

Understanding Oncogenic Tyrosine Kinase Fusion Driven Cancer: An Investigation into Inflammatory
Myofibroblastic Tumor and the Non-kinase Fusion Partner

By

Merrida Childress

Dissertation

Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements

For the degree of

DOCTOR OF PHILOSOPHY

in

Cancer Biology

December 15, 2018

Nashville, Tennessee

Approved:

Hal Moses M.D.

Scott Hiebert Ph.D.

Timothy Blackwell M.D.

Jason MacGurn Ph.D.

Christine M. Lovly M.D., Ph.D.

PREFACE

This dissertation is being submitted to the Faculty of the Graduate School of Vanderbilt University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The research contained within—involving a genomic investigation into Inflammatory Myofibroblastic Tumor and biochemical assessments of anaplastic lymphoma kinase (ALK) fusion proteins—was conducted by Merrida Childress, under the supervision of Dr. Christine M. Lovly, M.D., Ph.D. between 2014 and 2018. This written work is original, except where acknowledgements and references are made to previously published work. For example, chapter 3 has already been published:

Merrida A. Childress, Stephen M. Himmelberg, Huiqin Chen, Wanleng Deng, Michael A. Davies, Christine M. Lovly, ALK Fusion Partners Impact Response to ALK Inhibition: Differential Effects on Sensitivity, Cellular Phenotypes, and Biochemical Properties. *Mol Cancer Res.* 2018

TABLE OF CONTENTS

	Page
PREFACE	ii
LIST OF TABLES.....	vi
LIST OF FIGURES.....	viii
Chapter	
1 INTRODUCTION.....	1
1.1 Overview.....	1
1.2 Inflammatory Myofibroblastic Tumor.....	1
Clinical Features.....	1
Pathology	2
Molecular Features	3
Kinase Fusions in IMT	3
1.3 ALK Fusions	4
ALK Biology.....	5
ALK Rearranged Cancers.....	5
ALK Fusion Biology	7
Structure and Biological Function	7
Signaling.....	8
Therapeutic Targeting of ALK Fusions	10
1.4 Concluding Remarks	11
1.5 Purpose of These Studies.....	12
1.6 References	14
2 INFLAMMATORY MYOFIBROBLASTIC TUMOR: A GENOMIC AND TRNASCRIPTOMIC ANALYSIS.....	24
2.1 Abstract	24
2.2 Introduction.....	24
2.3 Materials and Methods	26

2.4 Results.....	28
Kinase Fusions Identified.....	28
Patient and Tumor Characterization	32
Pulmonary vs. Extra-pulmonary IMT Expression Analysis.....	35
Expression Analysis of RANBP2-ALK Tumors	36
Fusion Positive vs. Fusion Negative Expression Analysis	37
SNVs	39
2.5 Discussion	41
2.6 References	47
3 ALK FUSION VARIANTS	51
3.1 Abstract	51
Implications.....	51
3.2 Introduction.....	51
3.3 Materials and Methods	53
3.4 Results	57
Expression of ALK Fusion Variants in NIH 3T3 Cells	57
ALK Fusion Partners Confer Differential Oncogenic Properties	59
ALK Fusion Variants Exhibit Differential Response to ALK Tyrosine Kinase Inhibition.....	61
Tyrosine Kinase Activity of ALK Fusion Variants	62
Stability of ALK Fusion Proteins	64
3.5 Discussion	65
3.6 References	70
4 DISCUSSION AND FUTURE DIRECTIONS.....	75
4.1 Overview.....	75
4.2 Inflammatory Myofibroblastic Tumor.....	75
IMT Biology.....	75
Therapeutic Paradigms.....	76
4.3 ALK Fusions	77
Discussion and Unpublished Results.....	77

Ongoing Studies and Future Directions	81
4.4 Concluding Remarks	85
4.5 Methods.....	86
4.6 References	88
5 Appendix	91-212
Chapter II Appendix	91
Chapter III Appendix	188

LIST OF TABLES

Main Tables	Page
1.1 Response to ALK Tyrosine Kinase Inhibitor Therapy in ALK-rearranged Tumors	7
2.1 Annotation of 69 IMT Cases That Underwent Genomic Analyses as Part of This Study	31
2.2 Clinicopathological Features of Inflammatory Myofibroblastic Tumors According to Fusion Status.....	35
2.3 Clinicopathological Features of the Entire IMT Cohort.....	38
3.1 Properties of the ALK Fusion Variants Described in This Study	59
4.1 Enriched Categories from the Differentially Expressed Proteins Between the PRKAR1A-ALK NIH3T3 Cell Line and the WT ALK NIH3T3 Cell Line	78
4.2 Predictive Structural Models of ALK Fusion Variants and 'Linker' Region Lengths.....	83
Appendix Tables	Page
S2.1 KEGG and GO Term Enrichment Analysis for Differentially Expressed Genes in Pulmonary vs. Extra-pulmonary IMT	91
S2.2 KEGG and GO Term Enrichment Analysis for Differentially Expressed Genes in RANBP2-ALK Expressing IMTs vs. All Other IMTs.....	92
S2.3 KEGG and GO Term Enrichment Analysis for Differentially Expressed Genes in Fusion Positive vs. Fusion Negative IMT	96
S2.4 Signaling Pathway Impact Analysis for Differentially Expressed Genes in Pulmonary vs. Extra-pulmonary IMT	97
S2.5 Signaling Pathway Impact Analysis for Differentially Expressed Genes in RANBP2-ALK Expressing IMTs vs. All Other IMTs.....	98
S2.6 Signaling Pathway Impact Analysis for Differentially Expressed Genes in Fusion Positive vs. Fusion Negative IMT	100
S2.7 Cancer Associated SNVs Shared by at Least 25% of IMTs Evaluated.....	101
S2.8 All Cancer Associated SNVs Called for Case 1	103
S2.9 All Cancer Associated SNVs Called for Case 2	108
S2.10 All Cancer Associated SNVs Called for Case 3.....	112
S2.11 All Cancer Associated SNVs Called for Case 4.....	115
S2.12 All Cancer Associated SNVs Called for Case 5.....	119
S2.13 All Cancer Associated SNVs Called for Case 6.....	123

S2.14 All Cancer Associated SNVs Called for Case 7	127
S2.15 All Cancer Associated SNVs Called for Case 9	130
S2.16 All Cancer Associated SNVs Called for Case 10	133
S2.17 All Cancer Associated SNVs Called for Case 11	137
S2.18 All Cancer Associated SNVs Called for Case 12	142
S2.19 All Cancer Associated SNVs Called for Case 13	147
S2.20 All Cancer Associated SNVs Called for Case 14	152
S2.21 All Cancer Associated SNVs Called for Case 15	158
S2.22 All Cancer Associated SNVs Called for Case 65	164
S2.23 All Cancer Associated SNVs Called for Case 66	170
S2.24 All Cancer Associated SNVs Called for Case 67	174
S2.25 All Cancer Associated SNVs Called for Case 69	180
S2.26 Gene Ontology Terms for Commonly Shared SNVs	187
S3.1 MD Anderson RPPA Standard Antibody List	188
S3.2 Proteins Exhibiting Most Significantly Different Expression From the WT ALK Cell Line for Each ALK Fusion Cell Line	201
S3.3 ALK TKIs in Clinical Use	208
S3.4 Compilation of all Phenotypic Data Measurements for ALK Fusion Variants and ALK F1174L	209

LIST OF FIGURES

Main Figures	Page
1.1 The Diversity of ALK Fusions in Cancer	6
1.2 Structural and Biological Characteristics of ALK Fusions.....	9
2.1 Inflammatory Myofibroblastic Tumor is a Kinase Fusion Driven Disease.....	29
2.2 Kinase Fusions in IMT are Heterogeneous	30
2.3 Schematic Representation of Kinase Fusion Proteins Identified in This Study	33
2.4 Anatomical Locations of Primary IMTs Represented by Kinase Type and 5' Partner	34
2.5 Differential Expression Analysis of Pulmonary vs. Extra-pulmonary IMTs	36
2.6 Differential Expression Analysis of RANBP2-ALK Expressing IMTs vs. All Other IMTs.....	37
2.7 Differential Expression Analysis of Kinase Fusion Positive vs. Kinase Fusion Negative IMTs	40
2.8 Genes Harboring Cancer Associated SNVs Shared by at Least 25% of IMTs Evaluated.....	42
2.9 Shared SNVs Present in Fusion Negative Tumors	43
3.1 Schematic Representation of ALK Fusion Variants Utilized in This Study	58
3.2 Active ALK Variants Change the Protein Expression Profile of NIH3T3 Cells.....	60
3.3 ALK Fusion Variants Confer Differential Transformation Ability	61
3.4 ALK Fusion Variants Exhibit Differential Response to ALK TKIs	63
3.5 ALK Fusion Variants Exhibit Differences in Kinase Activity and Protein Turnover	65
3.6 The Heterogeneity of Tyrosine Kinase Fusion Partners.....	69
4.1 KIF5B-ALK is Less Stable Than Wild Type KIF5B.....	79
4.2 ALK Immunofluorescence in NIH3T3 Cells Stably Expressing ALK Variants.....	80
4.3 “Linker” Region of EML4-ALK Variants	81
4.4 Future Directions: Study Approach and Output.....	82
4.5 Naturally Occurring EML4-ALK and TFG-ALK Variants.....	84
Appendix Figures	Page
S3.1 Expression of ALK Fusion Variants	199
S3.2 Active ALK Variants Change the Protein Expression Profile of NIH3T3 Cells	200

S3.3	ALK Fusion Variants Exhibit Differential Response to ALK TKIs	206
S3.4	Immunoprecipitated ALK Fusion Variants for Kinase Assays	211
S3.5	ALK Fusion Variants Exhibit Differences in Protein Turnover and Sensitivity to Ganetespib212	

Chapter I

INTRODUCTION

1.1 Overview

This introductory chapter will provide background information necessary to understand the original research contained in this dissertation. The chapter is divided into three sections: the first is focused on Inflammatory Myofibroblastic Tumor (IMT), a sarcoma primarily driven by oncogenic kinase fusions; the second is focused on oncogenic kinase fusion biology with an emphasis on Anaplastic Lymphoma Kinase (ALK) fusions, and the third is reserved for concluding remarks. The first section will concentrate on our current knowledge of IMT including pathological and genetic descriptions. The second section will detail our current understanding of the biology of *ALK* fusions in cancer, including the signaling pathways that are activated downstream of ALK, the role of the 5' partner proteins in mediating ALK fusion oligomerization, subcellular localization, and kinase activation, and the clinical implications for targeting ALK fusions.

1.2 Inflammatory Myofibroblastic Tumor

Clinical Features

Inflammatory myofibroblastic tumor (IMT) is a rare, mesenchymal neoplasm that has a tendency for local recurrence. While it can occur at any age and in multiple anatomic locations, it arises primarily in the soft tissue and viscera of children, adolescents, and young to middle-aged adults. Tumors average 6 cm in size, but can range from 1 cm to >20 cm in diameter^{5,6}. IMTs in the abdomen or retroperitoneum, often present as multiple discrete masses in the same region^{1,5-7}. Surgical resection is the current standard of care for IMT. There is no systemic standard of care therapy for patients whose tumors cannot be resected. Although, regression and response to anti-tumor necrosis factor- α binding antibody, corticosteroids, non-steroidal anti-inflammatory drugs (NSAIDs), and chemotherapy have been reported in a few cases⁸⁻¹¹. A subset of patients experience an inflammatory clinical syndrome of fever, weight loss, growth failure, anemia, thrombocytosis, polyclonal hyperglobulinemia, and an elevated erythrocyte sedimentation rate^{12,13}. This syndrome relaxes following tumor resection, and a return of these clinical and laboratory abnormalities may indicate disease recurrence^{5,13,14}.

Recurrence rate appears to differ based on anatomical site. Primary tumors confined to the lung recur at a rate of <2%^{1,15,16} compared to 25% for extra-pulmonary disease⁵. Recurrence is most frequently seen in patients presenting with multi-focal intra-abdominal tumors and those in anatomical locations

where complete surgical resection is difficult. Distant metastasis of IMT is rare, striking <5% of patients^{5,17}. The most common sites of metastasis are lung and brain, followed by liver and bone¹⁸.

IMT is diagnosed by its pathology using a set of criteria set by the World Health Organization (WHO). The WHO defines IMT as “a lesion composed of a proliferation of myofibroblastic spindle and stellate cells with abundant eosinophilic cytoplasm mixed with infiltrative plasma, inflammatory cells, lymphocytes and eosinophils”^{19,20}. More recently a distinct subtype of IMT, known as epithelioid Inflammatory Myofibroblastic Sarcoma (eIMS), has been described. This subtype is characterized by a dominance of large epithelioid cells with only minor spindle cell components, noticeable nuclear atypia, and positive anaplastic lymphoma kinase (ALK) immunohistochemistry in a perinuclear pattern²¹, and it demonstrates a more clinically aggressive phenotype and dismal prognosis^{22,23}. Additionally, eIMS is almost exclusively found in the abdominal cavity and has a strong predominance in males²⁴.

Pathology

Three types of histological patterns are described; combinations of which are frequently seen within the same tumor: 1.) a myxoid/vascular type, characterized by loosely arranged, plump spindle tumor cells within in a myxoid, edematous, and vascular stroma, appearing like nodular fasciitis or granulation tissue; 2.) a compact, spindle cell pattern, characterized by tight proliferation of the spindle tumor cells with fascicle architecture and a collagen stroma; and 3.) a hypocellular fibrous pattern, characterized by the proliferation of elongated, plump spindle tumor cells within a dense collagenous stroma⁵. All histological patterns contain distinct infiltration of inflammatory cells. The inflammatory infiltrate in the myxoid regions often consists of greater neutrophils and eosinophils and fewer plasma cells than the other two patterns. The foci of compact, spindle cells typically have abundant plasma cells and lymphocytes closely mixed with the spindle cells; however, isolated lymphoid follicles and plasma cells are also commonly seen. The fibromatosis-like pattern contains scattered lymphocytes, plasma cells and eosinophils¹⁸.

Approximately 50% of IMTs contain dispersed “ganglion-like” cells- larger, polygonal shaped cells with copious amphophilic or eosinophilic cytoplasm with large, vesicular nuclei and obvious nucleoli^{6,25}. Foamy histiocytes are found in a minority of cases^{14,26}. Mitotic counts are typically low (0–2 mitoses per 10 HPF), and atypical mitoses are rare^{5,17,25,26}. While necrosis and vascular invasion have been reported in typical IMTs, they are also very rare^{5,27,28}. While the conventional IMT is considered low grade, IMTs may evolve into an atypical, higher grade, exhibiting increased cellularity,

numerous mitoses, atypical mitotic figures, distinct nuclear atypia, and necrosis^{5,17,29,30}, especially upon recurrence. The cellular morphologies of the higher grade tumors include hypercellular, spindle cell, epithelioid/histiocytoid, or round cell morphology^{5,17,29}.

Histologically, eIMS exhibits different cell types compared to classic IMT. eIMS is composed predominantly of large, epithelioid tumor cells and an inflammatory infiltrate dominated by neutrophils, in contrast to the spindled tumor cells with lymphocyte and plasma cell infiltration observed in the classic IMT²⁴. As this subtype of IMT tends to exhibit a more mixoid background and increased vasculature, it is most similar to first histological pattern described above.

Molecular Features

There are many fundamental aspects about IMT that remain unknown including the cell type that gives rise to IMT. However, distinct molecular markers that distinguish IMT from other tumor types have been established. Due to the myofibroblastic presentation, 80-90% of IMTs are positive for smooth muscle actin, 60-70% express desmin and calponin, and approximately 33% are reactive to keratin by immunohistochemistry (IHC). Given the often focal histology of these tumors, reactivity for these markers is can also be focal^{5,26,29,31}. Variable CD30 expression has been reported for eIMS²³.

Kinase Fusions in IMT

Rearrangements of the *ALK* gene on chromosome 2p23 have been reported in 50-60% of IMTs^{2,32,33}. Several *ALK* fusion partners have been identified in IMT, including *TPM-3/-4*, *ATIC*, *CLTC*, *CARS*, *FN1*, *PRKAR1A*, *EML4*, *RANBP2*, and many others^{2,34,35}. Unlike some other diseases, studies suggest that *ALK* rearranged IMT (except eIMS) generally exhibit a low risk toward metastasis; however, there does not appear to be a correlation with *ALK* reactivity and recurrence^{17,36,37}.

Importantly, *ALK* fusions have been validated as a therapeutic target in IMT. Results from the Children's Oncology Group study for crizotinib in *ALK* positive IMT reported an impressive overall response rate of 85%. A complete response was observed in 36% of patients and a partial response was observed in 50% of patients³⁸. This study demonstrates the importance of identifying *ALK* fusions in IMT and also highlights the importance of finding novel therapeutic targets in *ALK* negative IMT. Recently *ROS1*, *PDGFR β* , *NTRK3*, and *RET* gene alterations have also been reported, bringing kinase fusion detection in IMT to approximately 85%^{2,28,39}.

eIMS are also distinct in their specific expression of kinase fusions. 100% of eIMS cases reported to

date harbor either RANBP2-ALK or RRBP1-ALK fusions with distinctive nuclear membranous ALK IHC or cytoplasmic combined with intense nuclear membranous ALK IHC respectively⁴⁰.

Kinase fusion negative IMTs typically occur in older patients and have a more aggressive histological appearance^{2,17,28}. In addition, the frequency of metastasis appears to be greater in ALK negative IMT compared to ALK positive IMT, with the exception of eIMS. Little is known about the genomic landscape of IMT outside of the presence or absence of oncogenic kinase fusions, and as such, there are no currently available targeted therapies for kinase fusion negative patients.

One recent study of 25 distinct IMTs reported no difference in rate of PD-L1 expression between ALK positive and ALK negative patients, but that patients experiencing metastasis/recurrence (all ALK negative) exhibited greater PD-L1 expression compared to those that did not experience metastasis/recurrence. Additionally *ALK* fusion negative IMTs had greater CD8+ TILs, compared to *ALK* fusion positive tumors⁴¹. This early data suggests that ALK fusion negative IMTs may be good candidates for immune checkpoint blockade therapies, but further studies are needed to gain conclusive answers.

1.3 ALK Fusions

The study of oncogenic tyrosine kinase fusions began with the discovery of the t(9:22) translocation in chronic myelogenous leukemia and its chimeric protein product, BCR-ABL⁴². Since then, other liquid and solid tumors have been found to harbor chromosomal rearrangements that result in activated tyrosine kinase fusions. Notably, rearrangements involving *anaplastic lymphoma kinase (ALK)* have been found across all cell lineages, arising in hematologic, neural, epithelial, and mesenchymal malignancies⁴³. While these cancers can differ in the specific ALK kinase fusion they harbor, they share two features that underlie the importance of this class of oncogenes. First, no other known genetic drivers are found in concurrence with the ALK kinase fusion. This is likely attributed to the strength of the kinase to drive proliferative and anti-apoptotic pathways. Second, they all exhibit a dependency on the activated kinase for proliferation and survival. This dependency, or 'oncogene addiction', makes the cancer highly sensitive to small molecule tyrosine kinase inhibitors (TKIs) and has resulted in the development of several ALK TKIs, including crizotinib, alectinib, ensartinib, lorlatinib, and many others in clinical development⁴⁴⁻⁵⁰.

In this section, I will focus on the current knowledge of the functional and structural biology of ALK

fusion proteins and the therapeutic strategies aimed at targeting these oncogenic kinase fusions. I will discuss the functional domains present in the different ALK fusions, the signaling pathways activated downstream of the fusion, and the role of the 5' fusion partner in mediating fusion activation, subcellular localization, and oligomerization. Additionally, I will address the clinical features of ALK fusion positive (ALK+) malignancies including the types of patients in which they occur and the current therapeutic approaches to target these fusions.

ALK Biology

ALK was originally discovered as part of the characteristic t(2,5) chromosomal translocation found in anaplastic large cell lymphoma, which results in the *NPM-ALK* translocation. Since its discovery, many other *ALK* fusions have been found in several cancers, which initiated investigations into the function of wild-type ALK. Located on chromosome 2p23, wild-type ALK is a receptor tyrosine kinase that belongs to the insulin receptor superfamily. Midkine, pleiotrophin, and many others are proposed ALK ligands that have roles in cell survival, neural development, and tumorigenesis⁵¹⁻⁵⁴. However, the subject of activating ALK ligands is controversial as many of the studies are contradictory⁵⁵⁻⁵⁸. Expression and functional studies show *ALK* transcripts in the developing nervous system, which subside by 3 weeks of age^{59,60}. *ALK*^(-/-) knockout mice only exhibit slight behavioral changes with no identifiable developmental, anatomical or locomotive insufficiencies, leaving the exact function of the ALK receptor unclear⁶¹⁻⁶³.

Investigations into the signaling networks of the ALK receptor have revealed downstream activation of pathways controlling cell proliferation and survival such as JAK-STAT, PI3K-AKT, mTOR, SHH, JUNB, and MAPK⁶⁴⁻⁶⁷. ALK may also interact with several adaptor proteins including: FRS2/3, IRS1, SHC, and GRB2 and affects proteins such as BIM, p27, and cyclinD2⁶⁴⁻⁶⁷. Less understood ALK interactions include Nuclear Interacting Partner of ALK (NIPA), RAC1, CD42, BCAR1, SHP1/2, SRC, and phosphatidylinositol-3-phosphate 5-kinase type III (PINKFYVE)⁶⁸⁻⁷⁶. Notably, to date, ALK signaling has predominantly been studied in the setting of cancer activating mutations fusions, and amplifications and therefore, these results must be interpreted and applied with care.

