain	d-81%	d-60%	d-65%	d-76%	d-77%	d-69%
7	Colon	chr5,112151184,A,G	APC	Intronic	N/A	d-97%	d-77%	d-54%	d-91%	d-62%	d-64%
8	Colon	chr5,112151184,A,G	APC	Intronic	N/A	d-97%	d-77%	d-54%	d-91%	d-62%	d-64%
9	Colon	chr5,112151184,A,G	APC	Intronic	N/A	d-97%	d-77%	d-54%	d-91%	d-62%	d-64%
10	Lung	chr7,116412043,G,C	MET	Exonic	Non synonymous SNV	d-87%	d-74%	d-60%	d-83%	d-61%	d-51%
11	Lung	chr7,55259434,G,A	EGFR	Exonic	Non synonymous SNV	d-87%	d-52%	d-64%	d-67%	d-62%	d-50%
12	Breast	chr8,38271782,C,T	FGFR1	Exonic	Non synonymous SNV	d-87%	d-53%	d-77%	d-80%	d-62%	d-51%
13	Breast	chrX,66941751,C,G	AR	Exonic	Non synonymous SNV	d-82%	d-74%	n-60%	d-77%	d-61%	?-48%
14	Breast	chrX,66941751,C,G	AR	Exonic	Non synonymous SNV	d-82%	d-74%	n-60%	d-77%	d-61%	?-48%
15	Lung	chrX,70349258,C,T	MED12	Exonic	Non synonymous SNV	d-87%	n-56%	d-72%	d-56%	d-61%	d-51%

All other novel SNP mutations were found to be either Benign, Likely Benign, Uncertain pathogenicity, had mixed clinical evidence being marked as either Uncertain or Likely Benign, or returned no results in these databases. All of this data was imported back into RStudio to be further filtered. After removing Benign or Likely Benign mutations I was left with a list 374 novel SNP variants across 242 samples. Together these created 454 data points for analysis due to some mutations appearing in multiple samples and some samples having multiple novel SNP variants of interest.

The remaining data frame was combined with the oncology sample source and data from reported hotspot mutations for each sample in the dataset. The bar graph in Figure 1 highlights the percentage of novel SNPs for each oncology sample source. Lung and colon adenocarcinomas, along with breast cancer make up the majority of samples that have novel SNPs. Unknown source origin samples are also more likely to have a novel SNP. Unknown samples are samples that do not have a clear indication of what the source site is for the carcinoma. Many of these unknowns typically involve potential sources of cholangiocarcinoma, pancreaticobiliary, gallbladder, gastric and the pancreas.