ALK Rearranged Cancers

ALK rearrangements have been detected in hematologic, mesenchymal, neural, and epithelial cancers including anaplastic large cell lymphoma (ALCL), diffuse large B cell lymphoma (DLBCL), inflammatory myofibroblastic tumor (IMT), ependymoma-like glioma, non-small cell lung cancer

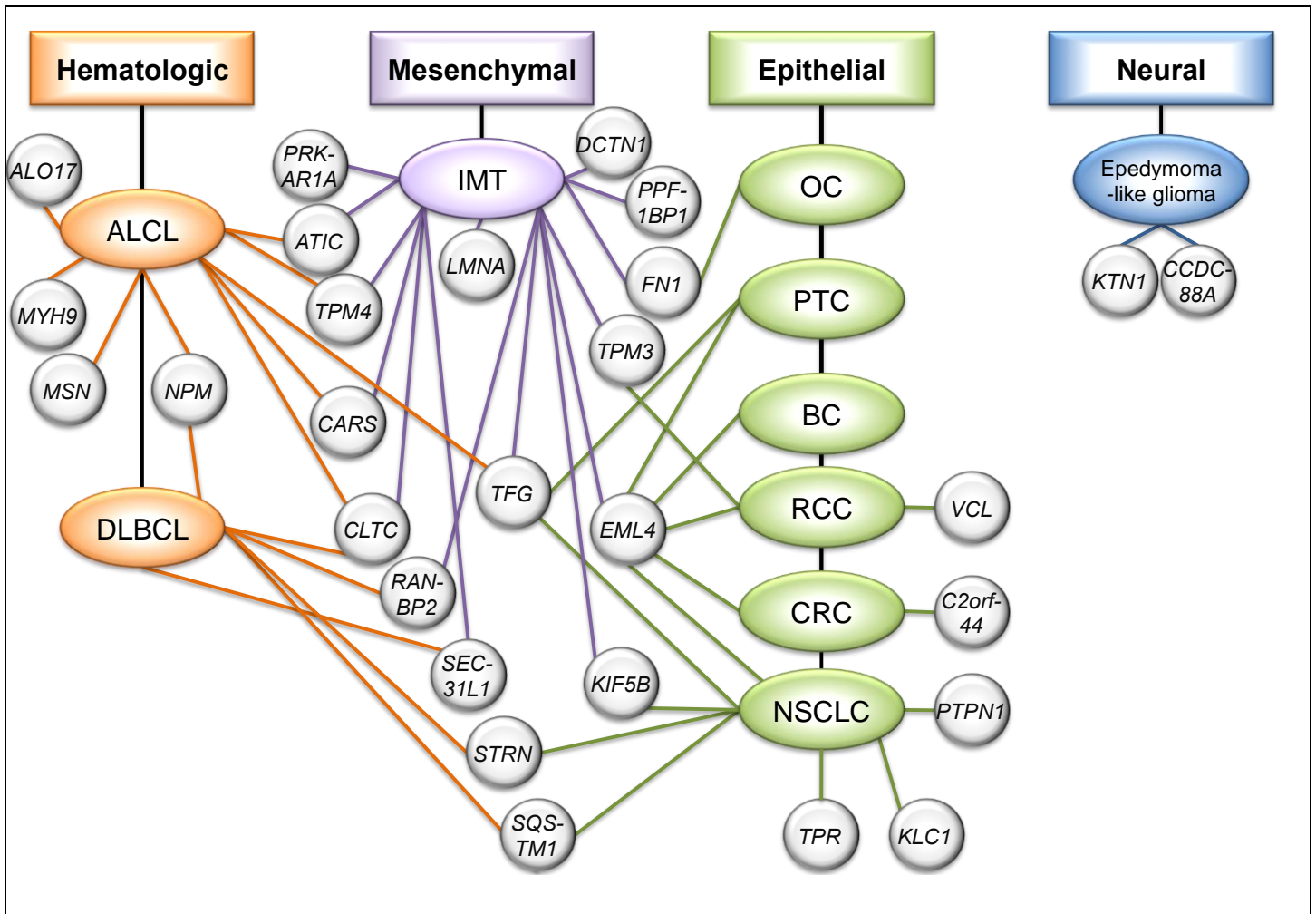


Figure 1.1: The Diversity of ALK fusions in Cancer. This diagram depicts the different tissues in which *ALK* rearranged cancers have been detected, the specific cancers within each tissue type, and the different *ALK* fusion partners that have been reported within each cancer type¹⁻⁴. Note: This diagram is not a comprehensive list of all *ALK* fusions or *ALK* fusion cancers.

(NSCLC), renal cell carcinoma (RCC), colorectal cancer (CRC), breast cancer, ovarian cancer, and esophageal cancer (**Figure 1.1**)^{3,43}. Though each malignancy has a particular 5' fusion partner that occurs most frequently, multiple different *ALK* fusions have been detected within the same disease. For example, *NPM-ALK* occurs in 80% of reported cases of *ALCL*^{77,78}), but *TFG-ALK*, *ATIC-ALK*, *CLTC-ALK*, *TPM3-ALK*, *MYH9-ALK*, *MSN-ALK*, and *ALO17-ALK* have also been reported⁴³(**Figure 1.1**). Interestingly, in all diseases, *ALK* positive patients tend to be younger than their *ALK* negative counterparts. However, the correlation between *ALK* positivity and prognostic outcome is disease dependent. *ALK* positive *DLBCL* and *NSCLC* tend to be more aggressive⁷⁹⁻⁸³. In *ALCL* and *IMT*, however, the *ALK* positive populations tend to have a better prognosis^{32,37}.

Disease	Response to ALK Targeted Therapy
Non-Small Cell Lung Cancer	+ (Phase 3 trial, FDA approved) ^{48,50}
Anaplastic Large Cell Lymphoma	+ (Phase 2/3 trial) ^{38,86}
Diffuse Large B Cell Lymphoma	+ (Case reports) ⁸⁷
Inflammatory Myofibroblastic Tumor	+ (Phase 1/2 trial) ^{38,86,88}
Spitz	ND
Renal Cell Carcinoma	+(phase 1 trial ongoing) ⁸⁹
Colorectal Cancer	+(phase 1 trial ongoing) ⁸⁹
Breast Cancer	ND
Ovarian Cancer	ND
Esophageal Cancer	ND
Papillary Thyroid Cancer	ND
Glioma	ND

Table 1.1: Response to ALK tyrosine kinase inhibitor therapy in ALK-rearranged tumors. This table summarizes the current clinical information known for patients with ALK-rearranged malignancies treated with ALK targeted therapies. + = a positive response. ND = not determined.

ALK TKIs have demonstrated great success in the clinic in certain ALK positive cancers, as discussed below^{84,85}. Unfortunately, however, the efficacy of ALK targeted therapy is still unknown for some tumor types harboring ALK fusions (**Table 1.1**). As better diagnostic tools enter the clinic, the identification of ALK positive cancers will likely continue to grow. Importantly, to better predict efficacy of ALK inhibition across diseases, we must improve our understanding of the biology of ALK fusions within each disease.

ALK Fusion Biology

Structure and Biological Function

In membrane spanning tyrosine kinases, the kinase domain is located toward the C-terminus while the N-terminus contains the inhibitory and trans-membrane domains. One shared feature of ALK kinase fusions is a largely conserved breakpoint within the C terminus of the receptor^{67,90,91}, just upstream of the kinase domain, resulting in the loss of the inhibitory juxtamembrane domain and ligand independent activation⁴. For ALK fusions, the genomic breakpoints have been found in the introns preceding exons 19 or 20, making exon 19 or 20 is the first exon of the ALK kinase domain^{2,4}.

Another characteristic of kinase fusions is the occurrence of multiple fusion partners within the same disease². Although these partners can vary, they share three features. First, the regulatory unit of the 5' fusion partner dictates the expression of the fusion, which may contribute to the level of active kinase^{92,93}. Secondly, most fusion partners^{92,93} contribute an oligomerization domain, which can aid in auto-activation of the kinase⁹⁴; although, this has not been verified for all fusion partners. The most

common oligomerization domains found in the fusion partners are coiled-coil domains (**Figure 1.2**)⁹⁵. For example, EML4-ALK, the most common ALK fusion detected in NSCLC, homodimerizes by virtue of a coiled-coil domain in EML4. Disruption of the coiled-coil domain abrogates the ability of EML4-ALK to transform cells⁹⁴. Furthermore, the extent of oligomerization may be important to transformation; other kinase fusions have been found to dimerize⁹⁶, trimerize^{97,98}, form tetramers⁹⁹, or require polymeric binding for transformation¹⁰⁰. Lastly, the 5' fusion partner also determines subcellular localization of the fusion, and this may have profound effects on the protein interactions that the fusion encounters, affecting activation, signaling, function, and degradation of the fusion. For example, a thorough analysis of each EML4-ALK variant revealed differences in their function and localization and implicated these properties in the variable sensitivity to Hsp90 inhibitors seen clinically⁹⁸. Additionally, for some fusions, subcellular localization controls fusion activation, as is the case for CLTC-ALK and MSN-ALK which congregate in clathrin coated vesicles and the plasma membrane respectively^{101,102}. While most ALK fusions appear pan-cytoplasmic, others like RANBP2-ALK and NPM-ALK have different localization, the effects of which have yet to be investigated (**Figure 1.2**)⁶⁷.

Some cancers may also exhibit effects from the loss of one wild-type allele of the partner gene. For example, *PRKAR1A* and *TPM3* are tumor suppressors, and as has been demonstrated for *KANK1* and *ETV6*, the remaining wild-type allele may be lost as a result of additional genetic or epigenetic alterations, resulting in no normal expression of the protein¹⁰³⁻¹⁰⁵. This has yet to be investigated for *PRKAR1A*. Additionally, if the remaining wild-type allele of the partner gene inhibits the fusion function, loss of the normal allele may exacerbate the disease as demonstrated for *TPM-ALK*^{106,107}. These kinds of interactions between the fusion and the WT partner protein have best been investigated in the case of NPM-ALK where studies have shown this interaction to affect nuclear translocation, activation, and response to ALK tyrosine kinase inhibitors¹⁰⁸⁻¹¹⁰.

Signaling

How different ALK fusions signal based on tissue of origin is only beginning to be elucidated. Of all known ALK fusions, NPM-ALK is the most investigated, since this was the first ALK fusion identified. Countless protein interactions involving NPM-ALK have been identified including those critical for transformation and migration, like activation of PLC γ , SRC, PI3K-AKT, STAT3, and SHH pathways^{67,111-114}, and less important interactions such as IRS-1 and SHC, which do not seem to be required for transformation⁵⁹. While NPM-ALK in has been found to activate PLC γ , PI3K-AKT, and STAT3 pathways, STAT3 activation seems to be the primary oncogenic mechanism in ALCL^{112,115-117}.

Figure 1.2

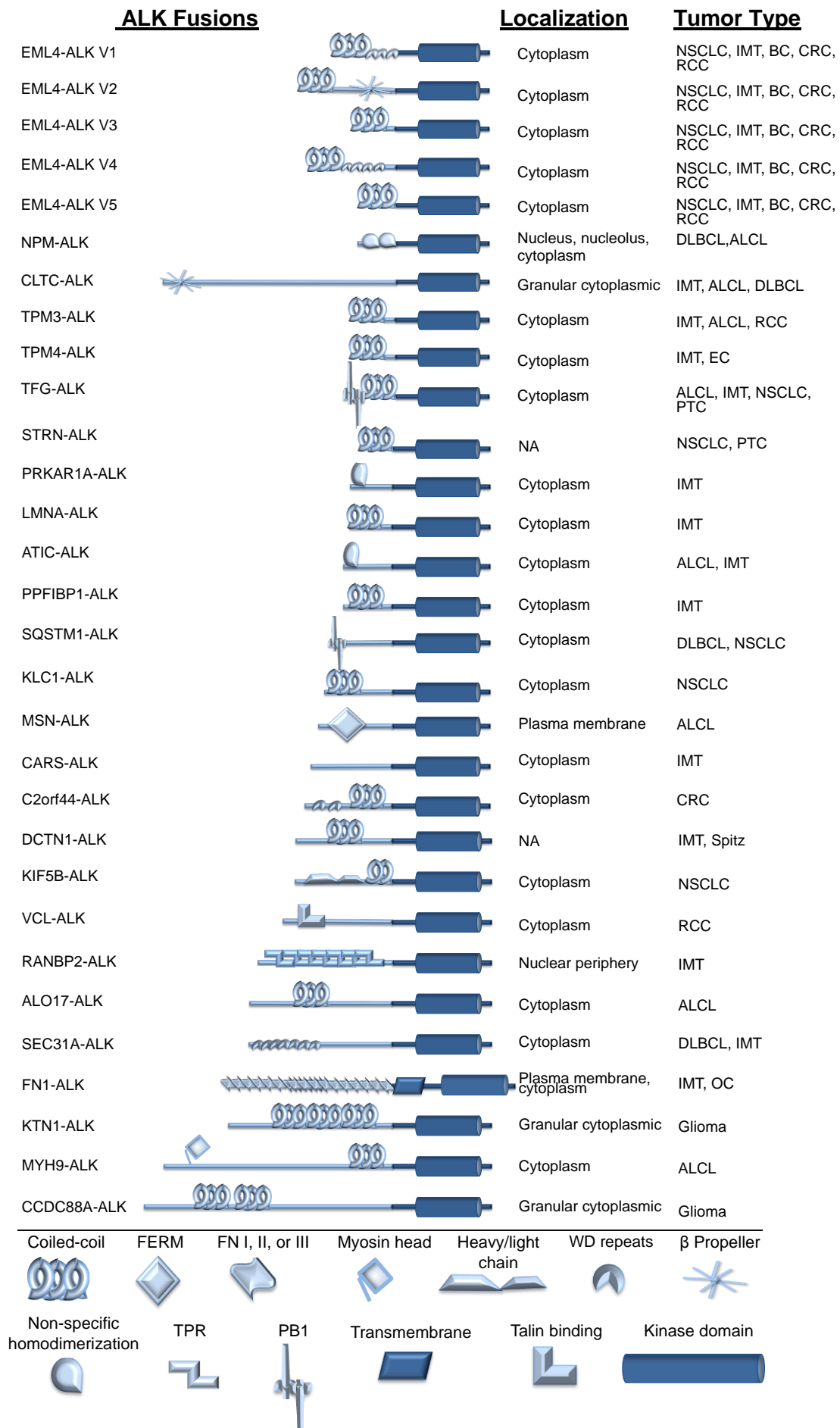


Figure 1.2. Structural and Biological Characteristics of ALK Fusions. This figure presents the majority of ALK fusions reported to date including the oligomerization domains present in the 5' fusion partner, the cellular localization of the fusion, and the cancers in which they have been detected to date. *Note that not all oligomerization domains have been experimentally validated.

Additionally, characterization of CLTC-ALK, the most commonly found ALK fusion in DLBCL, displayed increased STAT3 activity⁸¹. Further studies into EML4-ALK, the most commonly found ALK fusion in NSCLC, show similarities to NPM-ALK. In lung cancer cell lines and patient tumor samples, the MAP kinase, JAK-STAT, and PI3K-AKT pathways are all downstream of the EML4-ALK fusion proteins¹¹⁸. These studies, however, are investigating different ALK fusions. Whether or not the signaling networks of one particular ALK fusion changes with tissue type is a question yet to be explored.

Properties that may affect signaling between the various ALK fusions within the same disease are also not clear. For example, in IMT, it is unknown if the difference in cellular localization of RANBP2-ALK, which is at the periphery of the nucleus, compared to TFG-ALK, which is pan-cytoplasmic, has an effect on downstream signaling. Fusion partners found in ALCL are structurally and functionally different, yet their ALK protein interactions are similar^{90,119,120}. Signaling studies of ALCL fusions, however, reveal differences in pathway dependencies¹⁰⁶. Most studies focused on these questions, however, fail to study or validate the fusions in the corresponding cellular context making it difficult to draw clear conclusions or know the clinical relevance.

Therapeutic Targeting of ALK Fusions

ALK fusions have become the paradigm for development of targeted therapies since the time from discovery of the fusion in NSCLC, in 2007, to the initial FDA approval of the first ALK TKI, crizotinib, in 2011, showed the remarkable pace of bringing scientific discovery to the clinic. The phase I study of crizotinib in NSCLC was one of the first studies in which the biomarker, the *ALK* rearrangement, was identified prospectively before patients were enrolled onto the trial. In this study, the response rate to crizotinib in patients with heavily pre-treated ALK+ NSCLC was 60%; compared to only ~5% with traditional second- or third-line cytotoxic chemotherapy⁸⁴. After these initial Phase I results, a subsequent trial compared crizotinib in the first-line setting to standard chemotherapy for advanced NSCLC and showed a remarkable 55% increase in progression free survival over chemotherapy¹²¹. Thereafter, similar results have been published for patients with ALK+ ALCL and ALK+ IMT treated with crizotinib, overall demonstrating that ALK is an important therapeutic target in these

malignancies^{38,88}. Since the approval of crizotinib, four additional ALK TKIs (ceritinib, alectinib, brigatinib, and lorlatinib) with greater potency toward ALK have been granted FDA approval for use in ALK positive NSCLC^{45,48-50}.

Despite these promising results, we are still faced with difficulties in ALK detection and acquired drug resistance. The most widely used tools to detect ALK fusions in a patient's tumor are immunohistochemistry (IHC) and fluorescence *in situ* hybridization (FISH). All diagnostic tools have limitations; consequently, other forms of identifying ALK fusions in tumors, such as next generation sequencing methods, are increasingly being implemented¹²². Therapeutic drug resistance is another major barrier to overcome to improve outcomes for patients with ALK positive tumors¹²³. Mechanisms of acquired resistance to crizotinib are only beginning to be uncovered. These resistance mechanisms include: amplification of the *ALK* fusion, acquired mutations in the ALK kinase domain, up-regulation of 'bypass' signaling pathways to evade the drug-inhibited ALK fusion, and epithelial-to-mesenchymal transition. The most promising strategy to overcome crizotinib resistance to date involves the use of 'second or third-generation' ALK TKIs, such as ceritinib, alectinib, ensartinib, brigatinib, and lorlatinib, all more potent and more specific for the target (ALK). These drugs have proven to be effective against many ALK mutations that mediate crizotinib resistance¹²⁴. Strategies to combat ALK TKI resistance mediated by 'bypass' signaling pathways must include the use of combination therapies. We can also learn from the treatment of Acute Promyelocytic Leukemia, APLM, and investigate targeting the fusion partner as another way to combat resistance¹²⁵. Further studies like Richards et al. 2015, are needed to clarify and predict drug action as well as reveal avenues to block ALK activation by way of inhibiting the fusion partner.

1.4 Concluding Remarks

The cause of chromosomal rearrangements is a subject that has been investigated by many. The direct link between environmental factors such as ionizing radiation and chromosomal breaks in thyroid cancer has been well established while causes of other rearrangements remain less well known¹²⁵⁻¹²⁷. Interestingly, younger age may be a risk factor for some TK fusions such as ALK, ROS1, and RET, which occur more frequently in younger patients^{35,128-131}. Though the exact molecular mechanisms are unknown, the recurrence of similar fusions across many different cancers suggests a shared molecular deficiency. Along that line, some hypothesize that breaks are occurring at fragile DNA locus^{132,133}. Additionally, the occurrence and oncogenic potential of a particular kinase fusion seems to be tissue and cell-type dependent. In the case of CCDC6-RET, found in higher frequency in thyroid cancer than in breast cancer, *CCDC6* and *RET* were shown to be adjacent in the

nucleus of normal thyroid cells considerably more than in normal breast cells. Therefore, proximity may explain the higher frequency in thyroid cancer. However, while *CCDC6* and *RET* are juxtaposed in the nucleus of lymphoid cells just as frequently as in thyroid cells, *CCDC6-RET* is not seen in any lymphoid malignancies, and *CCDC6-RET* transgenic mice develop thyroid rather than lymphoid tumors^{134,135}. Cell-type dependencies are also seen for the fusion *BCR-ABL* where CML only occurs if the fusion is expressed in hematopoietic stem cells^{136,137}. These findings highlight the importance of the cellular context in which fusions occur. In-depth contextual studies are lacking for *ALK* fusions, and further investigation into the setting in which they display oncogenic activity is warranted. Understanding the background in which these fusions are oncogenic will allow us to more effectively inhibit the cancer and understand potential resistance mechanisms.

1.5 Purpose of These Studies

Beyond the presence or absence of a kinase fusion, nothing else is known about IMT on the genomic and transcriptomic level. The aim of these studies is to further reveal the genomic and transcriptomic landscape of IMTs and combine the data with the current histologic knowledge of the disease to gain insight into the pathogenesis or potential new therapies for patients with kinase fusion negative tumors or kinase fusion positive tumors which progress on targeted therapy.

While *ALK* fusions are validated targets for cancer therapy, our understanding of how to most effectively target these fusions clinically is lacking in comparison to other oncogenic drivers like mutant *EGFR* or *RAR α* fusions, where we know that different variants dictate drug sensitivity^{124, 125, 138-141}. For example, in *EGFR*-mutant NSCLC, we know that different *EGFR* alterations confer varying degrees of sensitivity or resistance to *EGFR* directed therapies¹³⁹⁻¹⁴². In addition, for other oncogenic fusions, such as the retinoic acid receptor alpha (*RAR α*) fusions found in APL, it is known that the particular gene fused to *RAR α* not only affects response to *RAR α* directed therapy but also can be a therapeutic target itself¹²⁵. Although many different 5' partner genes have been reported for *ALK*, even within the same disease, currently, there is very little data to address the question of how a different fusion partner may affect pretreatment clinical characteristics, disease responsiveness to targeted therapies, or acquired resistance. There are limited cell line models in which to study the various *ALK* fusions that occur in one particular disease; consequently, the role of the 5' fusion partners has not been systematically investigated. As more sophisticated next-generation sequencing technologies come to the forefront of clinical diagnostics, clinicians will know both the 5' partner and the 3' kinase involved in the fusion. Therefore, it is imperative that we determine the therapeutic implications of the various *ALK* fusions. The goal of these studies is to provide the specific pre-clinical

data needed to more precisely direct clinical treatment as well as potential innovative treatment strategies for ALK positive cancers such as IMT. In addition, these studies may have potential implications for other kinase fusion-driven malignancies.