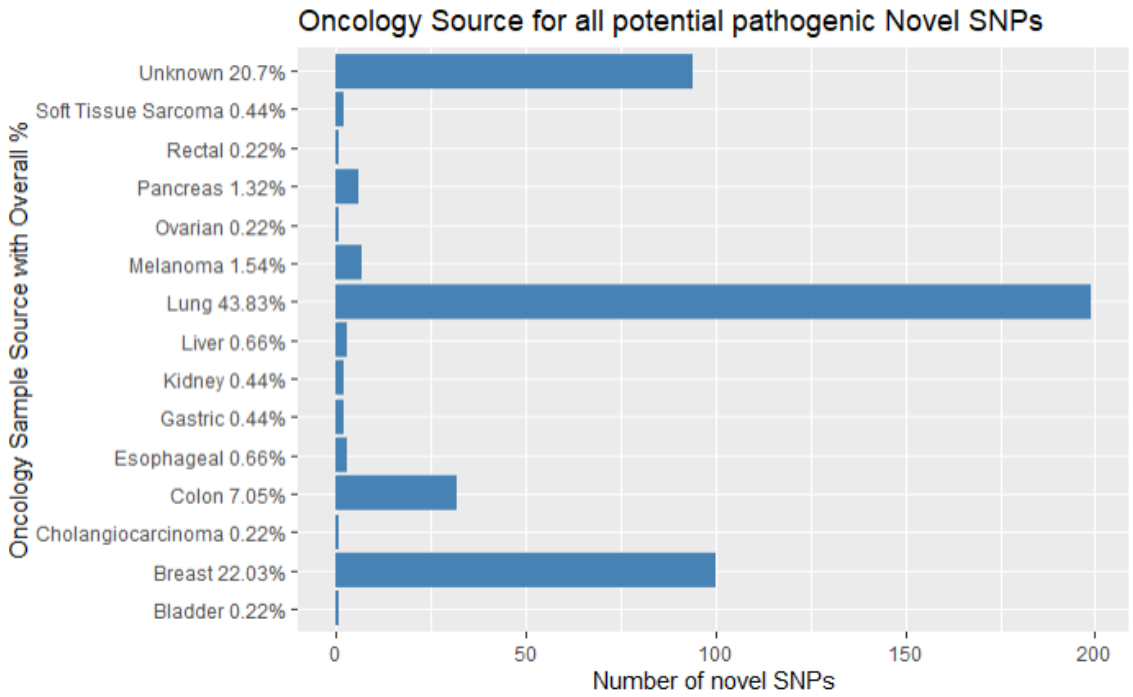

Figure 1: The number of potential pathogenic SNP variants in each oncology sample primary source Percentages of SNPs for each source placed in labels

In figure 2, the oncology source type for each sample containing a potentially pathogenic novel SNP is compared to the overall oncology source data from all samples during this same time frame, which was pulled from the lab's quality database. Most source types seem comparable between both datasets but there is a small but notable decrease of Lung adenocarcinoma sources in the potentially pathogenic novel SNPs as well as an increase in Colon adenocarcinomas. Unknown sources are also represented to a higher degree in the potentially pathogenic novel SNPs.

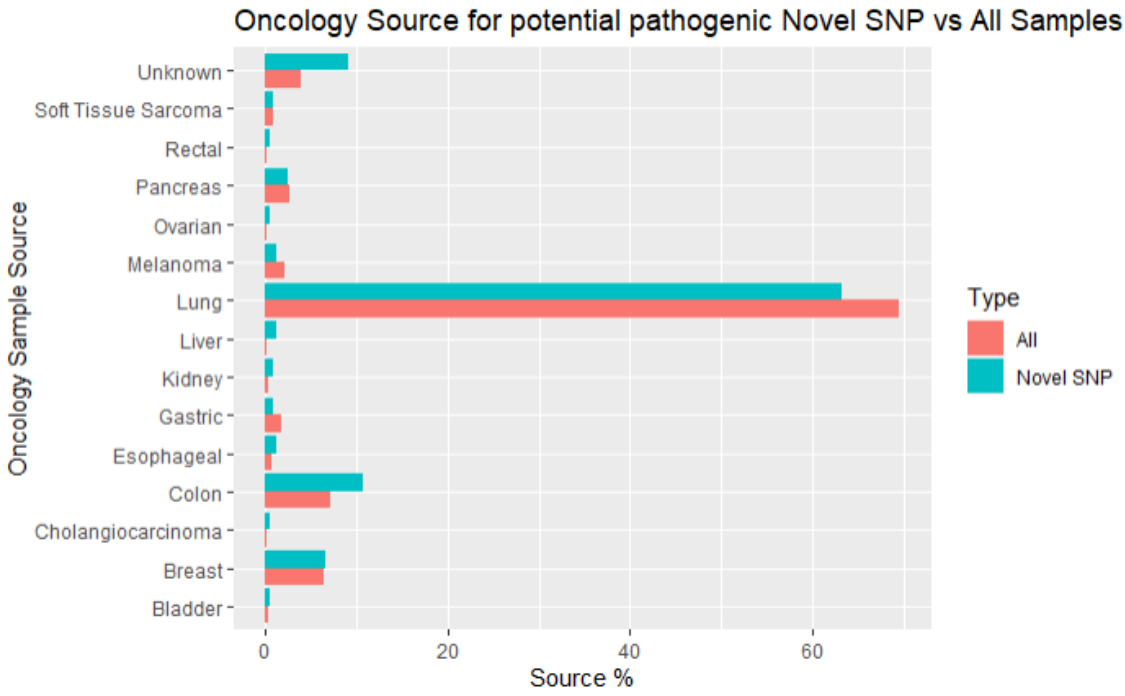

Figure 2: Comparison of the number of samples from each different primary oncology source in all samples in study versus samples with potential pathogenic SNP variants

Comparing both graphs, it's clear the number of samples that have multiple potential pathogenic novel SNPs are higher in samples from breast cancer and Unknown

carcinoma sources. Their representation among total potentially pathogenic novel SNPs is much higher than the number of samples total that contain these mutations. 20.7% and 22.03% of all potentially pathogenic novel SNPs are from Unknown and Breast sources respectively, while only 9.09% and 6.61%, respectively, of samples with these same sources contain these mutations.

In figure 3, the gene IDs of the potentially pathogenic novel SNPs were plotted against the oncology source from the number of samples that exhibited the mutations. By exhibiting this in a density plot, the relation to which potentially pathogenic novel SNP genes occur more often in specific cancer types can be more clearly discerned. While lung adenocarcinomas clearly are the most common cancer type in the sample pool, the novel SNP mutations most common in these lung samples appear to come from ALK, CDK4, EGFR, FGFR3, MET, AR, ERBB2 and MYCN genes. The only other higher density region centers around breast cancer source samples with genes FGFR1, MTOR and PDGFRA containing several potentially pathogenic novel SNPs in these samples.

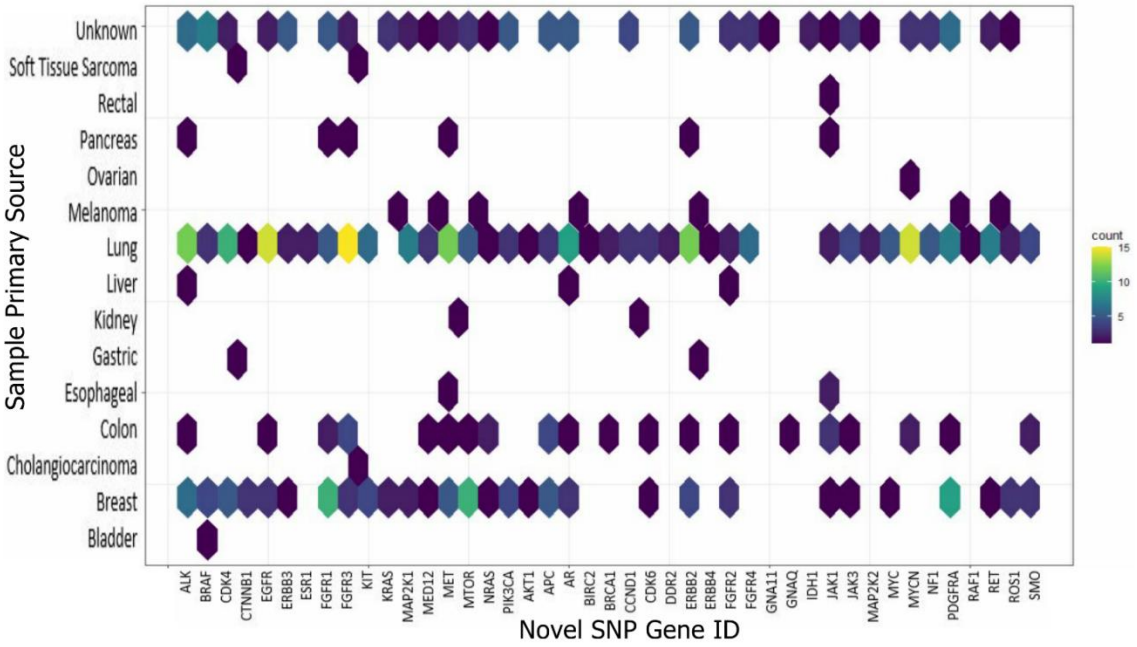

Figure 3: 2D density plot of genes containing potentially pathogenic novel SNPs in relation to the sample oncology source that exhibited those mutations