1.6 References

1. Davare MA, Tognon CE: Detecting and targetting oncogenic fusion proteins in the genomic era. *Biol Cell* 107:111-29, 2015
2. Lovly CM, Gupta A, Lipson D, et al: Inflammatory myofibroblastic tumors harbor multiple potentially actionable kinase fusions. *Cancer Discov* 4:889-95, 2014
3. Olsen TK, Panagopoulos I, Meling TR, et al: Fusion genes with ALK as recurrent partner in ependymoma-like gliomas: a new brain tumor entity? *Neuro Oncol*, 2015
4. Shaw AT, Hsu PP, Awad MM, et al: Tyrosine kinase gene rearrangements in epithelial malignancies. *Nat Rev Cancer* 13:772-87, 2013
5. Coffin CM, Watterson J, Priest JR, et al: Extrapulmonary inflammatory myofibroblastic tumor (inflammatory pseudotumor). A clinicopathologic and immunohistochemical study of 84 cases. *Am J Surg Pathol* 19:859-72, 1995
6. Meis JM, Enzinger FM: Inflammatory fibrosarcoma of the mesentery and retroperitoneum. A tumor closely simulating inflammatory pseudotumor. *Am J Surg Pathol* 15:1146-56, 1991
7. Gonzalez-Crussi F, deMello DE, Sotelo-Avila C: Omental-mesenteric myxoid hamartomas. Infantile lesions simulating malignant tumors. *Am J Surg Pathol* 7:567-78, 1983
8. Dishop MK, Warner BW, Dehner LP, et al: Successful treatment of inflammatory myofibroblastic tumor with malignant transformation by surgical resection and chemotherapy. *J Pediatr Hematol Oncol* 25:153-8, 2003
9. Germanidis G, Xanthakis I, Tsitouridis I, et al: Regression of inflammatory myofibroblastic tumor of the gastrointestinal tract under infliximab treatment. *Dig Dis Sci* 50:262-5, 2005
10. Su W, Ko A, O'Connell T, et al: Treatment of pseudotumors with nonsteroidal antiinflammatory drugs. *J Pediatr Surg* 35:1635-7, 2000
11. Williams ME, Longmaid HE, Trey G, et al: Renal failure resulting from infiltration by inflammatory myofibroblastic tumor responsive to corticosteroid therapy. *Am J Kidney Dis* 31:E5, 1998
12. Day DL, Sane S, Dehner LP: Inflammatory pseudotumor of the mesentery and small intestine. *Pediatr Radiol* 16:210-5, 1986
13. Souid AK, Ziemba MC, Dubansky AS, et al: Inflammatory myofibroblastic tumor in children. *Cancer* 72:2042-8, 1993
14. Pettinato G, Manivel JC, De Rosa N, et al: Inflammatory myofibroblastic tumor (plasma cell granuloma). Clinicopathologic study of 20 cases with immunohistochemical and ultrastructural observations. *Am J Clin Pathol* 94:538-46, 1990

15. Cerfolio RJ, Allen MS, Nascimento AG, et al: Inflammatory pseudotumors of the lung. *Ann Thorac Surg* 67:933-6, 1999
16. Janik JS, Janik JP, Lovell MA, et al: Recurrent inflammatory pseudotumors in children. *J Pediatr Surg* 38:1491-5, 2003
17. Coffin CM, Hornick JL, Fletcher CD: Inflammatory myofibroblastic tumor: comparison of clinicopathologic, histologic, and immunohistochemical features including ALK expression in atypical and aggressive cases. *Am J Surg Pathol* 31:509-20, 2007
18. Gleason BC, Hornick JL: Inflammatory myofibroblastic tumours: where are we now? *J Clin Pathol* 61:428-37, 2008
19. Dishop MK, Kuruville S: Primary and metastatic lung tumors in the pediatric population: a review and 25-year experience at a large children's hospital. *Arch Pathol Lab Med* 132:1079-103, 2008
20. Fletcher CD: The evolving classification of soft tissue tumours: an update based on the new WHO classification. *Histopathology* 48:3-12, 2006
21. Riddle NN, Gardner JM: The New Kids on the Block: Recently Characterized Soft Tissue Tumors. *Surg Pathol Clin* 8:467-91, 2015
22. Chen ST, Lee JC: An inflammatory myofibroblastic tumor in liver with ALK and RANBP2 gene rearrangement: combination of distinct morphologic, immunohistochemical, and genetic features. *Hum Pathol* 39:1854-8, 2008
23. Marino-Enriquez A, Wang WL, Roy A, et al: Epithelioid inflammatory myofibroblastic sarcoma: An aggressive intra-abdominal variant of inflammatory myofibroblastic tumor with nuclear membrane or perinuclear ALK. *Am J Surg Pathol* 35:135-44, 2011
24. Lee JC, Wu JM, Liao JY, et al: Cytopathologic features of epithelioid inflammatory myofibroblastic sarcoma with correlation of histopathology, immunohistochemistry, and molecular cytogenetic analysis. *Cancer Cytopathol* 123:495-504, 2015
25. Hussong JW, Brown M, Perkins SL, et al: Comparison of DNA ploidy, histologic, and immunohistochemical findings with clinical outcome in inflammatory myofibroblastic tumors. *Mod Pathol* 12:279-86, 1999
26. Ramachandra S, Hollowood K, Bisceglia M, et al: Inflammatory pseudotumour of soft tissues: a clinicopathological and immunohistochemical analysis of 18 cases. *Histopathology* 27:313-23, 1995
27. Warter A, Satge D, Roeslin N: Angioinvasive plasma cell granulomas of the lung. *Cancer* 59:435-43, 1987
28. Yamamoto H, Oda Y, Saito T, et al: p53 Mutation and MDM2 amplification in inflammatory myofibroblastic tumours. *Histopathology* 42:431-9, 2003

29. Donner LR, Trompler RA, White RR: Progression of inflammatory myofibroblastic tumor (inflammatory pseudotumor) of soft tissue into sarcoma after several recurrences. *Hum Pathol* 27:1095-8, 1996
30. Spencer H: The pulmonary plasma cell/histiocytoma complex. *Histopathology* 8:903-16, 1984
31. Meis-Kindblom JM, Kjellstrom C, Kindblom LG: Inflammatory fibrosarcoma: update, reappraisal, and perspective on its place in the spectrum of inflammatory myofibroblastic tumors. *Semin Diagn Pathol* 15:133-43, 1998
32. Coffin CM, Patel A, Perkins S, et al: ALK1 and p80 expression and chromosomal rearrangements involving 2p23 in inflammatory myofibroblastic tumor. *Mod Pathol* 14:569-76, 2001
33. Griffin CA, Hawkins AL, Dvorak C, et al: Recurrent involvement of 2p23 in inflammatory myofibroblastic tumors. *Cancer Res* 59:2776-80, 1999
34. Takeuchi K, Soda M, Togashi Y, et al: Identification of a novel fusion, SQSTM1-ALK, in ALK-positive large B-cell lymphoma. *Haematologica* 96:464-7, 2011
35. Takeuchi K, Soda M, Togashi Y, et al: RET, ROS1 and ALK fusions in lung cancer. *Nat Med* 18:378-81, 2012
36. Chun YS, Wang L, Nascimento AG, et al: Pediatric inflammatory myofibroblastic tumor: anaplastic lymphoma kinase (ALK) expression and prognosis. *Pediatr Blood Cancer* 45:796-801, 2005
37. Cook JR, Dehner LP, Collins MH, et al: Anaplastic lymphoma kinase (ALK) expression in the inflammatory myofibroblastic tumor: a comparative immunohistochemical study. *Am J Surg Pathol* 25:1364-71, 2001
38. Mosse YP, Voss SD, Lim MS, et al: Targeting ALK With Crizotinib in Pediatric Anaplastic Large Cell Lymphoma and Inflammatory Myofibroblastic Tumor: A Children's Oncology Group Study. *J Clin Oncol* 35:3215-3221, 2017
39. Antonescu CR, Suurmeijer AJ, Zhang L, et al: Molecular characterization of inflammatory myofibroblastic tumors with frequent ALK and ROS1 gene fusions and rare novel RET rearrangement. *Am J Surg Pathol* 39:957-67, 2015
40. Lee JC, Li CF, Huang HY, et al: ALK oncoproteins in atypical inflammatory myofibroblastic tumours: novel RRBP1-ALK fusions in epithelioid inflammatory myofibroblastic sarcoma. *J Pathol* 241:316-323, 2017
41. Cha YJ, Shim HS: PD-L1 expression and CD8+ tumor-infiltrating lymphocytes are associated with ALK rearrangement and clinicopathological features in inflammatory myofibroblastic tumors. *Oncotarget* 8:89465-89474, 2017
42. Nowell PC, Hungerford DA: Chromosome studies on normal and leukemic human leukocytes. *J Natl Cancer Inst* 25:85-109, 1960

43. Hallberg B, Palmer RH: Mechanistic insight into ALK receptor tyrosine kinase in human cancer biology. *Nat Rev Cancer* 13:685-700, 2013
44. Camidge DR, Bang YJ, Kwak EL, et al: Activity and safety of crizotinib in patients with ALK-positive non-small-cell lung cancer: updated results from a phase 1 study. *Lancet Oncol* 13:1011-9, 2012
45. Gettinger SN, Bazhenova LA, Langer CJ, et al: Activity and safety of brigatinib in ALK-rearranged non-small-cell lung cancer and other malignancies: a single-arm, open-label, phase 1/2 trial. *Lancet Oncol* 17:1683-1696, 2016
46. Horn L, Infante JR, Reckamp KL, et al: Ensartinib (X-396) in ALK-positive Non-Small Cell Lung Cancer: Results from a First-in-Human Phase I/II, Multicenter Study. *Clin Cancer Res*, 2018
47. Kim DW, Tiseo M, Ahn MJ, et al: Brigatinib in Patients With Crizotinib-Refractory Anaplastic Lymphoma Kinase-Positive Non-Small-Cell Lung Cancer: A Randomized, Multicenter Phase II Trial. *J Clin Oncol* 35:2490-2498, 2017
48. Peters S, Camidge DR, Shaw AT, et al: Alectinib versus Crizotinib in Untreated ALK-Positive Non-Small-Cell Lung Cancer. *N Engl J Med* 377:829-838, 2017
49. Shaw AT, Felip E, Bauer TM, et al: Lorlatinib in non-small-cell lung cancer with ALK or ROS1 rearrangement: an international, multicentre, open-label, single-arm first-in-man phase 1 trial. *Lancet Oncol* 18:1590-1599, 2017
50. Soria JC, Tan DSW, Chiari R, et al: First-line ceritinib versus platinum-based chemotherapy in advanced ALK-rearranged non-small-cell lung cancer (ASCEND-4): a randomised, open-label, phase 3 study. *Lancet* 389:917-929, 2017
51. Kadomatsu K, Muramatsu T: Midkine and pleiotrophin in neural development and cancer. *Cancer Lett* 204:127-43, 2004
52. Muramatsu T: Midkine and pleiotrophin: two related proteins involved in development, survival, inflammation and tumorigenesis. *J Biochem* 132:359-71, 2002
53. Stoica GE, Kuo A, Aigner A, et al: Identification of anaplastic lymphoma kinase as a receptor for the growth factor pleiotrophin. *J Biol Chem* 276:16772-9, 2001
54. Stoica GE, Kuo A, Powers C, et al: Midkine binds to anaplastic lymphoma kinase (ALK) and acts as a growth factor for different cell types. *J Biol Chem* 277:35990-8, 2002
55. Mathivet T, Mazot P, Vigny M: In contrast to agonist monoclonal antibodies, both C-terminal truncated form and full length form of Pleiotrophin failed to activate vertebrate ALK (anaplastic lymphoma kinase)? *Cell Signal* 19:2434-43, 2007
56. Moog-Lutz C, Degoutin J, Gouzi JY, et al: Activation and inhibition of anaplastic lymphoma kinase receptor tyrosine kinase by monoclonal antibodies and absence of agonist activity of pleiotrophin. *J Biol Chem* 280:26039-48, 2005
57. Motegi A, Fujimoto J, Kotani M, et al: ALK receptor tyrosine kinase promotes cell growth and neurite outgrowth. *J Cell Sci* 117:3319-29, 2004

58. Murali J, Benard A, Lourenco FC, et al: Anaplastic lymphoma kinase is a dependence receptor whose proapoptotic functions are activated by caspase cleavage. *Mol Cell Biol* 26:6209-22, 2006
59. Iwahara T, Fujimoto J, Wen D, et al: Molecular characterization of ALK, a receptor tyrosine kinase expressed specifically in the nervous system. *Oncogene* 14:439-49, 1997
60. Pulford K, Lamant L, Morris SW, et al: Detection of anaplastic lymphoma kinase (ALK) and nucleolar protein nucleophosmin (NPM)-ALK proteins in normal and neoplastic cells with the monoclonal antibody ALK1. *Blood* 89:1394-404, 1997
61. Bilsland JG, Wheeldon A, Mead A, et al: Behavioral and neurochemical alterations in mice deficient in anaplastic lymphoma kinase suggest therapeutic potential for psychiatric indications. *Neuropsychopharmacology* 33:685-700, 2008
62. Lasek AW, Lim J, Kliethermes CL, et al: An evolutionary conserved role for anaplastic lymphoma kinase in behavioral responses to ethanol. *PLoS One* 6:e22636, 2011
63. Weiss JB, Xue C, Benice T, et al: Anaplastic lymphoma kinase and leukocyte tyrosine kinase: functions and genetic interactions in learning, memory and adult neurogenesis. *Pharmacol Biochem Behav* 100:566-74, 2012
64. Mosse YP, Wood A, Maris JM: Inhibition of ALK signaling for cancer therapy. *Clin Cancer Res* 15:5609-14, 2009
65. Palmer RH, Vernersson E, Grabbe C, et al: Anaplastic lymphoma kinase: signalling in development and disease. *Biochem J* 420:345-61, 2009
66. Barreca A, Lasorsa E, Riera L, et al: Anaplastic lymphoma kinase in human cancer. *J Mol Endocrinol* 47:R11-23, 2011
67. Chiarle R, Voena C, Ambrogio C, et al: The anaplastic lymphoma kinase in the pathogenesis of cancer. *Nat Rev Cancer* 8:11-23, 2008
68. Ambrogio C, Voena C, Manazza AD, et al: The anaplastic lymphoma kinase controls cell shape and growth of anaplastic large cell lymphoma through Cdc42 activation. *Cancer Res* 68:8899-907, 2008
69. Ambrogio C, Voena C, Manazza AD, et al: p130Cas mediates the transforming properties of the anaplastic lymphoma kinase. *Blood* 106:3907-16, 2005
70. Bassermann F, von Klitzing C, Munch S, et al: NIPA defines an SCF-type mammalian E3 ligase that regulates mitotic entry. *Cell* 122:45-57, 2005
71. Colomba A, Courilleau D, Ramel D, et al: Activation of Rac1 and the exchange factor Vav3 are involved in NPM-ALK signaling in anaplastic large cell lymphomas. *Oncogene* 27:2728-36, 2008

72. Cussac D, Greenland C, Roche S, et al: Nucleophosmin-anaplastic lymphoma kinase of anaplastic large-cell lymphoma recruits, activates, and uses pp60c-src to mediate its mitogenicity. *Blood* 103:1464-71, 2004
73. Dupuis-Coronas S, Lagarrigue F, Ramel D, et al: The nucleophosmin-anaplastic lymphoma kinase oncogene interacts, activates, and uses the kinase PIKfyve to increase invasiveness. *J Biol Chem* 286:32105-14, 2011
74. Hegazy SA, Wang P, Anand M, et al: The tyrosine 343 residue of nucleophosmin (NPM)-anaplastic lymphoma kinase (ALK) is important for its interaction with SHP1, a cytoplasmic tyrosine phosphatase with tumor suppressor functions. *J Biol Chem* 285:19813-20, 2010
75. Ouyang T, Bai RY, Bassermann F, et al: Identification and characterization of a nuclear interacting partner of anaplastic lymphoma kinase (NIPA). *J Biol Chem* 278:30028-36, 2003
76. Voena C, Conte C, Ambrogio C, et al: The tyrosine phosphatase Shp2 interacts with NPM-ALK and regulates anaplastic lymphoma cell growth and migration. *Cancer Res* 67:4278-86, 2007
77. Amin HM, Lai R: Pathobiology of ALK+ anaplastic large-cell lymphoma. *Blood* 110:2259-67, 2007
78. Stein H, Foss HD, Durkop H, et al: CD30(+) anaplastic large cell lymphoma: a review of its histopathologic, genetic, and clinical features. *Blood* 96:3681-95, 2000
79. Choung HS, Kim HJ, Kim WS, et al: [Cytomorphology and molecular characterization of CLTC-ALK rearrangement in 2 cases of ALK-positive diffuse large B-cell lymphoma with extensive bone marrow involvement]. *Korean J Lab Med* 28:89-94, 2008
80. Lee HW, Kim K, Kim W, et al: ALK-positive diffuse large B-cell lymphoma: report of three cases. *Hematol Oncol* 26:108-13, 2008
81. Momose S, Tamaru J, Kishi H, et al: Hyperactivated STAT3 in ALK-positive diffuse large B-cell lymphoma with clathrin-ALK fusion. *Hum Pathol* 40:75-82, 2009
82. Reichard KK, McKenna RW, Kroft SH: ALK-positive diffuse large B-cell lymphoma: report of four cases and review of the literature. *Mod Pathol* 20:310-9, 2007
83. Stachurski D, Miron PM, Al-Homsi S, et al: Anaplastic lymphoma kinase-positive diffuse large B-cell lymphoma with a complex karyotype and cryptic 3' ALK gene insertion to chromosome 4 q22-24. *Hum Pathol* 38:940-5, 2007
84. Kwak EL, Bang YJ, Camidge DR, et al: Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N Engl J Med* 363:1693-703, 2010
85. Mosse YP, Lim MS, Voss SD, et al: Safety and activity of crizotinib for paediatric patients with refractory solid tumours or anaplastic large-cell lymphoma: a Children's Oncology Group phase 1 consortium study. *Lancet Oncol* 14:472-80, 2013
86. Balis FM, Thompson PA, Mosse YP, et al: First-dose and steady-state pharmacokinetics of orally administered crizotinib in children with solid tumors: a report on ADVL0912 from the

Children's Oncology Group Phase 1/Pilot Consortium. *Cancer Chemother Pharmacol* 79:181-187, 2017

87. Li J, Ouyang J, Zhou R, et al: Promising response of anaplastic lymphoma kinase-positive large B-cell lymphoma to crizotinib salvage treatment: case report and review of literature. *Int J Clin Exp Med* 8:6977-85, 2015
88. Schoffski P, Sufliarsky J, Gelderblom H, et al: Crizotinib in patients with advanced, inoperable inflammatory myofibroblastic tumours with and without anaplastic lymphoma kinase gene alterations (European Organisation for Research and Treatment of Cancer 90101 CREATE): a multicentre, single-drug, prospective, non-randomised phase 2 trial. *Lancet Respir Med* 6:431-441, 2018
89. Drilon A, Siena S, Ou SI, et al: Safety and Antitumor Activity of the Multitargeted Pan-TRK, ROS1, and ALK Inhibitor Entrectinib: Combined Results from Two Phase I Trials (ALKA-372-001 and STARTRK-1). *Cancer Discov* 7:400-409, 2017
90. Hernandez L, Bea S, Bellosillo B, et al: Diversity of genomic breakpoints in TFG-ALK translocations in anaplastic large cell lymphomas: identification of a new TFG-ALK(XL) chimeric gene with transforming activity. *Am J Pathol* 160:1487-94, 2002
91. Ou SH, Bartlett CH, Mino-Kenudson M, et al: Crizotinib for the treatment of ALK-rearranged non-small cell lung cancer: a success story to usher in the second decade of molecular targeted therapy in oncology. *Oncologist* 17:1351-75, 2012
92. Seo JS, Ju YS, Lee WC, et al: The transcriptional landscape and mutational profile of lung adenocarcinoma. *Genome Res* 22:2109-19, 2012
93. Wu YM, Su F, Kalyana-Sundaram S, et al: Identification of targetable FGFR gene fusions in diverse cancers. *Cancer Discov* 3:636-47, 2013
94. Soda M, Choi YL, Enomoto M, et al: Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature* 448:561-6, 2007
95. Knight MJ, Leetola C, Gingery M, et al: A human sterile alpha motif domain polymerizome. *Protein Sci* 20:1697-706, 2011
96. Baumann H, Kunapuli P, Tracy E, et al: The oncogenic fusion protein-tyrosine kinase ZNF198/fibroblast growth factor receptor-1 has signaling function comparable with interleukin-6 cytokine receptors. *J Biol Chem* 278:16198-208, 2003
97. Medves S, Noel LA, Montano-Almendras CP, et al: Multiple oligomerization domains of KANK1-PDGFRbeta are required for JAK2-independent hematopoietic cell proliferation and signaling via STAT5 and ERK. *Haematologica* 96:1406-14, 2011
98. Richards MW, O'Regan L, Roth D, et al: Microtubule association of EML proteins and the EML4-ALK variant 3 oncoprotein require an N-terminal trimerization domain. *Biochem J* 467:529-36, 2015
99. Zhao X, Ghaffari S, Lodish H, et al: Structure of the Bcr-Abl oncoprotein oligomerization domain. *Nat Struct Biol* 9:117-20, 2002

100. Tognon CE, Mackereth CD, Somasiri AM, et al: Mutations in the SAM domain of the ETV6-NTRK3 chimeric tyrosine kinase block polymerization and transformation activity. *Mol Cell Biol* 24:4636-50, 2004
101. Lahortiga I, Akin C, Cools J, et al: Activity of imatinib in systemic mastocytosis with chronic basophilic leukemia and a PRKG2-PDGFRB fusion. *Haematologica* 93:49-56, 2008
102. Tort F, Pinyol M, Pulford K, et al: Molecular characterization of a new ALK translocation involving moesin (MSN-ALK) in anaplastic large cell lymphoma. *Lab Invest* 81:419-26, 2001
103. Griesinger F, Janke A, Podleschny M, et al: Identification of an ETV6-ABL2 fusion transcript in combination with an ETV6 point mutation in a T-cell acute lymphoblastic leukaemia cell line. *Br J Haematol* 119:454-8, 2002
104. Medves S, Duhoux FP, Ferrant A, et al: KANK1, a candidate tumor suppressor gene, is fused to PDGFRB in an imatinib-responsive myeloid neoplasm with severe thrombocythemia. *Leukemia* 24:1052-5, 2010
105. Vu HA, Xinh PT, Masuda M, et al: FLT3 is fused to ETV6 in a myeloproliferative disorder with hypereosinophilia and a t(12;13)(p13;q12) translocation. *Leukemia* 20:1414-21, 2006
106. Armstrong F, Duplantier MM, Trepapat P, et al: Differential effects of X-ALK fusion proteins on proliferation, transformation, and invasion properties of NIH3T3 cells. *Oncogene* 23:6071-82, 2004
107. Armstrong F, Lamant L, Hieblot C, et al: TPM3-ALK expression induces changes in cytoskeleton organisation and confers higher metastatic capacities than other ALK fusion proteins. *Eur J Cancer* 43:640-6, 2007
108. Bischof D, Pulford K, Mason DY, et al: Role of the nucleophosmin (NPM) portion of the non-Hodgkin's lymphoma-associated NPM-anaplastic lymphoma kinase fusion protein in oncogenesis. *Mol Cell Biol* 17:2312-25, 1997
109. Ceccon M, Merlo MEB, Mologni L, et al: Excess of NPM-ALK oncogenic signaling promotes cellular apoptosis and drug dependency. *Oncogene* 35:3854-3865, 2016
110. Mason DY, Pulford KA, Bischof D, et al: Nucleolar localization of the nucleophosmin-anaplastic lymphoma kinase is not required for malignant transformation. *Cancer Res* 58:1057-62, 1998
111. Bai RY, Ouyang T, Miething C, et al: Nucleophosmin-anaplastic lymphoma kinase associated with anaplastic large-cell lymphoma activates the phosphatidylinositol 3-kinase/Akt antiapoptotic signaling pathway. *Blood* 96:4319-27, 2000
112. Chiarle R, Simmons WJ, Cai H, et al: Stat3 is required for ALK-mediated lymphomagenesis and provides a possible therapeutic target. *Nat Med* 11:623-9, 2005
113. Polgar D, Leisser C, Maier S, et al: Truncated ALK derived from chromosomal translocation t(2;5)(p23;q35) binds to the SH3 domain of p85-PI3K. *Mutat Res* 570:9-15, 2005

114. Singh RR, Cho-Vega JH, Davuluri Y, et al: Sonic hedgehog signaling pathway is activated in ALK-positive anaplastic large cell lymphoma. *Cancer Res* 69:2550-8, 2009
115. Amin HM, McDonnell TJ, Ma Y, et al: Selective inhibition of STAT3 induces apoptosis and G(1) cell cycle arrest in ALK-positive anaplastic large cell lymphoma. *Oncogene* 23:5426-34, 2004
116. Zamo A, Chiarle R, Piva R, et al: Anaplastic lymphoma kinase (ALK) activates Stat3 and protects hematopoietic cells from cell death. *Oncogene* 21:1038-47, 2002
117. Zhang Q, Raghunath PN, Xue L, et al: Multilevel dysregulation of STAT3 activation in anaplastic lymphoma kinase-positive T/null-cell lymphoma. *J Immunol* 168:466-74, 2002
118. Takezawa K, Okamoto I, Nishio K, et al: Role of ERK-BIM and STAT3-survivin signaling pathways in ALK inhibitor-induced apoptosis in EML4-ALK-positive lung cancer. *Clin Cancer Res* 17:2140-8, 2011
119. Ma Z, Cools J, Marynen P, et al: Inv(2)(p23q35) in anaplastic large-cell lymphoma induces constitutive anaplastic lymphoma kinase (ALK) tyrosine kinase activation by fusion to ATIC, an enzyme involved in purine nucleotide biosynthesis. *Blood* 95:2144-9, 2000
120. Trinei M, Lanfrancone L, Campo E, et al: A new variant anaplastic lymphoma kinase (ALK)-fusion protein (ATIC-ALK) in a case of ALK-positive anaplastic large cell lymphoma. *Cancer Res* 60:793-8, 2000
121. Solomon BJ, Mok T, Kim DW, et al: First-line crizotinib versus chemotherapy in ALK-positive lung cancer. *N Engl J Med* 371:2167-77, 2014
122. Weickhardt AJ, Aisner DL, Franklin WA, et al: Diagnostic assays for identification of anaplastic lymphoma kinase-positive non-small cell lung cancer. *Cancer* 119:1467-77, 2013
123. Hallberg B, Palmer RH: ALK and NSCLC: Targeted therapy with ALK inhibitors. *F1000 Med Rep* 3:21, 2011
124. Gainor JF, Dardaei L, Yoda S, et al: Molecular Mechanisms of Resistance to First- and Second-Generation ALK Inhibitors in ALK-Rearranged Lung Cancer. *Cancer Discov* 6:1118-1133, 2016
125. Melnick A, Licht JD: Deconstructing a disease: RARalpha, its fusion partners, and their roles in the pathogenesis of acute promyelocytic leukemia. *Blood* 93:3167-215, 1999
126. Duffy BJ, Jr., Fitzgerald PJ: Cancer of the thyroid in children: a report of 28 cases. *J Clin Endocrinol Metab* 10:1296-1308, 1950
127. Ito T, Seyama T, Iwamoto KS, et al: In vitro irradiation is able to cause RET oncogene rearrangement. *Cancer Res* 53:2940-3, 1993
128. Kazakov VS, Demidchik EP, Astakhova LN: Thyroid cancer after Chernobyl. *Nature* 359:21, 1992
129. Bergethon K, Shaw AT, Ou SH, et al: ROS1 rearrangements define a unique molecular class of lung cancers. *J Clin Oncol* 30:863-70, 2012