Last, the reported hotspot mutations found in all samples that contain these potentially pathogenic novel SNPs were imported manually from the lab quality database. These were inserted into a data frame with the Gene IDs of all of the related novel SNPs and displayed in a heatmap in Figure 4. The heatmap compares the genes from all hotspot mutations in the samples that also contain potentially pathogenic novel SNPs. Samples that contain both correlate to the values in the heatmap. Many samples did not have any hotspot mutations found. Those that do don't seem to have a strong correlation with any novel SNP except for potentially KRAS and EGFR hotspot mutations. KRAS hotspot mutations seem to contain a higher amount of FGFR3 potentially pathogenic novel SNPs. EGFR hotspot mutations contain a high amount of other EGFR potentially pathogenic novel SNPs.

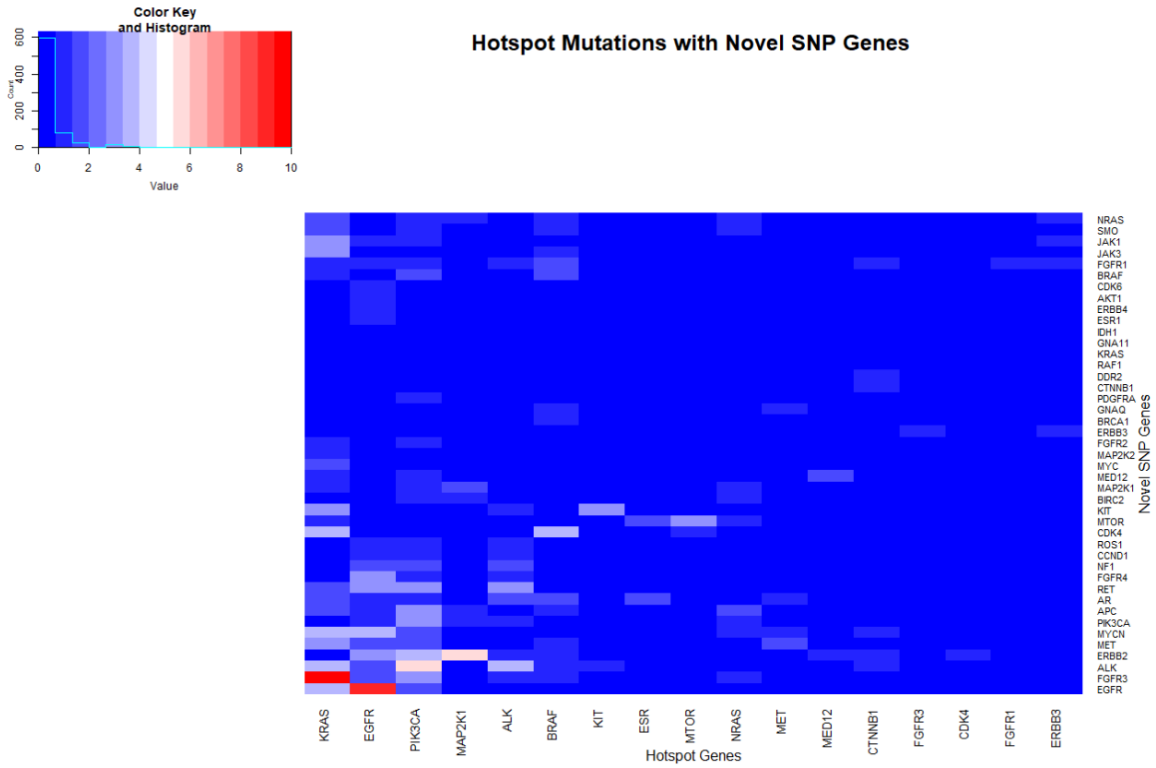

Figure 4: Heatmap of potentially pathogenic novel SNP Genes in relation to the Hotspot Mutations found in matching samples. Color value represents number of instances a variant from a hotspot gene matches with a gene from a novel SNP