130. Inamura K, Takeuchi K, Togashi Y, et al: EML4-ALK lung cancers are characterized by rare other mutations, a TTF-1 cell lineage, an acinar histology, and young onset. *Mod Pathol* 22:508-15, 2009
131. Shaw AT, Yeap BY, Mino-Kenudson M, et al: Clinical features and outcome of patients with non-small-cell lung cancer who harbor EML4-ALK. *J Clin Oncol* 27:4247-53, 2009
132. Wang R, Hu H, Pan Y, et al: RET fusions define a unique molecular and clinicopathologic subtype of non-small-cell lung cancer. *J Clin Oncol* 30:4352-9, 2012
133. Burrow AA, Williams LE, Pierce LC, et al: Over half of breakpoints in gene pairs involved in cancer-specific recurrent translocations are mapped to human chromosomal fragile sites. *BMC Genomics* 10:59, 2009
134. Gandhi M, Dillon LW, Pramanik S, et al: DNA breaks at fragile sites generate oncogenic RET/PTC rearrangements in human thyroid cells. *Oncogene* 29:2272-80, 2010
135. Nikiforova MN, Stringer JR, Blough R, et al: Proximity of chromosomal loci that participate in radiation-induced rearrangements in human cells. *Science* 290:138-41, 2000
136. Zhang Y, McCord RP, Ho YJ, et al: Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell* 148:908-21, 2012
137. Koschmieder S, Gottgens B, Zhang P, et al: Inducible chronic phase of myeloid leukemia with expansion of hematopoietic stem cells in a transgenic model of BCR-ABL leukemogenesis. *Blood* 105:324-34, 2005
138. Ren R: Mechanisms of BCR-ABL in the pathogenesis of chronic myelogenous leukaemia. *Nat Rev Cancer* 5:172-83, 2005
139. Ayeni D, Politi K, Goldberg SB: Emerging Agents and New Mutations in EGFR-Mutant Lung Cancer. *Clin Cancer Res* 21:3818-20, 2015
140. Gallant JN, Sheehan JH, Shaver TM, et al: EGFR Kinase Domain Duplication (EGFR-KDD) Is a Novel Oncogenic Driver in Lung Cancer That Is Clinically Responsive to Afatinib. *Cancer Discov* 5:1155-63, 2015
141. Meador CB, Jin H, de Stanchina E, et al: Optimizing the sequence of anti-EGFR-targeted therapy in EGFR-mutant lung cancer. *Mol Cancer Ther* 14:542-52, 2015
142. Wang Y, Li RQ, Ai YQ, et al: Exon 19 deletion was associated with better survival outcomes in advanced lung adenocarcinoma with mutant EGFR treated with EGFR-TKIs as second-line therapy after first-line chemotherapy: a retrospective analysis of 128 patients. *Clin Transl Oncol* 17:727-36, 2015

Chapter 2

INFLAMMATORY MYOFIBROBLASTIC TUMOR: A GENOMIC AND TRANSCRIPTOMIC ANALYSIS

2.1 Abstract

Inflammatory myofibroblastic tumor (IMT) is a rare, mesenchymal neoplasm occurring primarily in the soft tissue and viscera of children, adolescents, and young adults. Surgical resection is the current standard of care for IMT. There is no systemic standard of care therapy for patients whose tumors cannot be resected. ALK kinase fusions have been reported in 50-60% of IMT, and *ROS1*, *PDGFRB*, *RET*, and *NTRK3* fusions have also recently been detected in the ALK fusion negative cohort. Importantly, ALK fusions are validated as a therapeutic target in IMT, and there have also been reports of *ROS1* rearranged IMT responding to *ROS1* targeted therapy, demonstrating the importance of identifying kinase fusions in IMT. The genomic and transcriptomic landscape of IMT, outside of the presence or absence of oncogenic kinase fusions, remains largely unexplored; therefore, there is currently no rational to direct therapy, in particular, the use of targeted therapies for kinase fusion negative patients or kinase fusion positive patients who have progressed on targeted therapies. We hypothesized that further knowledge of the IMT transcriptome coupled with clinicopathologic data would provide insight into the pathogenesis of the disease and potential effective therapies. We performed targeted next generation sequencing (NGS) on 69 total samples and whole transcriptome RNA sequencing on 18 of the 69 samples. We identified kinase fusions in 84% of IMTs. Specific kinase fusions were not associated with anatomical location, age, or gender. Analysis of cancer associated SNVs revealed a set of 30 commonly shared variants involved in the VEGF pathway and chromatin remodeling. Three of four fusion negative tumors shared a mutation in *PIK3R1* previously reported in other cancers. These findings support the use of NGS diagnostics to select effective targeted therapeutic options for patients with IMT and reveal common pathway disruptions that could potentially be therapeutically targeted.

2.2 Introduction

Inflammatory myofibroblastic tumor (IMT) is a rare, mesenchymal neoplasm, with approximately 200 cases per year in the U.S. While it can occur at any age and in multiple anatomic locations, tumors arise primarily in the soft tissue and viscera of children, adolescents, and young adults¹⁻⁴. Surgical resection is the current standard of care for IMT. There is no systemic standard of care therapy for patients whose tumors cannot be resected. IMT has a tendency for local recurrence, most frequently

seen in patients presenting with multi-focal intra-abdominal tumors and those in anatomical locations where complete surgical resection is difficult. Distant metastasis of IMT is rare, affecting <5% of patients^{1,5}.

IMT is diagnosed by its pathology using a set of criteria set by the World Health Organization (WHO). The WHO defines IMT as “a lesion composed of a proliferation of myofibroblastic spindle and stellate cells with abundant eosinophilic cytoplasm mixed with infiltrative plasma, inflammatory cells, lymphocytes and eosinophils”^{6,7}. More recently a distinct subtype of IMT, known as epithelioid Inflammatory Myofibroblastic Sarcoma (eIMS), has been described. This subtype is characterized by a dominance of large epithelioid cells with only minor spindle cell components, noticeable nuclear atypia, and positive anaplastic lymphoma kinase (ALK) immunohistochemistry (IHC) in a perinuclear pattern⁸. eIMS has been shown to have a more clinically aggressive phenotype and dismal prognosis^{9,10}. Additionally, eIMS is almost exclusively found in the abdominal cavity and has a strong predominance in males¹¹.

Rearrangements involving *ALK* on chromosome 2p23 result in oncogenic fusion proteins and have been reported in 50-60% of IMTs¹²⁻¹⁴. Several *ALK* translocation partners have been identified in IMT, including *TPM-3/-4*, *ATIC*, *CLTC*, *CARS*, *FN1*, *PRKAR1A*, *EML4*, *RANBP2*, and many others¹⁴⁻¹⁶. EIMS are divergent in their specific expression of kinase fusions. 100% of eIMS cases reported to date harbor either *RANBP2-ALK* or *RRBP1-ALK* fusions with distinctive nuclear membranous ALK IHC or cytoplasmic combined with intense nuclear membranous ALK IHC respectively¹⁷. Recently *ROS1*, *PDGFR β* , *NTRK3*, and *RET* gene alterations have also been reported, bringing kinase fusion detection in IMT to approximately 85%^{14,18-20}. Importantly, ALK fusions are validated as a therapeutic target in IMT^{21,22}. There have also been reports of *ROS1* rearranged IMT responding to *ROS1* targeted therapy¹⁴, demonstrating the importance of identifying kinase fusions in ALK negative IMT. ALK fusion negative IMTs typically occur in older patients and have a more aggressive histological appearance^{5,19}. In addition, the frequency of metastasis appears to be greater in ALK negative IMT compared to ALK positive IMT, with the exception of eIMS.

The genomic and transcriptomic landscape of IMT, outside of the presence or absence of oncogenic kinase fusions, remains largely unexplored; therefore, there is currently no rational to direct therapy, in particular, the use of targeted therapies for kinase fusion negative patients or kinase fusion positive patients who have progressed on targeted therapies. The goal of this study is to further reveal the genomic and transcriptomic landscape of IMTs to gain insight into the pathogenesis of the disease

and potentially offer new therapeutic opportunities.

2.3 Materials and Methods

Patient Tumor Samples: IMT samples and associated patient characteristics were obtained and analyzed with an Institutional Review Board (IRB)-approved protocol (#090572). All clinical data was obtained and maintained according to HIPAA standards. All unique identifiers have been removed prior to publication.

Targeted NGS for Fusion Identification

IMT cases are described in Table 2.1.

IMT case numbers 6-12,15,17-25, 39, 61-63, 65, and 66: Following the manufacturer's protocol, RNA was extracted from formalin-fixed and paraffin embedded (FFPE) and fresh frozen (FF) samples using the Maxwell 16 LEV RNA FFPE Kit (Promega) and Maxwell RNA Kit (Promega) respectively. Fusions were identified using ArcherDx FusionPlex: Research Oncology Panel (ArcherDx) per manufacturer's protocol.

IMT case numbers 26-57: Fusions were identified using FoundationOne® assay, as previously reported¹⁴.

IMT case numbers 58-60, 62, 63: Fusions were identified using the Impact Assay²³ at Memorial Sloan Kettering Cancer Center.

IMT cases numbers 61 and 64: Fusions were identified using FoundationOne® assay²⁴.

RNA sequencing and Data Analysis

IMT case numbers 1-12, 66-69: RNA was extracted from FF samples. Libraries were prepared for whole transcriptome sequencing using Tru-Seq Stranded RNA library preparation. Sequencing was performed on the HiSeq2500 instrument (Illumina) with 100bp paired-end reads to an average sequencing depth of 50 million reads. Library preparation and sequencing was performed at the Broad Institute.

IMT case numbers 13, 14, 16, and 65: RNA was extracted from FF samples. Libraries were prepared

for whole transcriptome sequencing using ribo-reduction library preparation. Sequencing was performed on the HiSeq2500 instrument (Illumina) with 100bp paired-end reads to an average sequencing depth of 50 million reads. Library preparation and sequencing was performed at HudsonAlpha Research Institute.

Fusion Analysis

RNA sequencing fusion detection was performed using JAFFA²⁵ (version 1.09) and STAR-fusion (version 1.3.1) (reference genome hg19).

Differential Analysis

Illumina paired-end RNA sequencing data was aligned to the human genome assembly (reference genome hg38) using STAR²⁶ (version 2.5) with quantMode parameters set to TranscriptomeSAM for alignments translated into transcript coordinates. The transcriptomic alignments were run through RSEM²⁷ (version 1.2.31) command rsem-calculate-expression to calculate gene raw counts, TPM, FPKM and isoform expression. Differentially expressed genes were identified with R statistical software using the edgeR²⁸ package from Bioconductor. KEGG Pathway and Gene ontology enrichment analysis was conducted on those genes with a p-value <0.05 and an absolute Fold Change greater than 2. Identification of activated and inhibited pathway analysis was conducted using SPIA²⁹ (Signalling Pathway Impact Analysis).

SNV Analysis

Illumina paired-end RNA sequencing data was aligned to the human genome assembly (reference genome hg38) using STAR (version 2.5) 2-pass method. Alignments are sorted with SAMtools³⁰ (v. 1.3.1), duplicates marked with Picard Tools (version 2.4.1), reads are split and trimmed and mapping qualities reassigned with Genome Analysis Toolkit³¹ (version 3.6) using methods SplitNCigarReads and ReassignOneMappingQuality respectively. Variants are called with Haplotype Caller and annotated with ANNOVAR³².

Variants were filtered based on a list of 383 pre-defined cancer genes. The variants were maintained if they were exonic, had a damaging function, previously reported in COSMIC and were considered pathogenic in the ClinVar Database. Variants were further filtered and excluded if they had a variant allele frequency less than 0.05, depth less than 10, a prevalence of greater than 10% in the 1000 Genomes Project or if no annotations were present.

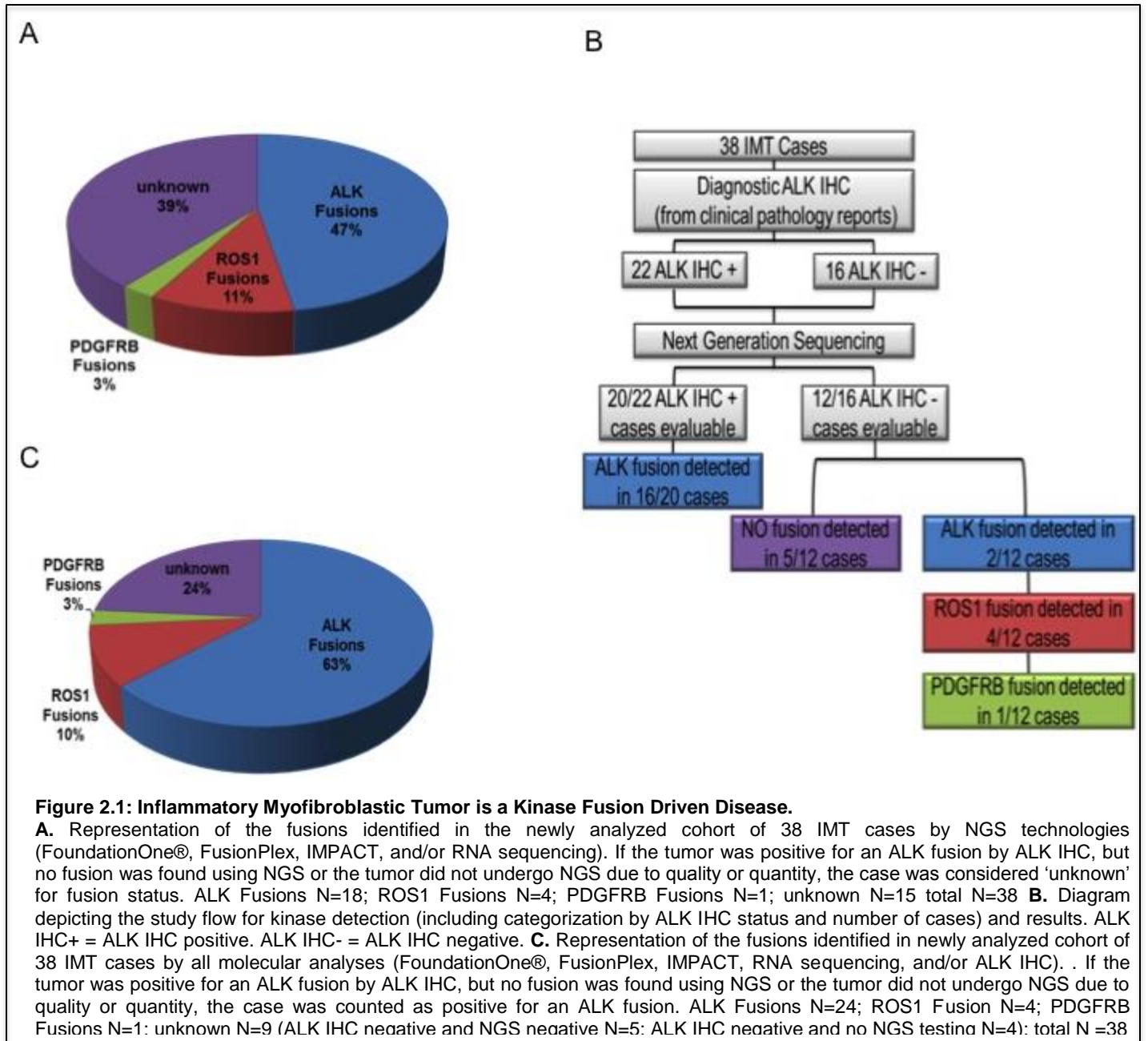
2.4 Results

Kinase Fusions Identified

In previous work, we identified *ROS1* and *PDGFRB* rearrangements that had never been reported before in IMT¹⁴. In total, approximately 85% of the IMTs assessed harbored actionable kinase fusions. More recently, others have reported *NTRK3* and *RET* fusions in IMT as well¹⁸⁻²⁰. Using established NGS platforms, we sought to identify fusions in an additional 38 distinct samples including, but not limited to, a cohort from the Children's Oncology Group and samples that were fusion negative by previous analyses. We identified *ALK*, *ROS1*, or *PDGFRB* rearrangements in 23 of the 38 samples (61%)(**Figure 2.1A**). *ALK* rearrangements were identified in 18 cases (47%). Both known and novel *ALK* fusion gene partners were detected. For example, *EEF1G-ALK* (detected in case 65) results in a novel fusion between *EEF1G*, a subunit in the complex responsible for aminoacyl t-RNA transport to the ribosome³³, on chromosome 11 and *ALK*. In addition, *KIF5B-ALK* (detected in case 18) results from a rearrangement between *KIF5B* on chromosome 10 and *ALK*. *KIF5B-ALK* has been previously detected in non-small cell lung cancer (NSCLC) but not IMT. *ROS1* rearrangements were detected in 4 cases (11% of the entire cohort). Interestingly, we detected a novel *PDGFRB* rearrangement, *NOTCH1-PDGFRB*, in one case (3% of the entire cohort). This fusion results from a chromosomal rearrangement between *NOTCH1* on chromosome 9 and *PDGFRB* on chromosome 5 and contains not only the entire *PDGFRB* tyrosine kinase domain but also all 34 exons of *NOTCH1*.

In 6/38 cases (cases numbers 16, 20-22, 59, and 60), *ALK* IHC was positive, but no *ALK* rearrangement was identified. We sought to understand the correlation between detection of *ALK* protein, through *ALK* IHC, and *ALK* rearrangements through genomic analysis. Of 22 cases that were positive for *ALK* expression by IHC, 16 had verified *ALK* rearrangements detected by NGS. For 2 of the 22 *ALK* IHC+ cases, there was insufficient tissue for NGS analysis. For 4 of the 22 *ALK* IHC+ cases (cases numbers 16, 21, 59, 60), no *ALK* rearrangement was detected through deep sequencing analysis (**Figure 2.1B**). One of these cases (#16) was from FF material and underwent RNA sequencing analysis. The other three cases were from FFPE and underwent targeted NGS via the FusionPlex platform (case #21) or MSKCC's Impact platform (case numbers 59 and 60). It is possible that an *ALK* rearrangement was not detected in these samples due to technical factors of the assay, low tumor content, or low expression of the fusion. By converse, there were 2 cases that were *ALK* IHC negative (case numbers 17 and 18) in which NGS analysis detected an *ALK* rearrangement.

This discrepancy is likely a result of low protein expression of the fusion or low tumor content on the analyzed tissue slide.



We acknowledge that ALK expression detected by IHC does not explicitly tell us that there is an ALK fusion present but rather only that ALK is expressed. However, for the purposes of clarity in this study, from this point on, we will make the assumption that ALK expression detected by IHC would also mean the tumor is ALK fusion positive, a reasonable assumption based on previous studies¹⁴. In total, considering NGS analyses and ALK IHC status, a total of 29/38 tumors (76%) harbored a kinase fusion (**Figure 2.1B,C**). In 5 of the 38 samples assessed by NGS, ALK IHC was negative and no rearrangement was identified by NGS (case numbers 6,12, 25, 61, 66). Three of the five samples underwent analysis by both RNA sequencing and the FusionPlex platform (case numbers 6,12, 66),

and one of the five by FusionPlex and FoundationOne (case 61). Four separate ALK IHC negative cases (case numbers 15, 23, 24, 68) either failed quality control (QC) or exhibited insufficient material and were unable to be evaluated by NGS.

Of those samples that were analyzed by multiple NGS methods, 6 samples displayed discordant results. In one case (case 7), no rearrangement was detected by the FusionPlex platform, but a *ROS1* fusion was detected by RNAseq. In the other case (Case #8), an *ALK* fusion was detected by

the FusionPlex platform, but the RNA sequencing data was of too poor quality to analyze. In case #39, no fusion was reported by FoundationOne¹⁴, but the *FN1-ALK* fusion harboring a non-canonical *ALK* breakpoint in intron 18 was identified using the FusionPlex platform. It is

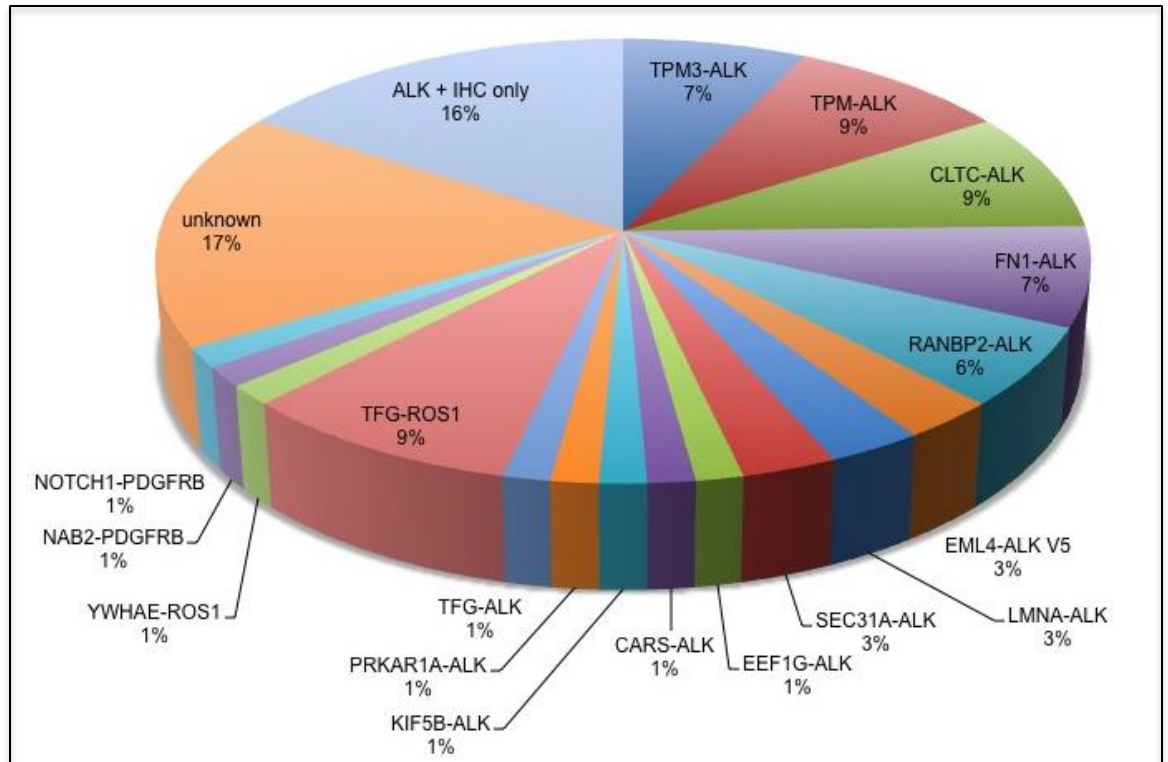


Figure 2.2: Kinase Fusions in IMT are Heterogeneous. Representation of the specific fusions detected in all (N=69) IMT cases by all molecular analyses (including those that only underwent ALK IHC testing). Fusion occurrence is represented as a percentage of all IMT cases.

important to note that the FoundationOne assay used in the previous study did not contain probes in intron 18 of *ALK*, but rather intron 19. This could be the reason it was not detected. Probes within intron 18 of *ALK* have now been added to the new FoundationOne assay to enable detection of *ALK* rearrangements with non-canonical breakpoints. The last two cases (#62 and 63) were negative for a fusion using MSKCC’s Impact assay, but *ALK* fusions were identified using the FusionPlex assay (Table 2.1).