CHAPTER FOUR: DISCUSSION

With over 1300 patient samples examined, over 1200 unique novel SNPs were found. Using PredictSNP2 and NCBI, 12 confirmed pathogenic novel SNPs were found with clinical evidence to support them. In total 374 novel SNPs showed the potential to be pathogenic with predictions of being deleterious by Predict SNP. Analysis of both the samples containing these novel SNPs and the total number of novel SNPs found, it is clear that breast cancer samples contain several of these potentially pathogenic novel SNPs per sample with the percent of the population from breast cancer samples being much higher in total SNPs vs samples that contained one or more novel SNPs. Further exploration of these novel SNPs could lead to more targets to report for breast cancer samples since current reporting focuses on PIK3CA and ERBB2 genes in this assay. Conversely, the far lower number of potentially pathogenic novel SNPs total found in lung adenocarcinomas versus the number of lung samples that showed a novel SNP suggests that the many target genes focused on in lung adenocarcinoma reduces the number of unexplored novel mutations.

Relationships between the specific genes of these potentially pathogenic novel SNPs and the sample oncology source or the associated hotspot mutations expressed in these samples show some correlation between a few novel SNP genes. In lung adenocarcinoma samples, EGFR hotspot mutations had a high correlation with novel

SNPs from EGFR genes. There were 25 total lung adenocarcinoma samples with EGFR hotspot mutations. Of these 25 samples, 8 also contained potentially pathogenic novel SNP EGFR variants giving a correlation of 32% between these two types of EGFR mutations in this study. These novel SNPs were also found at a higher rate in lung adenocarcinomas. The reason for this may be due to the higher rate of EGFR hotspot mutations found in lung adenocarcinomas leading to an increased sample bias in this data since lung adenocarcinomas make up a large percentage of samples in the dataset. There is also a high correlation found between KRAS hotspot mutations and novel SNPs from the FGFR3 gene. While KRAS hotspot mutations occur at a high rate in many of these different oncology samples, the relationship with FGFR3 cannot be easily written off.

For the 12 novel SNPs that have evidence of pathogenicity, 2 are related to non-cancer diseases. chr1:11174459C>T (MTOR:p.V2406M) has correlation to Macrocephaly, which is large head circumference [7]. chr8:38271782C>T (FGFR1:p.E692K) has correlation to arhinencephaly, which is a rare non-syndromic central nervous system malformation [8]. For the other 10 novel SNPs, each has correlation to specific carcinomas: chr10:43615610C>T (RET:p.R643*) is correlated to multiple endocrine neoplasia type 2 (MEN2) [9]; chr15:66727451A>C (MAP2K1:p.Q56P) is related to lung adenocarcinomas [10] ; chr15:66729175G>A (MAP2K1:p.G128D) has correlation with malignant melanoma [11]; chr17:41203103C>A (BRCA1:p.G1770A) is related to breast and ovarian malignancy [12]; chr4:55593604G>A (KIT:p.W553*) is related to Gastrointestinal stromal tumors (GIST) [9]; chr5:112151184A>G (APC:c.835-8A>G) is linked to colorectal cancer [13];

chr7:116412043G>C (MET:p.D1028H) has evidence that links it to lung adenocarcinoma [14]; chr7:55259434G>A (EGFR:p.R564H) has correlation with lung mesothelioma [15]; chrX:66941751C>G (AR:c.2395C>G) is linked with breast cancer and Reifenstein syndrome [16]; chrX:70349258C>T (MED12:p.L1224F) is correlated with pancreatic cancer [17].

MAP2K1:p.Q56P, APC:c.835-8A>G, MET:p.D1028H, EGFR:p.R564H, AR:c.2395C>G, and MED12:p.L1224F were all identified in samples that were sourced from the same carcinomas that they are linked to. This means 9 of 15 (60%) total samples that contained these pathogenic novel SNP mutations were found in samples matching their reported pathogenicity. 9 of 12 (75%) total samples with pathogenic novel SNPs linked to cancer were found in samples matching their reported pathogenicity. This helps support the evidence these novel SNPs have a likely relation to the carcinoma they are found in. Further investigation and at minimum, inclusion into oncology reports as variants of unknown significance (VUS) seems to be supported by the evidence. The MET:p.D1028H variant has published evidence of trial responsiveness to the drug crizotinib in a lung adenocarcinoma patient which suggests it could be a viable mutation for targeted therapy [14].