Regarding the frequency of the various fusion partner genes, previous studies have postulated that *TPM3* and *TPM4* are the most common *ALK* fusion partners in IMT³⁴. When combining our entire cohort of IMT samples, including those previously published in 2014 (total N=69), we indeed verify

Case #	Sample ALK IHC		Fusion	5' Kinase		ArcherDx	RNAseq	FM1	IMPACT
	Type	Status		Exons	Exons				
1	FF	ALK +	RANBP2-ALK	18	20-29	N	Y-none	N	N
2	FF	ALK +	RANBP2-ALK	18	20-29	N	Y- RANBP2-ALK	N	N
3	FF	ALK +	RANBP2-ALK	18	20-29	N	Y-none	N	N
4	FF	ALK +	CLTC-ALK	31	20-29	N	Y- CLTC-ALK	N	N
5	FF	ALK +	SEC31L1-ALK	22	20-29	N	Y-none	N	N
6	FF	NA	NA	1	3-18	Y- none	Y-none	N	N
7	FF	ALK -	TFG-ROS1	4	36-43	Y-none	Y- TFG-ROS1	N	N
8	FF	ALK +	FN1-ALK	35	19-29	Y- FN1-ALK	Y-poor QC	N	N
9	FF	ALK -	TFG-ROS1	4	35-43	Y- TFG-ROS1	Y- TFG-ROS1	N	N
10	FF	ALK +	TPM3-ALK	6	20-29	Y- TPM3-ALK	Y- TPM3-ALK	N	N
11	FF	ALK +	TPM4-ALK	7	20-29	Y-TPM4-ALK	Y-TPM4-ALK	N	N
12	FF	NA	NA	NA	NA	Y-none	Y-none	N	N
13	FF	ALK +	TPM4-ALK	7	20-29	N	Y- TPM4-ALK	N	N
14	FF	ALK +	TPM4-ALK	7	20-29	N	Y- TPM4-ALK	N	N
15	FF	ALK -	NA	NA	NA	Y- poor QC	N	N	N
16	FF	ALK +	ALK	NA	NA	N	Y-none	N	N
17	FFPE	ALK -	FN1-ALK	29	18-29	Y-FN1-ALK	N	N	N
18	FFPE	ALK -	KIF5B-ALK	24	20-29	Y-KIF5B-ALK	N	N	N
19	FFPE	ALK -	TFG-ROS1	4	35-43	Y-TFG-ROS1	N	N	N
20	FFPE	ALK +	ALK	NA	NA	N-poor QC	N	N	N
21	FFPE	ALK +	ALK	NA	NA	Y-none	N	N	N
22	FFPE	ALK +	ALK	NA	NA	N- IM	N	N	N
23	FFPE	ALK -	NA	NA	NA	N-IM	N	N	N
24	FFPE	ALK -	NA	NA	NA	N-poor QC	N	N	N
25	FFPE	ALK -	NA	NA	NA	Y-none	N	N	N
26	FFPE	ALK -	NA	NA	NA	N	N	Y-none	N
27	FFPE	ALK -	NA	NA	NA	N	N	Y-none	N
28	FFPE	ALK -	YWHAE-ROS1	4	36-43	N	N	Y- YWHAE-ROS1	N
29	FFPE	ALK -	EML4-ALK	2	20-29	N	N	Y- EML4-ALK	N
30	FFPE	ALK -	TFG-ROS1	4	36-43	N	N	Y-TFG-ROS1	N
31	FFPE	ALK -	NA	NA	NA	N	N	Y-none	N
32	FFPE	ALK -	TPM3-ALK	7	20-29	N	N	Y-TPM3-ALK	N
33	FFPE	ALK -	NAB2-PDGFRB	7	12-23	N	N	Y- NAB2-PDGFRB	N
34	FFPE	ALK +	RANBP2-ALK	18	20-29	N	N	Y- RANBP2-ALK	N
35	FFPE	ALK +	LMNA-ALK	2	20-29	N	N	Y- LMNA-ALK	N
36	FFPE	ALK +	TPM3-ALK	7	20-29	N	N	Y- TPM3-ALK	N
37	FFPE	ALK +	TPM4-ALK	7	20-29	N	N	Y- TPM4-ALK	N
38	FFPE	ALK +	TPM4-ALK	7	20-29	N	N	Y- TPM4-ALK	N
39	FFPE	ALK +	FN1-ALK	19	19-29	Y-FN1-ALK	N	Y-none	N
40	FFPE	ALK +	ALK	NA	NA	N	N	Y-fail	N
41	FFPE	ALK +	ALK	NA	NA	N	N	Y-none	N
42	FFPE	ALK +	EML4-ALK	2	20-29	N	N	Y- EML4-ALK	N
43	FFPE	ALK +	TPM3-ALK	7	20-29	N	N	Y- TPM3-ALK	N
44	FFPE	ALK +	TPM3-ALK	7	20-29	N	N	Y- TPM3-ALK	N
45	FFPE	ALK +	ALK	NA	NA	N	N	Y-fail	N
46	FFPE	ALK +	SEC31L1-ALK	22	20-29	N	N	Y- SEC31L1-ALK	N
47	FFPE	ALK +	ALK	NA	NA	N	N	Y-fail	N
48	FFPE	ALK +	TFG-ALK	6	20-29	N	N	Y- TFG-ALK	N
49	FFPE	ALK +	FN1-ALK	23	19-29	N	N	Y- FN1-ALK	N
50	FFPE	ALK +	CLTC-ALK	31	20-29	N	N	Y- CLTC-ALK	N
51	FFPE	ALK +	CLTC-ALK	31	20-29	N	N	Y- CLTC-ALK	N
52	FFPE	ALK +	FN1-ALK	23	19-29	N	N	Y- FN1-ALK	N
53	FFPE	ALK +	ALK	NA	NA	N	N	Y-none	N
54	FFPE	ALK +	CLTC-ALK	31	20-29	N	N	Y- CLTC-ALK	N
55	FFPE	ALK +	CLTC-ALK	31	20-29	N	N	Y- CLTC-ALK	N
56	FFPE	ALK +	PRKAR1A-ALK	5	20-29	N	N	Y- PRKAR1A-ALK	N
57	FFPE	ALK -	TFG-ROS1	4	36-43	N	N	Y- TFG-ROS1	N
58	FFPE	ALK +	LMNA-ALK	2	20-29	N	N	Y-none	Y- LMNA-ALK
59	FFPE	ALK +	ALK	NA	NA	N	N	N	Y-none
60	FFPE	ALK +	ALK	NA	NA	N	N	N	Y-none
61	FFPE	ALK -	NA	NA	NA	Y-none	N	Y-none	N
62	FFPE	ALK +	CARS-ALK	18	20-29	Y-CARS-ALK	N	N	Y-none
63	FFPE	ALK +	TPM4-ALK	7	20-29	Y-TPM4-ALK	N	N	Y-none
64	FFPE	ALK -	TFG-ROS1			N	N	Y- TFG-ROS1	N
65	FF	ALK +	EEF1G-ALK	6	20-29	Y-EEF1G-ALK	Y- EEF1G-ALK	N	N
66	FF	ALK -	NA	NA	NA	Y-none	Y- none	N	N

Case #	Sample ALK IHC		Fusion	5' Kinase		ArcherDx	RNAseq	FM1	IMPACT
	Type	Status		Exons	Exons				
67	FF	ALK +	<i>CLTC-ALK</i>	31	20-29	N	Y- <i>CLTC-ALK</i>	N	N
68	FF	ALK -	NA	NA	NA	N	N	N	N
69	FF	ALK -	<i>NOTCH1-PDGFRB</i>	34	12-23	N	Y- <i>NOTCH1-PDGFRB</i>	Y- <i>NOTCH1-PDGFRB</i>	N

Table 2.1: Annotation of 69 IMT cases that underwent genomic analyses as part of this study.

This table details the method(s) through which each case was analyzed. Y = Yes, meaning that the sample underwent the noted type of analysis, as defined in the column heading. N= No, meaning that the sample did not undergo the noted type of analysis, as defined in the column heading. FF = fresh frozen. FFPE = formalin fixed and paraffin embedded. FM1 = Foundation Medicine FoundationOne® assay. NA = not available. ALK- = ALK IHC negative. ALK+ = ALK IHC positive. IM = insufficient material

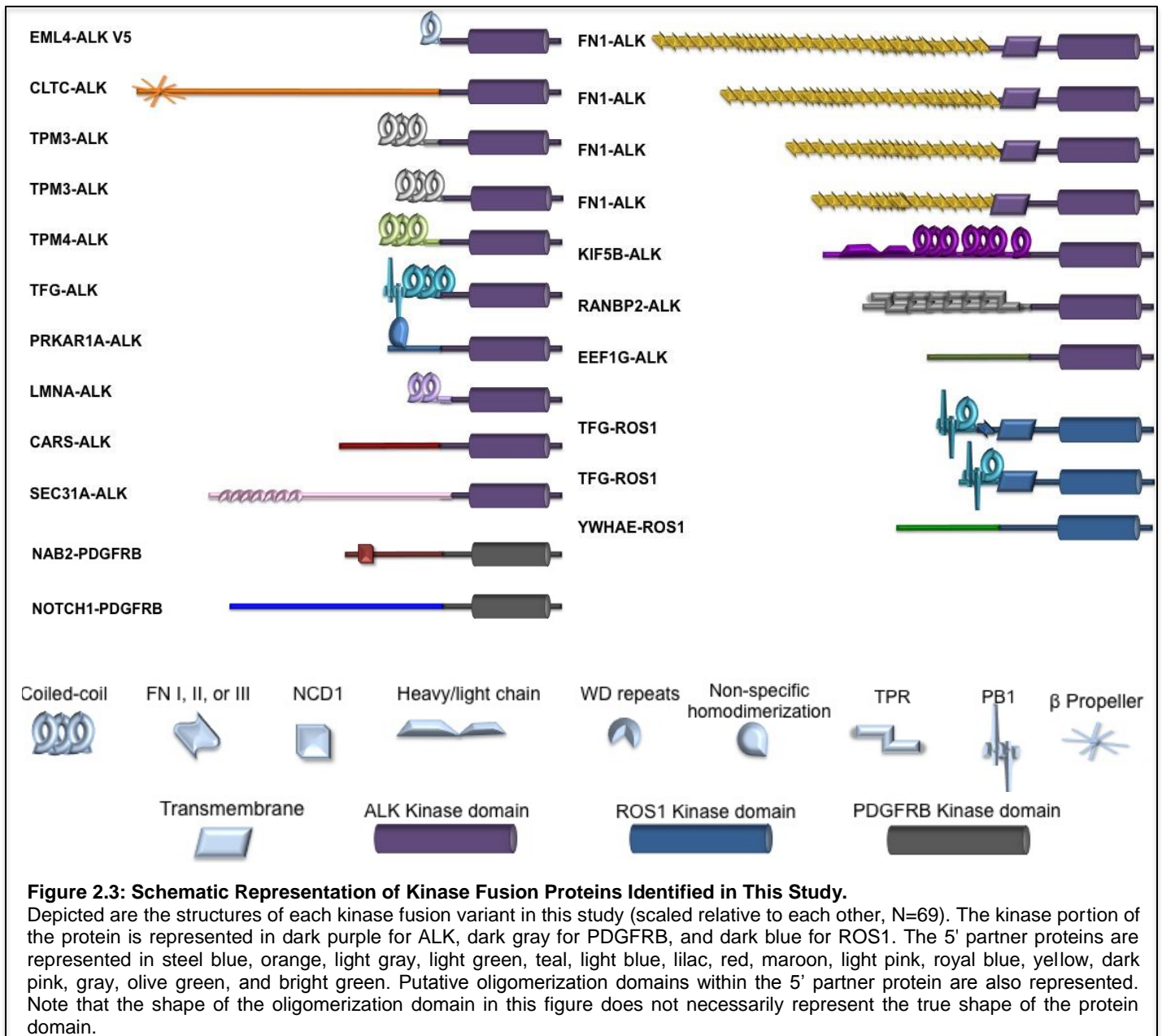
that *TPM3/4-ALK* are the most common fusions, totaling 11 cases when combining both *TPM3-ALK* (5 cases) and *TPM4-ALK* (6 cases) (**Figure 2.2**). The next most common fusion partners in our study were *CLTC* (6 cases), *TFG-ROS1* (6 cases), and *FN1-ALK* (5 cases). Of note, 4 cases in our total cohort harbored *RANBP2-ALK* indicating that these are likely eIMS since the *RANBP2-ALK* fusion, has thus far been detected exclusively in the eIMS subtype^{11,17}.

Starting with the genomic breakpoints obtained through our NGS analyses, we translated each transcript and subsequently constructed schematic representations for each putative fusion protein (**Figure 2.3, Table 2.1**). All transcripts resulted in an in frame predicted fusion protein. Putative oligomerization domains and the tyrosine kinase domains were mapped for each case. We made several interesting observations with this analysis. First, we noted that a given fusion partner may have multiple distinct breakpoints, resulting in variable lengths of the partner gene fused to the kinase. For example, we detected four different *FN1-ALK* variants that differ by the breakpoint in *FN1* as well as *ALK*. Importantly, they all retain the transmembrane domain of *ALK*, and one also retained an extra-cellular region of *ALK*, exon 18 (case 17). Next, we noted overlap of the 5' partner between multiple kinases, such as *TFG-ALK* and *TFG-ROS1*. Interestingly, although *TFG* was identified here as a 5' fusion partner for both *ALK* and *ROS1*; the breakpoints in *TFG* differed between the *ALK* fusion and the *ROS1* fusions. Additionally, the breakpoint in *ROS1* differed between the *TFG-ROS1* fusions identified, resulting in two different *TFG-ROS1* variants.

Patient and Tumor Characterization

Consistent with current literature, we did not find a gender predilection for IMT in our study. Of the 69 distinct patient samples that we have evaluated to date, 49% were from male patients, and 51% were from female patients (**Table 2.2**). In addition, we found that the proportion of tumors harboring a kinase fusion was similar between male and female patients with kinase fusions being found in 88% of tumors from male patients and 77% of female patient's tumors. When considering all NGS analyses as well as ALK IHC, 84% of the 69 distinct tumors were positive for a kinase fusion. *ALK*

fusions occurred in 70% of all tumors (52% male, 48% female); *ROS1* fusions occurred in 10% of tumors (57% male, 43% female), and *PDGFRB* fusions occurred in 3% (50% male, 50% female) (Table 2.2). Studies to date have reported age/fusion associations considering only the ALK kinase



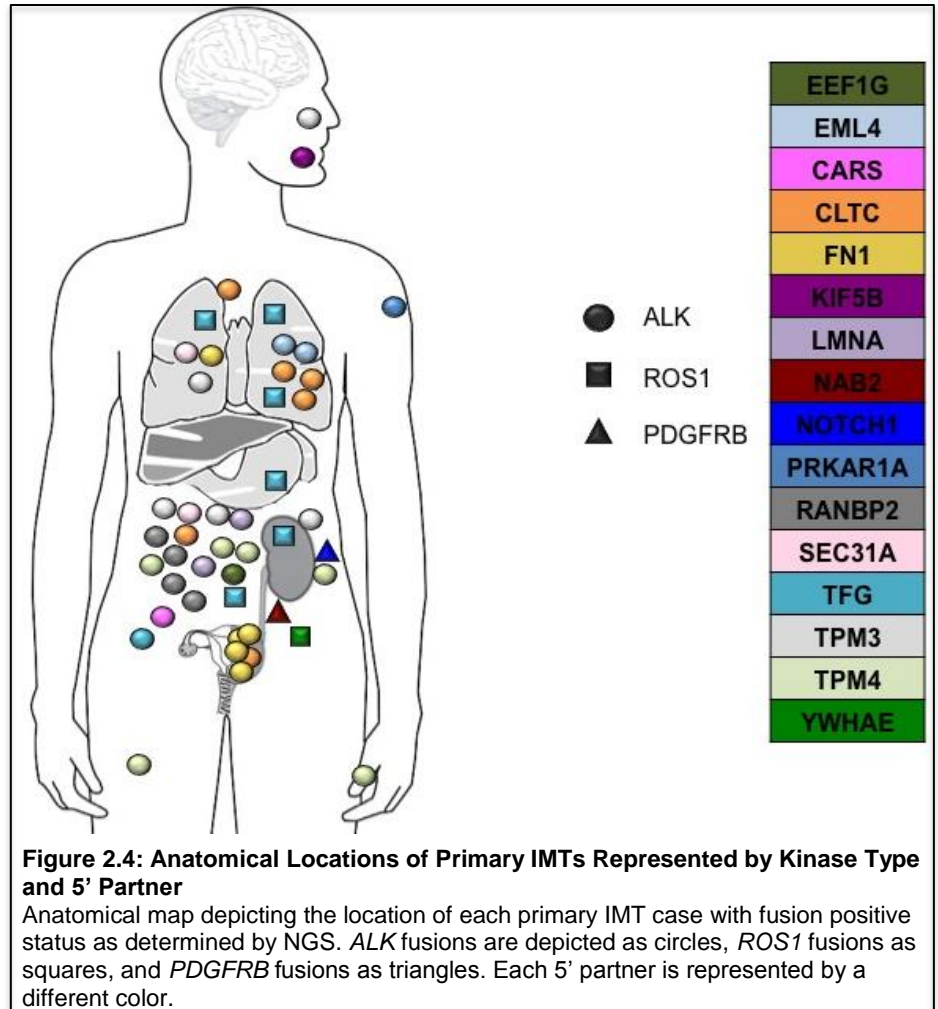
and with detection methodologies less sensitive than NGS, such as ALK IHC^{5,19}. Consistent with many other cancer types, these studies report a greater prevalence of ALK fusions in younger patients. In our study we did not find such associations. The average age of our overall patient cohort was 11 years. Those harboring *ALK* fusions exhibited a mean age of 11 years, as did both kinase fusion positive and kinase fusion negative cohorts. The average age of those harboring *ROS1* fusions was 10 years. Patients with *PDGFRB* tumors averaged 7 years of age. There was also no

relationship between age, sex, and fusion status identified.

While IMTs can occur at any anatomic location, the lungs and the abdomen are the most commonly reported disease sites^{1,35,36}. Surgical resection is the standard of care for all IMTs. However, surgical resection may present significant clinical challenges depending on the size and location of the tumor in the context of other vital anatomic structures. As a result, physicians may sometimes be faced with the decision to give neo-adjuvant therapy to

facilitate resection. Therefore, we wanted to know if intra-thoracic or intra-abdominal tumors specifically showed a tendency for kinase fusions, and if so which ones. We chose to focus on intra-thoracic tumors since, in our study, 23% of tumors originated in the lungs (**Table 2.3**). Similar to IMTs overall, we found that 82% of pulmonary IMTs were positive for expression of a potential kinase fusion (by NGS or IHC). 63% of which harbored *ALK* fusions as determined by NGS or were positive by ALK IHC, and 19% harbored *ROS1* fusions (**Table 2.2**). Since crizotinib targets both ALK and ROS1, 82% of patients with pulmonary IMT may experience a positive response to crizotinib treatment.

We also sought to evaluate the specific fusion reported in relation to the anatomical location. While there does appear to be some anatomical preference for certain fusions, such as *FN1-ALK* which was reported in the bladder in 4/5 cases, we did not find that any particular fusion strictly adhered to one organ site or that any one organ harbored only one type of fusion (**Figure 2.4, Table 2.3**). In agreement with current literature, however, *RANBP2-ALK* positive tumors were all reported in the abdomen^{11,17}, excluding this fusion from the lungs.



	<i>ALK</i> N=48	<i>ROS1</i> N=7	<i>PDGFRB</i> N=2	Fusion neg N=12	Fusion pos N=57	Total IMTs N=69
Clinicopathological Parameters						
age (years, mean)	11	10	6.5	11	11	11
age (years, mean) of males	10	10	1.5	8	10	9
age (years, mean) of females	12	10	12	13	12	12
sex (male)% of total IMTs	36	6	1	6	43	49
pulmonary (% of total IMTs)	10 (14)	3 (4)	0	3 (4)	13(19)	16(23)
sex (male)% of fusion	52	57	50	33	53	49
pulmonary (% of fusion)	10 (21)	3 (43)	0	3 (25)	13(23)	16
sex (male)% of male	74	12	3	12	88	100
pulmonary (% of pulmonary)	10 (63)	3 (19)	0	3 (19)	13(81)	16

Table 2.2: Clinicopathological Features of Inflammatory Myofibroblastic Tumors According to Fusion Status. All 69 IMT cases analyzed to date are represented. Cases positive by ALK IHC were considered ALK positive and fusion positive regardless of fusion status determined by NGS. Neg= negative. Pos = positive.