CHAPTER FIVE: CONCLUSION

While only 12 novel SNP mutations were found with clinical evidence to support being pathogenic, most evidence supported their clinical relevance to the cancer type they were found in. These 10 with evidence to carcinoma related mutations should be included in the NGS reports sent to oncologists so any potential treatments being studied to target these mutations, such as with crizotinib in the MET:p.D1028H variant, could be utilized by the oncologist in patient care. Furthermore, if new trials are started up at a later date that target any of these mutations, their inclusion in the report can support the oncologist to look at these trials without any new testing. For the other 362 potentially pathogenic novel mutations that do not have current clinical evidence, these can be a source for new studies on their relevance to the cancer samples they are found in. In particular, additional novel EGFR SNP mutations found in patients with identified hotspot EGFR mutations suggest multiple potentially relevant mutations in this gene that may influence how effective targeted therapies work for these hotspot EGFR mutations.

The high correlation of novel FGFR3 mutations in samples with hotspot KRAS mutations should also be studied further as well. FGFR3 is a growth factor receptor and can be tied into several cancer pathways including those that include ras mutations like KRAS. In one study, KRAS and FGFR3 mutations were shown to cooperate together to promote tumorigenesis in the skin and the lung [19]. These novel FGFR3 mutations were

found at a high rate in lung adenocarcinomas, further supporting them as potential therapy targets. These findings support the notion that many of these novel SNP mutations have some relevance to these carcinomas from which they were derived, and more effort to include them in NGS reporting should be taken.

REFERENCES

1. Shah D, Masters G. (2021) Hotspot Testing Versus Next-Generation Sequencing for NSCLC (2017). ASCO Connection. Available at: https://connection.asco.org/magazine/current-controversies-oncology/hotspot-testing-versus-next-generation-sequencing-nsclc
2. Williams, H.L., Walsh, K., Diamond, A., Oniscu, A., and Deans, Z.C. (2018). Validation of the Oncomine™ focus panel for next-generation sequencing of clinical tumor samples. Virchows Arch *473*, 489–503.
3. Bendl, J., Musil, M., Štourač, J., Zendulka, J., Damborský, J., and Brezovský, J. (2016). PredictSNP2: A Unified Platform for Accurately Evaluating SNP Effects by Exploiting the Different Characteristics of Variants in Distinct Genomic Regions. PLoS Comput Biol *12*, e1004962.
4. Nussinov, R., Jang, H., Tsai, C.-J., and Cheng, F. (2019). Review: Precision medicine and driver mutations: Computational methods, functional assays and conformational principles for interpreting cancer drivers. PLoS Comput Biol *15*. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6438456/ [Accessed April 8, 2021].
5. Li, W., She, H., Tian, X., Neja, M., Harris, A., and Li, Y. (2019). Abstract 3007: Validation of an NGS panel assay for detection of hotspot cancer somatic mutations. Cancer Res *79*, 3007–3007.

6. Oncomine Focus Assay - US Available at: //www.thermofisher.com/us/en/home/clinical/preclinical-companion-diagnostic-development/oncomine-oncology/oncomine-focus-assay.html
7. NM_004958.4(MTOR):c.7216G>A (p.Val2406Met) AND Macrocephalus (2019). Available at: https://www.ncbi.nlm.nih.gov/clinvar/RCV000768372/.
8. NM_023110.2(FGFR1):c.2074G>A (p.Glu692Lys) AND Holoprosencephaly sequence (2016). Available at: https://www.ncbi.nlm.nih.gov/clinvar/47231103/.
9. Nykamp, K., Anderson, M., Powers, M., Garcia, J., Herrera, B., Ho, Y.-Y., Kobayashi, Y., Patil, N., Thusberg, J., Westbrook, M., *et al.* (2017). Sherloc: a comprehensive refinement of the ACMG-AMP variant classification criteria. Genet Med *19*, 1105–1117.
10. Marks, J.L., Gong, Y., Chitale, D., Golas, B., McLellan, M.D., Kasai, Y., Ding, L., Mardis, E.R., Wilson, R.K., Solit, D., *et al.* (2008). Novel MEK1 mutation identified by mutational analysis of epidermal growth factor receptor signaling pathway genes in lung adenocarcinoma. Cancer Res *68*, 5524–5528.
11. Emery, C.M., Vijayendran, K.G., Zipser, M.C., Sawyer, A.M., Niu, L., Kim, J.J., Hatton, C., Chopra, R., Oberholzer, P.A., Karpova, M.B., *et al.* (2009). MEK1 mutations confer resistance to MEK and B-RAF inhibition. Proc Natl Acad Sci U S A *106*, 20411–20416.
12. Bouwman, P., van der Gulden, H., van der Heijden, I., Drost, R., Klijn, C.N., Prasetyanti, P., Pieterse, M., Wientjens, E., Seibler, J., Hogervorst, F.B.L., *et al.* (2013). A high-throughput functional complementation assay for classification of BRCA1 missense variants. Cancer Discov *3*, 1142–1155.
13. Terlouw, D., Suerink, M., Boot, A., Wezel, T. van, Nielsen, M., and Morreau, H. (2020). Recurrent APC Splice Variant c.835-8A>G in Patients With Unexplained Colorectal

Polyposis Fulfilling the Colibactin Mutational Signature. Gastroenterology *159*, 1612-1614.e5.

14. Waqar, S.N., Cottrell, C.E., and Morgensztern, D. (2015). MET mutation associated with responsiveness to crizotinib. J Thorac Oncol *10*, e29–e31.
15. Foster, J.M., Radhakrishna, U., Govindarajan, V., Carreau, J.H., Gatalica, Z., Sharma, P., Nath, S.K., and Loggie, B.W. (2010). Clinical implications of novel activating EGFR mutations in malignant peritoneal mesothelioma. World J Surg Oncol *8*, 88.
16. Wooster, R., Mangion, J., Eeles, R., Smith, S., Dowsett, M., Averill, D., Barrett-Lee, P., Easton, D.F., Ponder, B.A., and Stratton, M.R. (1992). A germline mutation in the androgen receptor gene in two brothers with breast cancer and Reifenstein syndrome. Nat Genet *2*, 132–134.
17. Chang, M.T., Asthana, S., Gao, S.P., Lee, B.H., Chapman, J.S., Kandoth, C., Gao, J., Socci, N.D., Solit, D.B., Olshen, A.B., *et al.* (2016). Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. Nat Biotechnol *34*, 155–163.
18. Kris, M.G., Johnson, B.E., Berry, L.D., Kwiatkowski, D.J., Iafrate, A.J., Wistuba, I.I., Varella-Garcia, M., Franklin, W.A., Aronson, S.L., Su, P.-F., et al. (2014). Using multiplexed assays of oncogenic drivers in lung cancers to select targeted drugs. JAMA 311, 1998–2006.
19. Ahmad, I., Singh, L.B., Foth, M., Morris, C.-A., Taketo, M.M., Wu, X.-R., Leung, H.Y., Sansom, O.J., and Iwata, T. (2011). K-Ras and β-catenin mutations cooperate with Fgfr3 mutations in mice to promote tumorigenesis in the skin and lung, but not in the bladder. Disease Models & Mechanisms 4, 548–555.

BIOGRAPHY

Luke R Grissom graduated from Bayside High School, Virginia Beach, Virginia, in 2002. He received his Bachelor of Science from Old Dominion University in 2007. He was employed as a Forensic Laboratory Specialist for 5 years with the Virginia Department of Forensic Science and then as a Molecular Technologist with Sentara Norfolk General Hospital for over 5 years.