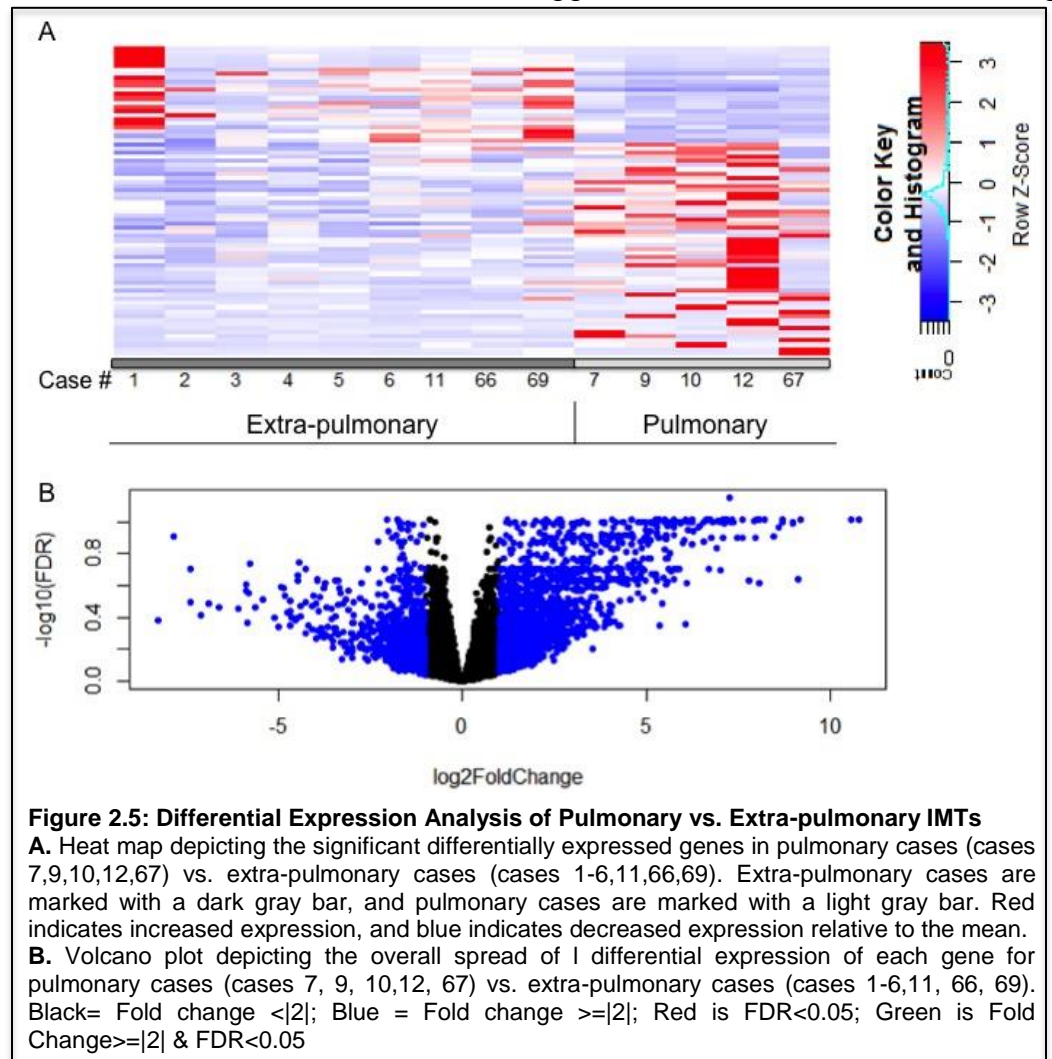
Pulmonary vs. Extra-pulmonary IMT Expression Analysis

Rate of local recurrence, but not metastasis, appears to differ based on anatomical site. Primary tumors confined to the lung recur at a rate of <2%^{4,36,37} compared to 25% for extra-pulmonary disease¹. Given this documented difference in pulmonary vs. extra-pulmonary IMT, we interrogated our RNAseq data for differences in gene expression between these cohorts. We performed a differential expression analysis on 15 samples (case numbers 1-7, 9-12, 66-69), 5 of which were pulmonary, and 10 were extra-pulmonary. We identified a cohort of 74 significant differentially expressed genes (DEGs) (FC=>|2|, p=<0.05) (**Figure 2.5A,B**). (No genes were exhibited significantly different expression when adjusting the analysis for multiple comparisons (FDR<=0.05), likely due to small sample size.) KEGG Pathway and Gene Ontology (GO) term enrichment analyses revealed up-regulation of pathways in protein processing in the endoplasmic reticulum (ER) and negative regulators of ER stressed-induced apoptosis and down-regulation of inflammatory responses (**Table S2.1**). Signaling Pathway Impact Analysis (SPIA) of the 74 genes revealed inhibition of the cell cycle and activation of neuroactive ligand-receptor interactions in pulmonary samples compared to extra-pulmonary samples (**Table S2.4**).

Expression Analysis of RANBP2-ALK Tumors

The recently described subtype of IMT, eIMS, differentiates itself from conventional IMT histologically and clinically. These tumors have been known to exhibit a more aggressive clinical behavior including

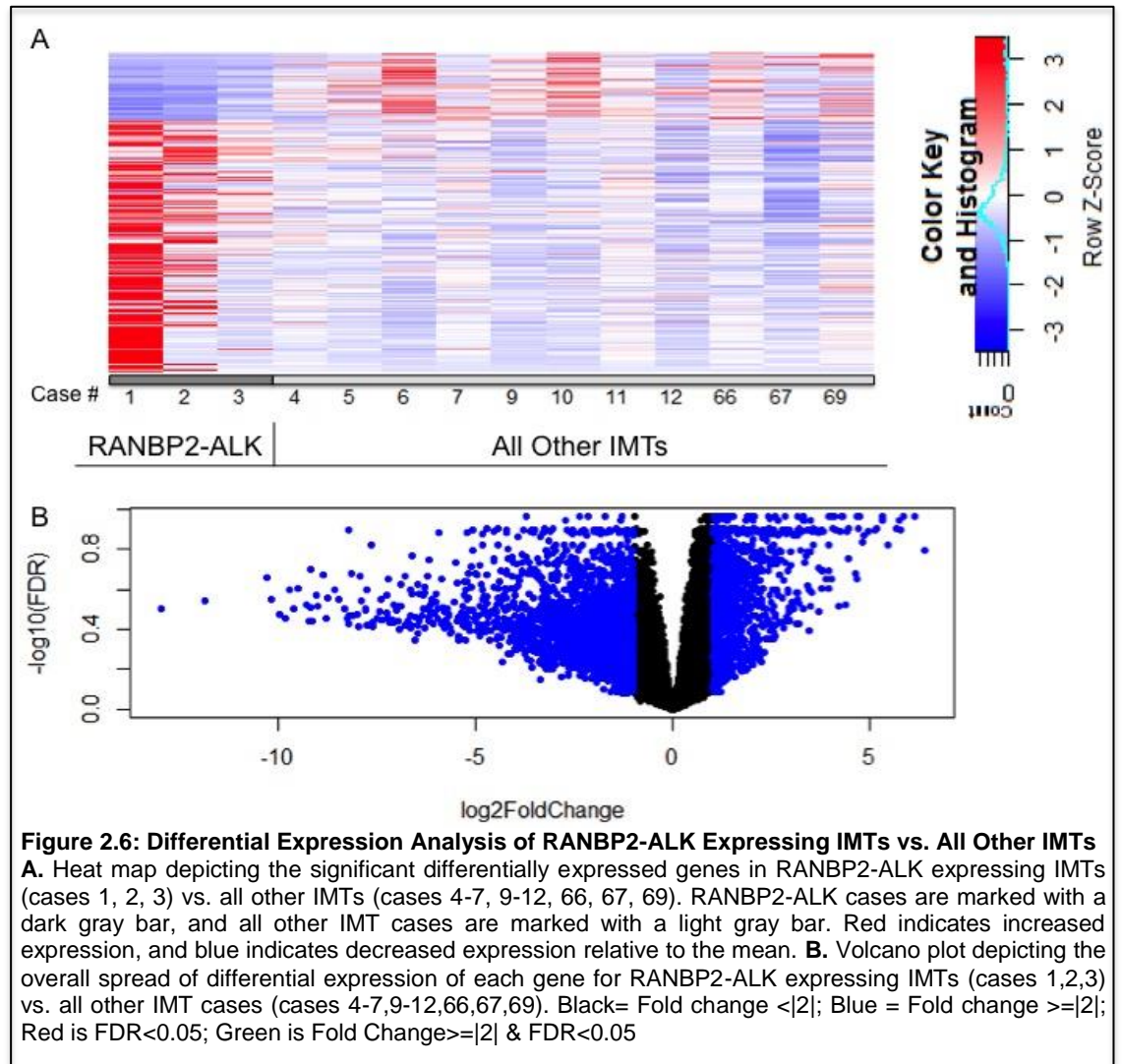
the prevalence for metastasis⁸⁻¹¹ and resistance to ALK TKI therapy^{10,38,39}. Therefore RNAseq data was interrogated to identify DEGs when comparing cases harboring the RANBP2-ALK fusion described in eIMS and other IMTs. Differential expression analysis was performed on a cohort of 15 samples (case numbers 1-7, 9-12, 66-69), of which 3 harbored *RANBP2-ALK*. We identified 235 DEGs ($FC \geq |2|$, $p < 0.05$) (Figure 2.6A, B). (No



genes were exhibited significantly different expression when adjusting the analysis for multiple comparisons ($FDR \leq 0.05$), likely due to small sample size.) KEGG Pathway and GO term enrichment analyses showed up-regulation of lysosomal and phagocytic pathways and processes in the inflammatory response and leukocyte migration and down-regulation of complement cascades, extracellular matrix organization, and cell adhesion (Table S2.2). SPIA revealed several impacted pathways including inhibition of Neuroactive ligand-receptor interactions and ECM-receptor interactions and activation of Fc gamma R-mediated phagocytosis, Fc epsilon RI signaling, and chemokine signaling (Table S2.5).

Fusion Positive vs. Fusion Negative Expression Analysis

In our studies, we have consistently found that approximately 85% of IMTs harbor therapeutically targetable kinase fusions (as determined by NGS or ALK IHC). There are no data to aid in directing clinical treatment for the remaining 15%; therefore, we sought to identify expression differences between fusion positive and fusion negative samples that may give insight into pathogenesis and potential therapeutic treatments.



Differential expression analysis was performed on a cohort of 15 samples (case numbers 1-7, 9-12, and 66-69), 3 of which were negative for a kinase fusion, and 12 were positive for a kinase fusion. We identified a group of 145 DEGs (FC=>|2|, p=<0.05) (**Figure 2.7A,B**). (Only one gene, *GBE1*, exhibited significantly different expression when adjusting the analysis for multiple comparisons (FDR<=0.05), likely due to small sample size). KEGG Pathway and Gene Ontology term enrichment analyses revealed up-regulation of cellular components related to extracellular space and exosomes and down-regulation of pathways related to the biosynthesis of antibiotics and cellular components involved in a proteinaceous extracellular matrix in fusion positive samples (**Table S2.3**). SPIA results showed activation of PPAR signaling and cytokine-cytokine receptor interactions in fusion negative samples (**Table S2.6**).

Case #	Age	Sex	Anatomical Location	Fusion
1	9	M	abdomen	<i>RANBP2-ALK</i>
2	14	M	abdomen	<i>RANBP2-ALK</i>
3	11	M	abdomen	<i>RANBP2-ALK</i>
4	7	F	chest	<i>CLTC-ALK</i>
5	0.5	F	abdomen	<i>SEC31L1-ALK</i>
6	9	M	eye, orbit-right	NA
7	13	M	lung	<i>TFG-ROS1</i>
8	4	M	bladder	<i>FN1-ALK</i>
9	4	F	lung	<i>TFG-ROS1</i>
10	6	M	Lung, lower left lobe	<i>TPM3-ALK</i>
11	0.5	M	Right posterior thigh	<i>TPM4-ALK</i>
12	17	F	lung upper lobe	NA
13	3	F	abdomen	<i>TPM4-ALK</i>
14	5	M	abdomen	<i>TPM4-ALK</i>
15	9	F	kidney, left	NA
16	13	M	bladder	<i>ALK (IHC)</i>
17	8	F	bladder	<i>FN1-ALK</i>
18	0.5	M	floor of mouth	<i>KIF5B-ALK</i>
19	14	M	kidney	<i>TFG-ROS1</i>
20	16	M	bladder	<i>ALK (IHC)</i>
21	2	M	lung, left	<i>ALK (IHC)</i>
22	8	F	abdomen	<i>ALK (IHC)</i>
23	8	M	lung	NA
24	10	M	abdomen	NA
25	16	F	small bowel	NA
26	14	F	mesentery	NA
27	16	F	mesentery	NA
28	22	F	buttock/pelvis	<i>YWHAE-ROS1</i>
29	38	F	lung	<i>EML4-ALK</i>
30	8	M	mesentery	<i>TFG-ROS1</i>
31	5	M	lung	NA
32	41	M	nasopharynx	<i>TPM3-ALK</i>
33	12	F	peritoneum	<i>NAB2-PDGFRB</i>
34	6	M	omentum	<i>RANBP2-ALK</i>
35	7	F	mesentery	<i>LMNA-ALK</i>
36	2	F	mesentery	<i>TPM3-ALK</i>
37	3	F	mesentery	<i>TPM4-ALK</i>
38	29	M	mesentery	<i>TPM4-ALK</i>
39	36	F	lung	<i>FN1-ALK</i>
40	13	M	lung	<i>ALK (IHC)</i>
41	2	M	bladder	<i>ALK (IHC)</i>
42	11	F	lung	<i>EML4-ALK</i>
43	7	M	mesentery	<i>TPM3-ALK</i>
44	20	F	mesentery	<i>TPM3-ALK</i>

Case #	Age	Sex	Anatomical Location	Fusion
45	1	M	mesentery	ALK (IHC)
46	6	F	lung	<i>SEC31L1-ALK</i>
47	4	M	mesentery	ALK (IHC)
48	14	M	pelvis	<i>TFG-ALK</i>
49	26	F	bladder	<i>FN1-ALK</i>
50	26	F	bladder	<i>CLTC-ALK</i>
51	14	M	mesentery	<i>CLTC-ALK</i>
52	18	F	bladder	<i>FN1-ALK</i>
53	10	M	mesentery	ALK (IHC)
54	9	F	lung	<i>CLTC-ALK</i>
55	4	F	lung	<i>CLTC-ALK</i>
56	0.5	F	shoulder	<i>PRKAR1A-ALK</i>
57	6	M	lung	<i>TFG-ROS1</i>
58	9	M	small bowel	<i>LMNA-ALK</i>
59	17	F	bladder serosa and pelvis	ALK (IHC)
60	5	M	small bowel mesentery	ALK (IHC)
61	14	F	stomach	NA
62	16	F	pelvic and mesentery	<i>CARS-ALK</i>
63	9	M	thumb	<i>TPM4-ALK</i>
64	4	F	stomach	<i>TFG-ROS1</i>
65	0.5	F	abdomen	<i>EEF1G-ALK</i>
66	12	F	uterus	NA
67	9	F	lung	<i>CLTC-ALK</i>
68	4	F	abdomen	NA
69	1.5	M	abdomen	<i>NOTCH1-PDGFRB</i>

Table 2.3: Clinicopathological Features of the entire IMT Cohort. All 69 IMT cases analyzed to date are represented. Cases positive by ALK IHC only were considered ALK positive and noted ALK as the fusion by IHC.

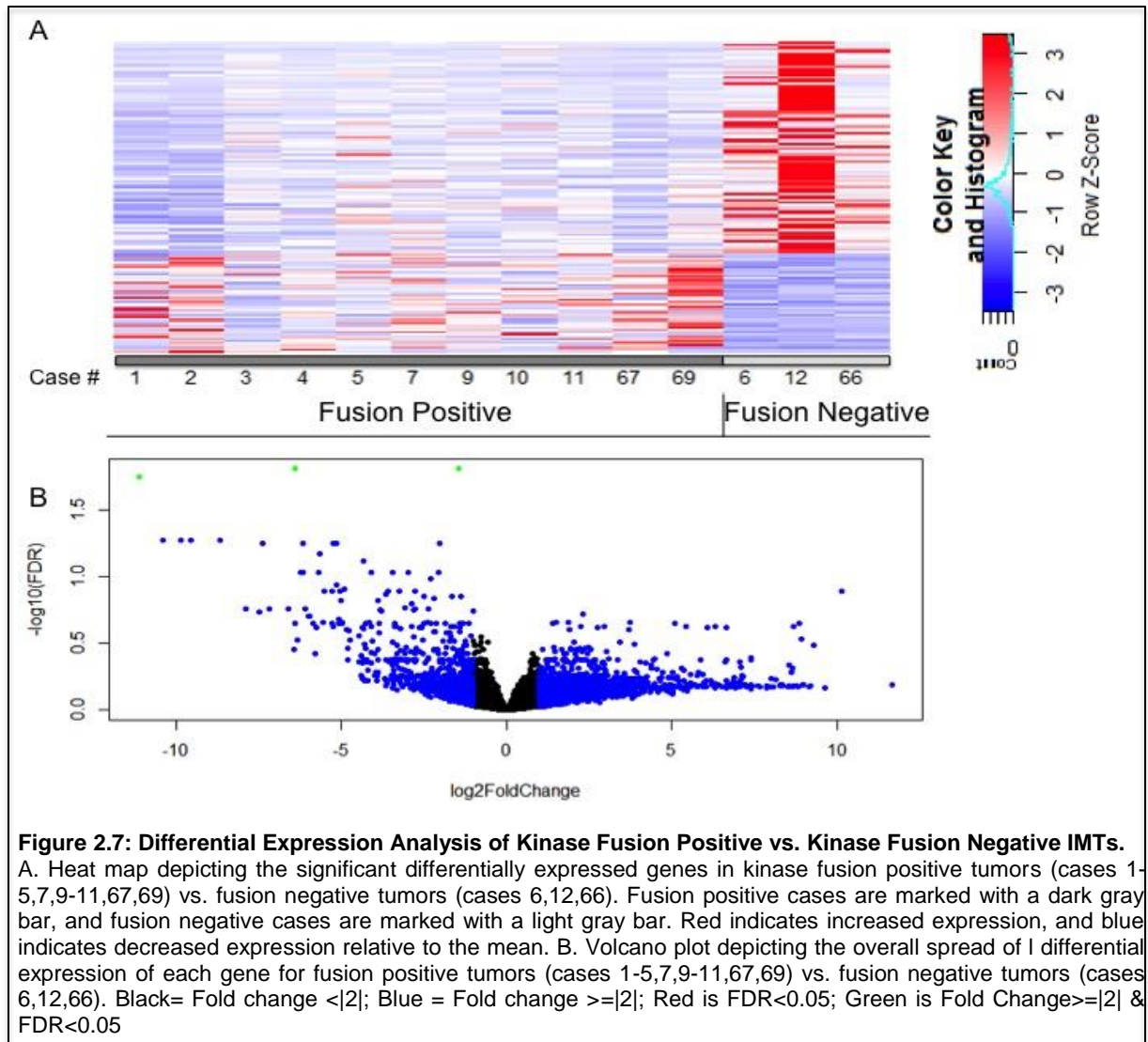
SNVs

Aside from the presence of kinase fusions, little is known regarding the mutational landscape of IMT. This results in the use of therapies without much rationale to support them in the ~15% of patients whose tumors are kinase fusion negative or for kinase fusion positive patients whose disease has progressed on targeted therapy. Using RNAseq of 18 distinct patient samples, we sought to identify cancer associated single nucleotide variants (SNVs) that may give insight into the pathogenesis of IMT and nominate potential therapies. We found that each sample harbored between 20 and 52 cancer associated SNVs. We noted that many of the SNVs were common amongst the samples. There were 30 SNVs shared by at least 25% of the samples (N>=5) (**Figure 2.8A, Table S2.7**), including a *RPL22* frameshift insertion (44 A duplication, K15 frame shift) that occurred in every sample with a variant allele frequency (VAF) <40% in all samples, and *TP53* C98G (P33R) that

occurred in 17/18 samples. *ROS1* fusions shared the least number of common SNVs (17/30). SNVs excluding *ROS1* samples were in *SETD2*, *ATRX*, *DHX15*, *CROCC*, *IGLL5*, *TET2*, *NUP98*, *LRIG1*, *KMT2A*, *GAK*, *KMT2C*, *MGA*, and *ZFH3*. We also noted that a *NCOR1* frameshift (636A duplication, V213 frame shift) consisted of a low percentage of ALK samples relative to the other 29 commonly shared SNVs. The common SNVs more exclusive to fusion positive tumors were *DHX15* frameshift (1224A duplication, Q409 frame shift), *CROCC* G2072A (R691H), *RHOA* A485G (K162R), *TSC1* T965C (M322T) and *KMT2A* frameshift (9308 A duplication, Q3103 frame shift) while those more

exclusive to fusion negative tumors were *MTAP* G166A (V56I) and *ZFH3* G247T (A83S) (Figure 2.8B). Both *MTAP* and *ZFH3* are considered tumor suppressors.

In order to further determine



significance of these common SNVs, we assessed the variant allele frequencies (VAFs). Interestingly, while the *RPL22* SNV was present in all samples, it always occurred as one of the lowest allele frequencies (VAF < 40% (Tables S2.8-S2.25)). The SNVs consistently having the highest allele frequencies, homozygous in most cases, were *TP53*, *DDX31*, *IL7R*, *ZNF384*, *SETD2*, and *ATRX* (Tables S2.8-S2.25). We wanted to know if there was a common pathway in which these shared SNVs occurred. GO term enrichment analysis of the 30 shared SNVs revealed a striking involvement

of the VEGF pathway and chromatin remodeling proteins, including histone methylation (**Table S2.26**).

We also explored additional SNVs occurring in the majority (2/4) of fusion negative samples. We found that 3 of 4 fusion negative samples shared the *PIK3R1* G978A (M326I) mutation, and half shared the *EGFR* G1562A (R51K) mutation. Both of these aberrations occurred with high allele frequencies (VAF > 50% (**Figure 2.9, Tables S2.8-S2.25**)).

2.5 Discussion

It is well established that *ALK* fusions occur in ~50% of IMT¹²⁻¹⁴, and clinical studies of *ALK* inhibitors have proven effective in *ALK* positive IMT^{21,22}. Recent NGS studies of IMT have revealed the presence of targetable kinase fusions in *ALK* negative tumors including *ROS1*, *PDGFRB*, *RET*, and *NTRK3*^{14,18,19}. In targeted NGS studies of relatively large cohorts of this rare tumor, the majority of tumors have been found to harbor a kinase fusion¹⁴, emphasizing the importance of utilizing NGS diagnostic tools to improve clinical decision making for this disease. Nothing yet is known on the genomic or transcriptomic level for those that do not harbor kinase fusions or for kinase fusion positive tumors which progress of therapy, highlighting the need for further genomic and tumor expression studies that may help us understand IMT pathogenesis and identify rational therapeutic options.

Our study aimed to increase the knowledge surrounding this rare and heterogeneous tumor, coupling clinicopathologic and transcriptomic data, specifically regarding the presence of kinase fusions, differences in gene expression between cohorts of pre-determined clinical significance, and the presence of cancer associated SNVs.

Consistent with our previous report, we have identified targetable kinase fusions in 84% of our entire IMT cohort (N=69) including *ALK*, *ROS1*, and *PDGFRB* kinases (as determined by NGS and *ALK* IHC). We did not find any predilection for different fusions (as determined by the kinase type or 5' partner) for specific anatomical locations, age, or gender (**Table 2.2**).

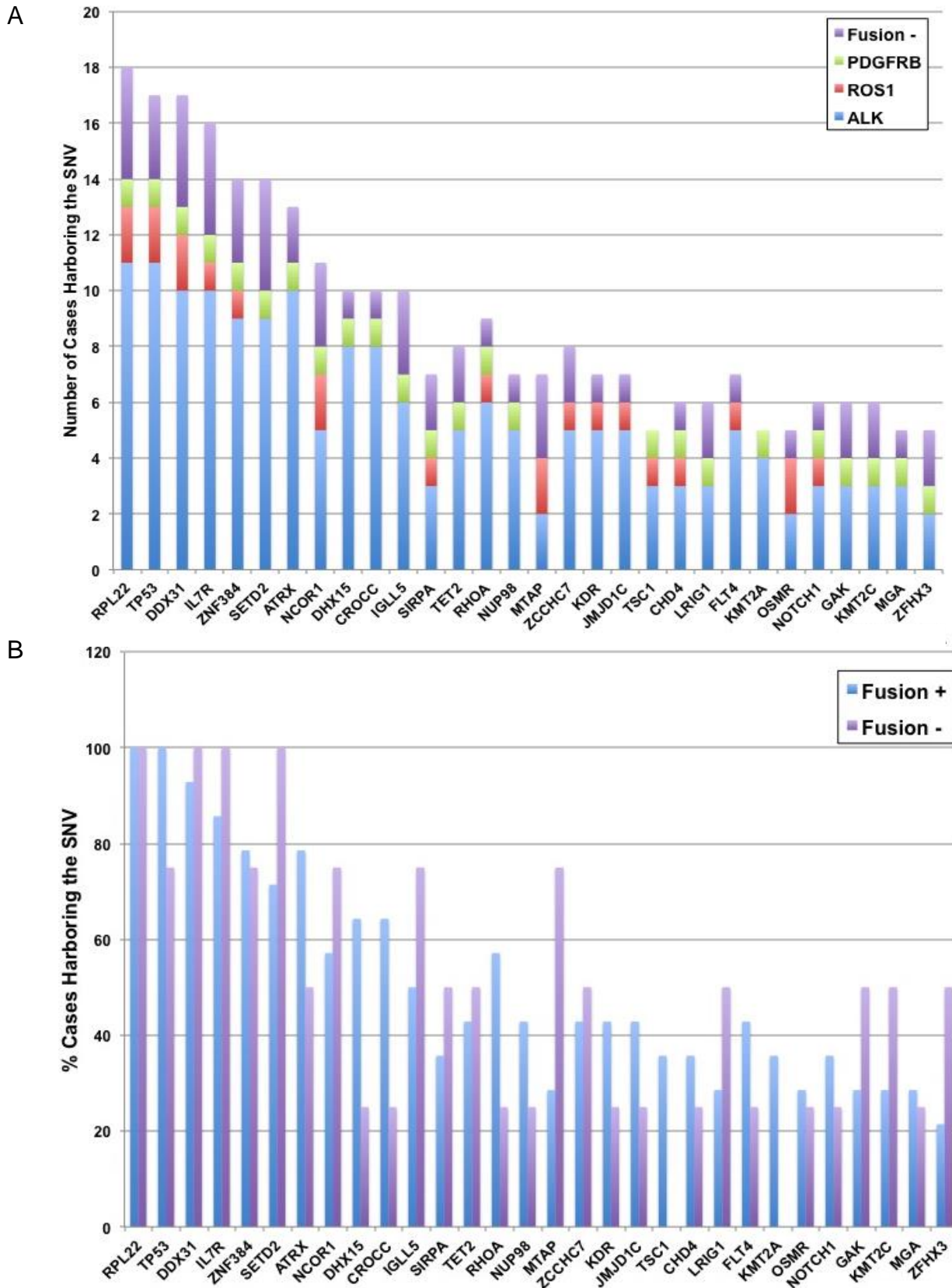


Figure 2.8: Genes Harboring Cancer Associated SNVs Shared by at Least 25% of IMTs Evaluated.

All 18 IMTs undergoing RNAseq analysis were assessed for SNVs. Variants were filtered based on a list of 383 pre-defined cancer genes. The variants were maintained if they were exonic, had a damaging function, previously reported in COSMIC, and were considered pathogenic in the ClinVar Database. Variants were then excluded if they had a variant allele frequency below 0.05, depth below 10, a prevalence of greater than 10% in the 1000 Genomes Project or if no annotations were present. Genes harbouring SNVs shared by at least 25% of samples are displayed. **A.** The number of positive tumors for each fusion category is represented. ALK N=11; ROS1 N=2; PDGFRB N=1; Fusion negative N=4 **B.** The percent of positive tumors within each category is represented. Fusion positive N=14; Fusion negative N=4.

Previous literature reports show that older patients (>40 years) tend have ALK negative tumors^{5,19}. The average patient age in our cohort was 11 years with the oldest patient being 41 years of age. We did not find a correlation between the presence/absence of a kinase fusion and age, or for anatomical location or gender (Table 2.2).

We report several novel fusion findings, including: (a) a new ALK fusion partner, *EEF1G*; (b) a new *PDGFRB* fusion partner, *NOTCH1*; (c) novel ALK fusion partner for IMT specifically, including *KIF5B-ALK* which also occurs in NSCLC; and

(d) novel variants of previously reported ALK fusions, including 4 variants of *FN1-ALK* which differ based on both the *FN1* and *ALK* genomic breakpoints (Table 2.1, Figure 2.3). The full biological and clinical significance of different fusion variants that vary by the specific 5' partner protein (such as *TFG-ALK* vs. *TPM3-ALK*) or the length of the same 5' partner protein (such as the four different *FN1-ALK* variants identified in this study) is unknown, but previous clinical and pre-clinical studies suggest that there are biochemical differences that affect efficacy of TKIs and Hsp90 inhibitors⁴⁰⁻⁴⁶. We also made an interesting observation regarding fusions in which *TFG* is the 5' protein. *TFG* was a 5' partner for both *ALK* and *ROS1*; however, the breakpoints in *TFG* differed between *TFG-ALK* and *TFG-ROS1*. The *TFG* breakpoint in *TFG-ROS1* disrupts the coiled-coil oligomerization domain resulting in a shorter coiled-coil domain for *TFG-ROS1*. The significance of this difference is unknown. It is possible that this breakpoint is required for an in-frame fusion. Another factor to consider is the retention of the transmembrane domain of *ROS1* in the *TFG-ROS1* fusions. The

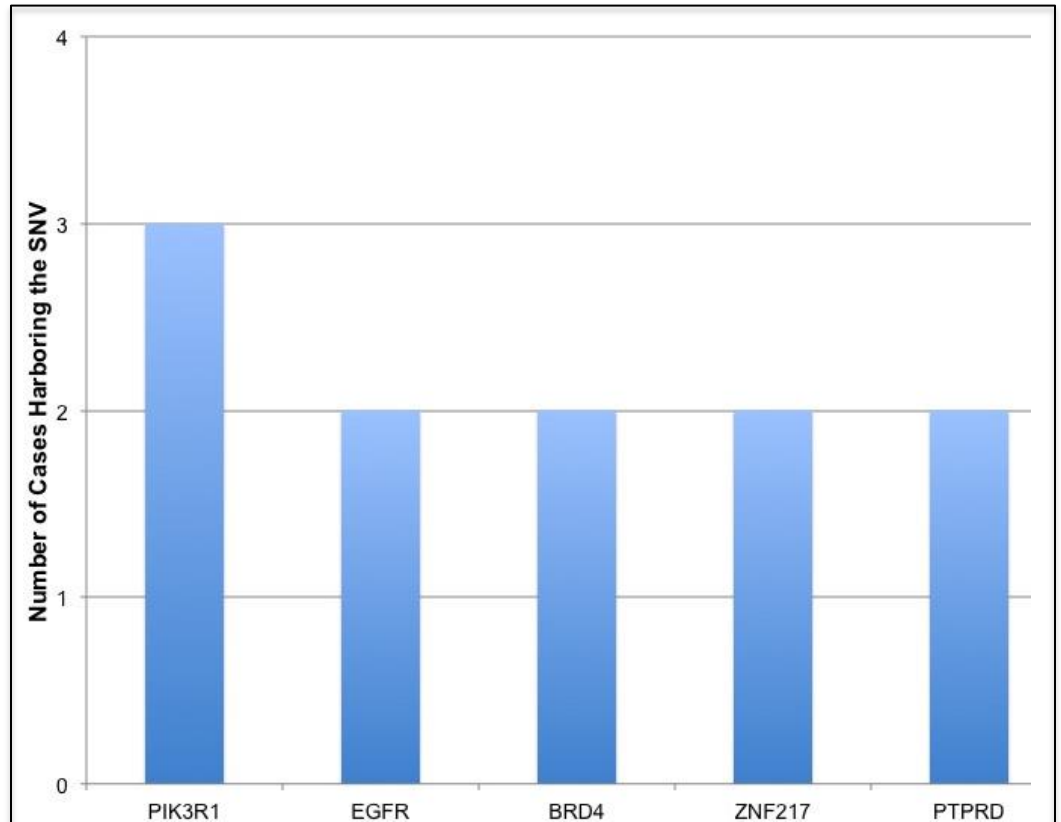


Figure 2.9: Shared SNVs Present in Fusion Negative Tumors.

All 18 IMTs undergoing RNAseq analysis were assessed for SNVs. Variants were filtered based on a list of 383 pre-defined cancer genes. The variants were maintained if they were exonic, had a damaging function, previously reported in COSMIC, and were considered pathogenic in the ClinVar Database. Variants were then excluded if they had a variant allele frequency below 0.05, depth below 10, a prevalence of greater than 10% in the 1000 Genomes Project or if no annotations were present. Depicted are genes harbouring SNVs shared by at least 50% of fusion negative tumors, beyond the 30 most commonly shared SNVs in all IMT cases. The number of tumors harboring the shared SNV is represented. Total N=4.

presence of a transmembrane domain with the shorter coiled-coil may be more favorable in terms of disorder. The role of disorder in oncogenic kinase fusions and how it relates to the findings in this study are further discussed in chapter 4.

There have been different clinical observations reported regarding pulmonary vs. extra-pulmonary tumors^{1,4,36,37}. Additionally, these tumors may bring clinical challenges due to the location, resulting in a physician needing to consider systemic therapies over surgical resection. We are able to report that, similar to IMT overall, 82% of pulmonary IMTs also were kinase fusion positive (**Table 2.2**). While this was a small cohort for comparison, all of the pulmonary fusions were either ALK or ROS1. Given the potential challenges in operative management of these tumors, and the observation that pulmonary tumors are less likely to recur^{1,36,37}, physicians may consider neo-adjuvant crizotinib treatment for patients with pulmonary IMTs when up-front resection is not possible.

Studying the IMT transcriptome presents many challenges and limitations including the gap in knowledge regarding cell of origin and tumor rarity resulting in small cohorts. The former results in an imprecise normal control for comparison. Because of this, we chose to perform comparisons by grouping our samples based on previously described clinical categories such as RANBP2-ALK harboring IMTs (likely eIMS) vs. conventional IMT, pulmonary vs. extra-pulmonary, and fusion positive vs. fusion negative. Unfortunately, due to small sample size, made even smaller when split into the comparison groupings, no genes were identified as being significant from differential expression analyses after adjusting for multiple-test correction. As IMT is rare, it can be difficult to obtain enough samples to comparatively evaluate by this method, so we chose to continue the analysis with differentially expressed genes exhibiting a $p \text{ value} \leq 0.05$ and a $FC \geq |2|$ to infer those genes that are biologically significant. Regarding differential expression between extra-pulmonary vs. pulmonary IMT, SPIA showed greater activation of the cell cycle in extra-pulmonary tumors. IMTs tend to grow quite large^{1,3} and therefore have a growing demand for nutrients. This may be better supplied to extra-pulmonary (abdominal) tumors resulting in greater activation of the cell cycle. For the comparison of RANBP2-ALK IMTs vs. all other IMTs, SPIA showed inhibition of extra-cellular matrix interactions in RANBP2-ALK tumors and activation of chemokines. This is interesting since eIMS are known to be clinically more aggressive with a tendency toward metastasis and have a different inflammatory profile^{8,10,11,47}. When assessing the signaling pathway impact of differentially expressed genes between fusion negative and fusion positive samples, we identified activation of the PPAR pathway in fusion negative samples, specifically through PPAR γ . The PPAR pathway, involved

in metabolism and inflammation^{48,49}, is targeted by some anti-diabetic drugs, and is a potential target for anti-cancer therapies⁴⁹.

Our analysis of cancer associated SNVs revealed a set of 30 variants commonly shared between at least 25% of samples. Importantly, all but two of the most commonly shared SNVs occur in both fusion negative and fusion positive samples indicating that this cohort of SNVs is likely indicative of IMT pathogenesis as a disease. Further analysis of this gene set revealed significant involvement of the VEGF pathway and regulation of chromatin (**Table S2.26**). Given the average size of IMTs^{1,3}, and therefore demand for blood supply, this is not necessarily surprising. Functions of these genes include but are not limited to histone methylation⁵⁰, chromosome segregation⁵¹, and centrosome cohesion⁵². Many of the genes in this set are involved in the p53 pathway, and *TP53* C98G (P33R) was present in 17/18 samples. A mutation in ribosomal protein RPL22 was present in all samples, although with low allele frequency. Interestingly, two of the other shared SNVs occurred in RNA helicases, *DDX31* and *DHX15*, which are involved in ribosome biogenesis⁵³. Some have postulated that stress in ribosomal biogenesis may allow for selective pressures toward cells that override pathways regulating ribosome biogenesis such as mTOR⁵⁴. Interestingly, *DDX31* is also involved in p53 regulation through interactions with NPM1⁵⁵. Taken together, this suggests IMTs are undergoing stress in ribosome biogenesis and may even suggest this as an early cancer-initiating step in IMT. Developing drugs targeting ribosome biogenesis is already underway.

Further assessment of shared SNVs in the fusion negative cohort revealed a potentially significant mutation in *PIK3R1*. According to cBioPortal, this particular mutation, M326I, has been reported in several cancers such as adrenocortical carcinoma, chondroblastic osteosarcoma, renal cell carcinoma, neuroblastoma, acute myeloid leukemia, mixed germ cell tumor, and hepatoblastoma. Deletion of this amino acid results in truncation and has been reported in glioblastoma multiforme and renal cell carcinoma. It is considered a loss of function and oncogenic^{56,57}. This mutation was present in all but one fusion negative sample and consistently at high allele frequency.

In summary, our findings support the use of NGS diagnostics to help direct therapy for IMT. Other rational therapeutic options, though not tested clinically in this tumor type, could include agents targeting angiogenesis and the VEGF pathway, such as bevacizumab, or epigenetic modulators, such as azacitidine or valproic acid. Unfortunately, there are currently no published IMT cell lines or mouse models in which to study the efficacy of these agents against IMT. In order to better understand IMT pathogenesis and study drug candidates, we must collaborate globally to work to

develop patient cell line models and overcome this limitation to better study the disease and fully understand this rare tumor.

2.6 References

1. Coffin CM, Watterson J, Priest JR, et al: Extrapulmonary inflammatory myofibroblastic tumor (inflammatory pseudotumor). A clinicopathologic and immunohistochemical study of 84 cases. *Am J Surg Pathol* 19:859-72, 1995
2. Gonzalez-Crussi F, deMello DE, Sotelo-Avila C: Omental-mesenteric myxoid hamartomas. Infantile lesions simulating malignant tumors. *Am J Surg Pathol* 7:567-78, 1983
3. Meis JM, Enzinger FM: Inflammatory fibrosarcoma of the mesentery and retroperitoneum. A tumor closely simulating inflammatory pseudotumor. *Am J Surg Pathol* 15:1146-56, 1991
4. Davare MA, Tognon CE: Detecting and targetting oncogenic fusion proteins in the genomic era. *Biol Cell* 107:111-29, 2015
5. Coffin CM, Hornick JL, Fletcher CD: Inflammatory myofibroblastic tumor: comparison of clinicopathologic, histologic, and immunohistochemical features including ALK expression in atypical and aggressive cases. *Am J Surg Pathol* 31:509-20, 2007
6. Dishop MK, Kuruvilla S: Primary and metastatic lung tumors in the pediatric population: a review and 25-year experience at a large children's hospital. *Arch Pathol Lab Med* 132:1079-103, 2008
7. Fletcher CD: The evolving classification of soft tissue tumours: an update based on the new WHO classification. *Histopathology* 48:3-12, 2006
8. Riddle NN, Gardner JM: The New Kids on the Block: Recently Characterized Soft Tissue Tumors. *Surg Pathol Clin* 8:467-91, 2015
9. Chen ST, Lee JC: An inflammatory myofibroblastic tumor in liver with ALK and RANBP2 gene rearrangement: combination of distinct morphologic, immunohistochemical, and genetic features. *Hum Pathol* 39:1854-8, 2008
10. Marino-Enriquez A, Wang WL, Roy A, et al: Epithelioid inflammatory myofibroblastic sarcoma: An aggressive intra-abdominal variant of inflammatory myofibroblastic tumor with nuclear membrane or perinuclear ALK. *Am J Surg Pathol* 35:135-44, 2011
11. Lee JC, Wu JM, Liao JY, et al: Cytopathologic features of epithelioid inflammatory myofibroblastic sarcoma with correlation of histopathology, immunohistochemistry, and molecular cytogenetic analysis. *Cancer Cytopathol* 123:495-504, 2015
12. Coffin CM, Patel A, Perkins S, et al: ALK1 and p80 expression and chromosomal rearrangements involving 2p23 in inflammatory myofibroblastic tumor. *Mod Pathol* 14:569-76, 2001
13. Griffin CA, Hawkins AL, Dvorak C, et al: Recurrent involvement of 2p23 in inflammatory myofibroblastic tumors. *Cancer Res* 59:2776-80, 1999
14. Lovly CM, Gupta A, Lipson D, et al: Inflammatory myofibroblastic tumors harbor multiple potentially actionable kinase fusions. *Cancer Discov* 4:889-95, 2014

15. Takeuchi K, Soda M, Togashi Y, et al: Pulmonary inflammatory myofibroblastic tumor expressing a novel fusion, PPFIBP1-ALK: reappraisal of anti-ALK immunohistochemistry as a tool for novel ALK fusion identification. *Clin Cancer Res* 17:3341-8, 2011
16. Takeuchi K, Soda M, Togashi Y, et al: RET, ROS1 and ALK fusions in lung cancer. *Nat Med* 18:378-81, 2012
17. Lee JC, Li CF, Huang HY, et al: ALK oncoproteins in atypical inflammatory myofibroblastic tumours: novel RRBP1-ALK fusions in epithelioid inflammatory myofibroblastic sarcoma. *J Pathol* 241:316-323, 2017
18. Antonescu CR, Suurmeijer AJ, Zhang L, et al: Molecular characterization of inflammatory myofibroblastic tumors with frequent ALK and ROS1 gene fusions and rare novel RET rearrangement. *Am J Surg Pathol* 39:957-67, 2015
19. Yamamoto H, Oda Y, Saito T, et al: p53 Mutation and MDM2 amplification in inflammatory myofibroblastic tumours. *Histopathology* 42:431-9, 2003
20. Pavlick D, Schrock AB, Malicki D, et al: Identification of NTRK fusions in pediatric mesenchymal tumors. *Pediatr Blood Cancer* 64, 2017
21. Mosse YP, Voss SD, Lim MS, et al: Targeting ALK With Crizotinib in Pediatric Anaplastic Large Cell Lymphoma and Inflammatory Myofibroblastic Tumor: A Children's Oncology Group Study. *J Clin Oncol* 35:3215-3221, 2017
22. Schoffski P, Sufliarsky J, Gelderblom H, et al: Crizotinib in patients with advanced, inoperable inflammatory myofibroblastic tumours with and without anaplastic lymphoma kinase gene alterations (European Organisation for Research and Treatment of Cancer 90101 CREATE): a multicentre, single-drug, prospective, non-randomised phase 2 trial. *Lancet Respir Med* 6:431-441, 2018
23. Cheng DT, Mitchell TN, Zehir A, et al: Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): A Hybridization Capture-Based Next-Generation Sequencing Clinical Assay for Solid Tumor Molecular Oncology. *J Mol Diagn* 17:251-64, 2015
24. Frampton GM, Fichtenholtz A, Otto GA, et al: Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat Biotechnol* 31:1023-31, 2013
25. Davidson NM, Majewski IJ, Oshlack A: JAFFA: High sensitivity transcriptome-focused fusion gene detection. *Genome Med* 7:43, 2015
26. Dobin A, Davis CA, Schlesinger F, et al: STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15-21, 2013
27. Li B, Dewey CN: RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323, 2011
28. Robinson MD, McCarthy DJ, Smyth GK: edgeR: a Bioconductor package for differential

- expression analysis of digital gene expression data. *Bioinformatics* 26:139-40, 2010
29. Tarca AL, Draghici S, Khatri P, et al: A novel signaling pathway impact analysis. *Bioinformatics* 25:75-82, 2009
 30. Li H, Handsaker B, Wysoker A, et al: The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078-9, 2009
 31. DePristo MA, Banks E, Poplin R, et al: A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43:491-8, 2011
 32. Wang K, Li M, Hakonarson H: ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38:e164, 2010
 33. Sheu GT, Traugh JA: Recombinant subunits of mammalian elongation factor 1 expressed in *Escherichia coli*. Subunit interactions, elongation activity, and phosphorylation by protein kinase CKII. *J Biol Chem* 272:33290-7, 1997
 34. Lawrence B, Perez-Atayde A, Hibbard MK, et al: TPM3-ALK and TPM4-ALK oncogenes in inflammatory myofibroblastic tumors. *Am J Pathol* 157:377-84, 2000
 35. Coffin CM, Dehner LP, Meis-Kindblom JM: Inflammatory myofibroblastic tumor, inflammatory fibrosarcoma, and related lesions: an historical review with differential diagnostic considerations. *Semin Diagn Pathol* 15:102-10, 1998
 36. Cerfolio RJ, Allen MS, Nascimento AG, et al: Inflammatory pseudotumors of the lung. *Ann Thorac Surg* 67:933-6, 1999
 37. Janik JS, Janik JP, Lovell MA, et al: Recurrent inflammatory pseudotumors in children. *J Pediatr Surg* 38:1491-5, 2003
 38. Kimbara S, Takeda K, Fukushima H, et al: A case report of epithelioid inflammatory myofibroblastic sarcoma with RANBP2-ALK fusion gene treated with the ALK inhibitor, crizotinib. *Jpn J Clin Oncol* 44:868-71, 2014
 39. Sasaki T, Okuda K, Zheng W, et al: The neuroblastoma-associated F1174L ALK mutation causes resistance to an ALK kinase inhibitor in ALK-translocated cancers. *Cancer Res* 70:10038-43, 2010
 40. Childress MA, Himmelberg SM, Chen H, et al: ALK Fusion Partners Impact Response to ALK Inhibition: Differential Effects on Sensitivity, Cellular Phenotypes, and Biochemical Properties. *Mol Cancer Res*, 2018
 41. Woo CG, Seo S, Kim SW, et al: Differential protein stability and clinical responses of EML4-ALK fusion variants to various ALK inhibitors in advanced ALK-rearranged non-small cell lung cancer. *Ann Oncol* 28:791-797, 2017
 42. Seo S, Woo CG, Lee DH, et al: The clinical impact of an EML4-ALK variant on survival following crizotinib treatment in patients with advanced ALK-rearranged non-small-cell lung cancer. *Ann Oncol* 28:1667-1668, 2017

43. Richards MW, O'Regan L, Roth D, et al: Microtubule association of EML proteins and the EML4-ALK variant 3 oncoprotein require an N-terminal trimerization domain. *Biochem J* 467:529-36, 2015
44. Richards MW, Law EW, Rennalls LP, et al: Crystal structure of EML1 reveals the basis for Hsp90 dependence of oncogenic EML4-ALK by disruption of an atypical beta-propeller domain. *Proc Natl Acad Sci U S A* 111:5195-200, 2014
45. Yoshida T, Oya Y, Tanaka K, et al: Differential Crizotinib Response Duration Among ALK Fusion Variants in ALK-Positive Non-Small-Cell Lung Cancer. *J Clin Oncol* 34:3383-9, 2016
46. Lin JJ, Zhu VW, Yoda S, et al: Impact of EML4-ALK Variant on Resistance Mechanisms and Clinical Outcomes in ALK-Positive Lung Cancer. *J Clin Oncol* 36:JCO2017762294, 2018
47. Chen Y, Takita J, Choi YL, et al: Oncogenic mutations of ALK kinase in neuroblastoma. *Nature* 455:971-4, 2008
48. Tyagi S, Gupta P, Saini AS, et al: The peroxisome proliferator-activated receptor: A family of nuclear receptors role in various diseases. *J Adv Pharm Technol Res* 2:236-40, 2011
49. Vella V, Nicolosi ML, Giuliano S, et al: PPAR-gamma Agonists As Antineoplastic Agents in Cancers with Dysregulated IGF Axis. *Front Endocrinol (Lausanne)* 8:31, 2017
50. Park IY, Powell RT, Tripathi DN, et al: Dual Chromatin and Cytoskeletal Remodeling by SETD2. *Cell* 166:950-962, 2016
51. De La Fuente R, Baumann C, Viveiros MM: Role of ATRX in chromatin structure and function: implications for chromosome instability and human disease. *Reproduction* 142:221-34, 2011
52. Bahe S, Stierhof YD, Wilkinson CJ, et al: Rootletin forms centriole-associated filaments and functions in centrosome cohesion. *J Cell Biol* 171:27-33, 2005
53. Martin R, Straub AU, Doebele C, et al: DExD/H-box RNA helicases in ribosome biogenesis. *RNA Biol* 10:4-18, 2013
54. Goudarzi KM, Lindstrom MS: Role of ribosomal protein mutations in tumor development (Review). *Int J Oncol* 48:1313-24, 2016
55. Fukawa T, Ono M, Matsuo T, et al: DDX31 regulates the p53-HDM2 pathway and rRNA gene transcription through its interaction with NPM1 in renal cell carcinomas. *Cancer Res* 72:5867-77, 2012
56. Gao J, Aksoy BA, Dogrusoz U, et al: Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 6:p11, 2013
57. Cerami E, Gao J, Dogrusoz U, et al: The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2:401-4, 2012

Chapter 3

ALK FUSION VARIANTS

Adapted from: Merrida A. Childress, Stephen M. Himmelberg, Huiqin Chen, Wanleng Deng, Michael A. Davies, Christine M. Lovly, ALK Fusion Partners Impact Response to ALK Inhibition: Differential Effects on Sensitivity, Cellular Phenotypes, and Biochemical Properties. *Mol Cancer Res.* 2018

3.1 Abstract

Oncogenic tyrosine kinase fusions involving the anaplastic lymphoma kinase (ALK) are detected in numerous tumor types. Although more than 30 distinct 5' fusion partner genes have been reported, treatment of *ALK*-rearranged cancers is decided without regard to which 5' partner is present. There is little data addressing how the 5' partner affects the biology of the fusion or responsiveness to ALK tyrosine kinase inhibitors (TKIs). Based on the hypothesis that the 5' partner influences the intrinsic properties of the fusion protein, cellular functions that impact oncogenic potential, and sensitivity to ALK TKIs, clonal 3T3 cell lines stably expressing seven different ALK fusion variants were generated. Biochemical and cellular assays were used to assess the efficacy of various ALK TKIs in clinical use, transformative phenotypes, and biochemical properties of each fusion. All seven ALK fusions induced focus formation and colonies in soft agar, albeit to varying degrees. IC50s were calculated for different ALK TKIs (crizotinib, ensartinib, alectinib, lorlatinib) and consistent differences (5-10 fold) in drug sensitivity were noted across the seven ALK fusions tested. Finally, biochemical analyses revealed negative correlations between kinase activity and protein stability. These results demonstrate that the 5' fusion partner plays an important biological role that affects sensitivity to ALK TKIs.

Implications

This study shows that the 5' ALK fusion partner influences ALK TKI drug sensitivity. As many other kinase fusions are found in numerous cancers, often with overlapping fusion partners, these studies have ramifications for other kinase-driven malignancies.

3.2 Introduction

Genomic rearrangements involving the gene which encodes anaplastic lymphoma kinase (*ALK*) have been described in a broad spectrum of malignancies including anaplastic large cell lymphoma (ALCL), diffuse large B cell lymphoma, inflammatory myofibroblastic tumor (IMT), glioma, non-small cell lung cancer (NSCLC), colorectal, breast, ovarian, and esophageal cancer⁶. *ALK* rearrangements

have been identified in up to 8% of NSCLC cases⁷ and in up to 50% of IMTs, a soft tissue tumor predominantly diagnosed in children⁸. The resulting ALK fusion proteins all retain the entire kinase domain of ALK at the C-terminus, and the N-terminus consists of an entirely different protein. These fusion proteins are validated therapeutic targets. Several large international trials have now validated that patients with ALK positive (ALK+) lung cancer derive improved clinical outcomes from treatment with ALK TKIs (**Supplementary Table S3.1**), leading to FDA approval of agents such as crizotinib, ceritinib, alectinib, brigatinib, and lorlatinib. Similar trials have also been completed in ALK+ IMT and ALCL⁹, and there have been case reports of response to ALK TKI therapy in patients with renal cell carcinoma and colon carcinoma harboring *ALK* fusions¹⁰.

Although ALK fusions are validated targets for cancer therapy, precision regarding how to best target these fusions clinically is lacking in comparison to other oncogenic drivers like mutant Epidermal Growth Factor Receptor (EGFR) or retinoic acid receptor alpha (*RARα*) fusions, where it is known that different variants dictate drug sensitivity^{11,12}. For example, in *EGFR*-mutant NSCLC, it is known that different *EGFR* kinase domain mutations confer varying degrees of sensitivity or resistance to EGFR directed therapies¹¹. Likewise, for *RARα* fusions found in subtypes of leukemia, it is known that the particular gene fused to *RARα* not only affects response to therapy but also can be a therapeutic target itself¹². Greater than 30 distinct *ALK* fusion partners have been identified, including *TPM-3/-4*, *CLTC*, *LMNA*, *PRKAR1A*, *EML4*, *RANBP2*, *TFG*, *FN1*, *KIF5B*, and many others^{5,8}. Although many different 5' partner genes have been reported for ALK, even within the same cancer type, there is currently little data to address the question of how different fusion partners may affect pretreatment clinical characteristics, disease responsiveness to targeted therapies, or acquired resistance. Therefore, treatment of ALK+ cancers is currently decided without considering which fusion partner is present. However, recent retrospective clinical studies in ALK+ NSCLC suggest a difference in progression-free survival based on the specific *ALK* fusion variant present¹³⁻¹⁵. As next-generation sequencing technologies continue to be approved by regulatory agencies, clinicians will know both the 5' partner and the 3' kinase involved in the fusion. Therefore, it is imperative that the nuances of the various ALK fusions are better understood, including determining the therapeutic implications of the various ALK fusions to bring more precision to patient care. Herein, we sought to test the hypothesis that the 5' ALK fusion partner influences the intrinsic properties of the fusion protein as well as the cellular functions that impact overall oncogenic potential and sensitivity to ALK targeted therapy.

3.3 Materials and Methods

Cell Culture: NIH 3T3 cells were a kind gift from Dr. William Pao¹⁶. NIH3T3 cells stably expressing ALK F1174L were a kind gift of Dr. Marc Ladanyi¹⁷. Cell line authentication was not performed after receipt of the cells. NIH3T3 cells were maintained in DMEM (Mediatech, Corning, NY, USA), supplemented with 10% FBS (Atlanta Biologicals, Flowery Branch, GA, USA), and penicillin (100U/mL)/streptomycin (100µg/mL) (Mediatech) at 37°C, 5% CO₂ for maintenance culture and all experiments unless noted otherwise. Plat GP cells were obtained from Cell Biolabs and cultured in RPMI 1640 (Mediatech), supplemented with 10% FBS (Atlanta Biologicals), and 2µg/mL blasticidin (Invitrogen, Grand Island, NY, USA). All cell lines were routinely evaluated for mycoplasma contamination. The latest date these cell lines were tested was September 2017.

Expression Constructs cDNAs for *EML4-ALK E13:A20* (variant 1, V1), *EML4-ALK E6a/b:A20* (variant 3, V3), *KIF5B-ALK*, *TFG-ALK*, *PRKAR1A-ALK*, *FN1-ALK*, *RANBP2-ALK*, and wild-type (WT) *ALK* receptor were synthesized by GeneArt, (ThermoFisher, Grand Island, NY, USA) and subcloned into the pMXs-Puro retroviral vector (Cell Biolabs, San Diego, CA, USA) using EcoRI and NotI or PacI. Genomic breakpoints we previously reported for *TFG-ALK*, *PRKAR1A-ALK*, and *RANBP2-ALK* were used to generate these cDNAs⁸. The *EML4-ALK V1*, *EML4-ALK V3*, *KIF5B-ALK*, and *FN1-ALK* cDNA sequences were generated using previously described breakpoints¹⁸⁻²¹.

Viral Transduction and Clonal Selection Retroviral vectors were individually transfected into the Plat GP packaging cell line (HEK293 cells stably expressing a gag-pol internal ribosome entry site)(Cell Biolabs). Viral media was harvested 48 hours post transfection and pelleted. The viral pellet was re-suspended in DMEM and applied to NIH 3T3 cells. Transduced NIH 3T3 cells were treated with 2µg/mL puromycin (Invitrogen) beginning at 48 hours post transduction for a minimum of two weeks. Single cell clones were grown until confluent in a 6-well plate at which time half of the cells were harvested for lysate, and half were frozen and stored for later culture. Lysate for each clone was run on the same gel and probed for total ALK expression (Cell Signaling, #3333). Clones exhibiting relatively equal expression for all ALK fusions were selected for expansion, except *EML4-ALK V3*, for which the lowest expressing clone of 20 tested (data not shown) was selected. Cells were maintained in DMEM, supplemented with 10% FBS, penicillin (100U/mL)/streptomycin (100µg/mL), and 2µg/mL puromycin at 37°C, 5% CO₂ for maintenance culture and experiments unless noted otherwise. Cells were cultured and used for experiments for up to 12 weeks after thawing. All cell lines were routinely evaluated for mycoplasma contamination. The latest date these cell lines were tested was September

2017.

Immunoblot and Antibodies The following primary antibodies were obtained from Cell Signaling Technology (Danvers, MA, USA): ALK mAb Rabbit (#3333), ALK mAb Mouse (3791S), ALK (D5F3) mAb Rabbit (#3633) and pALK Y1604 (#3341S). The actin antibody (A2066) was purchased from Sigma-Aldrich (St. Louis, MO, USA). The following secondary antibodies were obtained from LiCor (Lincoln, NE, USA): IRDye 680RD Goat anti-Mouse (#92668070), and IRDye 800 Goat anti-Rabbit (#92632211). For infrared (IR) immunoblot, cells were harvested, washed in PBS, and lysed in KRAS buffer [50 mM Tris-HCl, 150 mM NaCl, 5 mM MgCl₂, 1% triton X-100, 0.5% Na Deoxy Cholate, 0.1% SDS, 40 nmol/L sodium fluoride, 1mM sodium orthovanidate, and complete protease inhibitors (Roche Diagnostics, Indianapolis, IN, USA)]. Signal detection was obtained using the LiCor Odyssey scanner system and quantified using iMageStudio Lite software (LiCor).

Reverse Phase Protein Lysate Microarrays (RPPA) NIH 3T3 cell lines stably expressing ALK fusion variants, WT ALK, ALK F1174L, and empty vector control cells were grown in normal culture conditions. Lysates were harvested using the lysis buffer recommended by the MD Anderson RPPA core [1% Triton X-100, 50 mM HEPES, pH 7.4, 150 mM NaCl, 1.5 mM MgCl₂, 1 mM EGTA, 100 mM NaF, 10 mM Na pyrophosphate, 1mM sodium orthovanidate, 10% glycerol, containing freshly added protease and phosphatase inhibitors]. Protein concentrations were adjusted to 1.5µg/µL. Samples were prepared as instructed by the MD Anderson RPPA Core using 4xSDS sample buffer [40% Glycerol, 8% SDS, 0.25M Tris-HCL, pH 6.8. (β-mercaptoethanol added at 1/10 of the volume immediately before use)], boiled for 5 minutes, and stored at -80°C until sample submission. Three separate biological replicates for each cell line type were tested. All of the samples were assayed for the expression of 304 phospho-protein or total protein analytes using the reverse phase protein lysate microarrays (RPPA), as described (<https://www.mdanderson.org/research/research-resources/core-facilities/functional-proteomics-rppa-core/rppa-process.html>).

RPPA Analysis: The loading control normalized and log₂ transformed RPPA data were imported and analyzed using R packages. Unsupervised hierarchical clustering analysis was performed to check the data quality and the association between cell lines. The heat map shows that the empty vector transferred and wild type cell line are well separated with other cell lines.

RPPA Statistical analysis: One-Way analysis of variance (ANOVA) was used to assess the differences in protein expressions between cell lines on a feature-by-feature basis. First, for one

protein at a time, we carried out an over-all F test to detect any significant difference among the means of all the groups. Next, for the proteins identified in this process, we then compared between desired cell line groups to identify the sources of difference. The R library “multcomp” was used for this purpose. Note that the fold change (FC) values were calculated as the estimated ratio between the 2 groups in comparison, with the following conventional modification: For the ratios > 1 (up-regulation), FCs were noted as the same as the ratio. For the ratios ≤ 1 (down-regulation), FCs were noted as the negative inverse of the ratio. Furthermore, to account for multiple testing, we estimated the false discovery rates (FDR) of the overall test of the model using the Benjamini-Hochberg method. The criteria of protein selection for each pairwise comparison were: 1. Significant in overall F-test (FDR adjusted p-value <0.05); 2. Significant in pairwise comparison. (Raw p-value <0.05). 3. The FC between groups is at least 1.5.

In order to calculate pathway activity scores, RPPA data were median-centered and normalized by standard deviation across all samples for each component to obtain the relative protein level. The pathway score is then the sum of the relative protein level multiplied by its weight of all components in a particular pathway. The similar comparisons as protein level were applied to the pathway scores.

Focus Formation NIH 3T3 cell lines stably expressing ALK fusion variants were trypsinized to disrupt cell adhesions, seeded at 4000 cells/well in 24 well plates, and allowed to grow for 8 days in a cell culture incubator containing an IncuCyte ZOOM (Essen BioSciences, Ann Arbor, MI, USA). Foci were quantified manually from daily images acquired by the IncuCyte ZOOM.

Soft Agar Assays NIH 3T3 cell lines stably expressing ALK fusion variants were trypsinized to disrupt cell adhesions. 1.5mL of 0.5% agar/complete media were added to each well in a 6 well plate and allowed to solidify. Cells were seeded on top of the 0.5% agar/complete media layer at a density of 5000 cells/well in complete media containing 0.3% agar. After the agar solidified, 200 μ L of media was added on top and changed every week. Cultures were grown at 37°C in 5% CO₂ for 6 weeks. Colony formation was quantified using the GelCount system (Oxford Optronix, United Kingdom).

Inhibitors: Crizotinib, alectinib, lorlatinib, and ganetespib were purchased from Selleck Chemicals (Houston, TX, USA). Ensartinib was obtained from Xcovery (Palm Beach Gardens, FL). Cycloheximide was purchased from Sigma-Aldrich (St. Louis, MO, USA).

Drug Response Curves NIH 3T3 cell lines stably expressing ALK fusion variants were seeded at 2000 cells/well in 96 well plates using the Multidrop Combi dispenser (ThermoFisher). Cells were

treated with increasing doses of each inhibitor. 72 hours post drug treatment propidium iodide (PI) (Sigma-Aldrich) and Hoechst 33342 (Invitrogen) were added to a concentration of 0.4 μ g/mL and 4 μ g/mL respectively using the Multidrop Combi dispenser and allowed to incubate for 20 minutes at 37°C. For automated high-content imaging, the ImageXpress Micro XL (Molecular Devices, San Jose, CA) is integrated with a Thermo F3 robotic arm. Whole-well imaging was captured with a 4x objective using DAPI and Texas Red filters. Total nuclei and PI+ cells were counted using the multi-wavelength cell scoring application module in the MetaXpress software. Dose response curves, using live cell counts (total nuclei - PI+), and IC50s were generated using GraphPad Prism™ version 7 (GraphPad software, La Jolla, CA, USA).

Kinase Assays ALK variants were immunoprecipitated from NIH 3T3 cells lines stably expressing ALK fusion variants and controls using a monoclonal anti-ALK antibody (Cell Signaling #3633) Lysates were pre-cleared with Protein A Sepharose beads (Invitrogen). After 1 hour in primary antibody, Protein A Sepharose beads (Invitrogen) were added for 1 hour. Beads were washed 4x in PBS as directed in the manufacturer's protocol. Kinase assays were performed using the Universal Kinase Assay Kit (Takara, Clontech, Japan) according to the manufacturer's instructions. An aliquot of the immunoprecipitate, equal to the amount used in each *in vitro* kinase assay, was assessed by immunoblot for total ALK (Cell Signaling #3333). After completion of the *in vitro* kinase assay, the data was analyzed with a two-step normalization process: (1) Total ALK expression for each variant was calculated using iMageStudio Lite (LiCor) and normalized to (divided by) the WT ALK expression. (2) Kinase activity of each fusion was then normalized to (divided by) expression and represented as relative to WT ALK. Kinase assays were performed in the absence of known ALK ligand.

Protein Stability Assays NIH 3T3 cell lines stably expressing ALK fusion variants were seeded in 6 well plates. 24 hours after seeding, cells were treated with 50 μ g/mL cycloheximide (Sigma-Aldrich) for 2, 8, 16, or 24 hours, harvested for lysate, and probed for total ALK protein (Cell Signaling, #3191S) by IR Western blot. Total ALK variant expression was normalized to the actin loading control.

Data Analysis and Statistical Considerations Each experiment was performed with three biological replicates and repeated at least three times. IR Western blots were visualized using the LiCor Odyssey scanner system and quantified using ImageStudio Lite (LiCor), and ALK phosphorylation and total ALK protein expression were normalized to the actin loading control. Kinase assays were measured using the Synergy MX microplate reader (BioTek). All graphs and statistical analyses were

generated using GraphPad Prism™ software and a p value < 0.05 as the threshold for statistical significance. For comparing more than two conditions, the Kruskal-Wallis or Tukey tests were used.

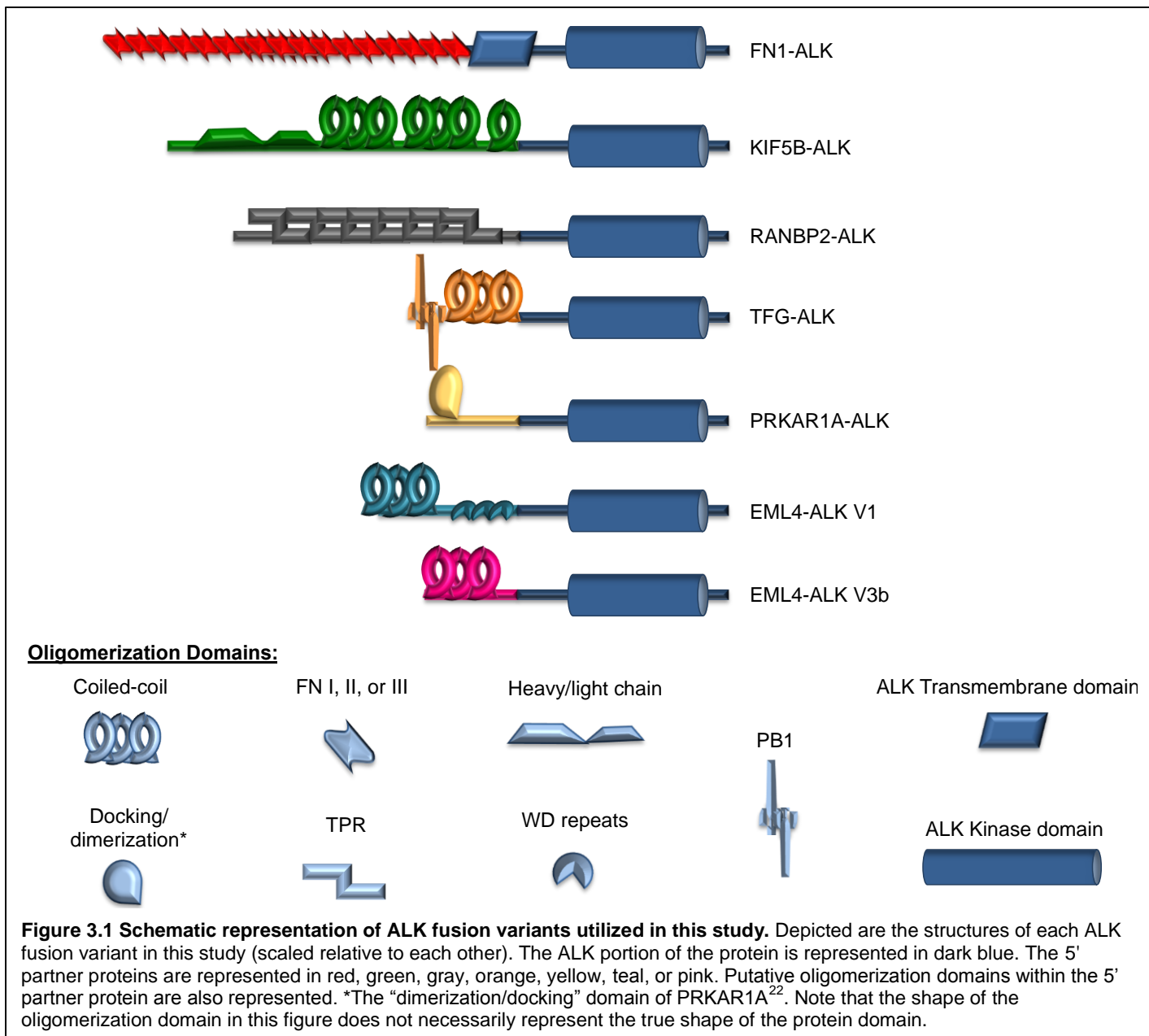
3.4 Results

Expression of ALK fusion variants in NIH3T3 cells

At present, there are limited cell lines that harbor endogenous *ALK* fusions, and these models are limited to *EML4-ALK* and *NPM-ALK*. To study the effects of multiple 5' partners, we selected seven different *ALK* fusions for exogenous expression: *FN1-ALK*, *KIF5B-ALK*, *RANBP2-ALK*, *TFG-ALK*, *PRKAR1A-ALK*, *EML4-ALK E13:A20* (variant 1, V1; exon 13 of *EML4* fused to exon 20 of *ALK*), and *EML4-ALK E6b:A20* (variant 3b, V3b, exon 6b of *EML4* fused to exon 20 of *ALK*) (**Table 3.1, Fig. 3.1**). These fusion variants have all been detected in IMT and/or NSCLC^{8,19,21}, the two tumor types with the most clinical data regarding the efficacy of ALK inhibition.

Of the greater than 30 ALK fusion partners which have been described, we selected these particular variants because they produce fusion proteins of different molecular weights which differ in protein subdomain structure, particularly with respect to the putative oligomerization domains (**Fig. 3.1**), and subcellular localizations (**Table 3.1**). Full-length ALK receptor, either WT or F1174L mutated, were used as controls. The F1174L mutation in the ALK kinase domain confers ligand-independent, constitutive activity of the receptor, representing a positive comparator for oncogenic activity while the full length WT ALK receptor is expected to be inactive unless ligand is added^{39,40}.

A challenge to consider in studying *ALK* fusion variants comparatively is that each variant is under the control of the 5' fusion partner's regulatory elements, which vary from fusion partner to fusion partner and cell type to cell type. To address this issue, all variants were placed under the control of the LTR in the PMXs-Puro retroviral vector. We stably transduced the seven distinct ALK variants (**Table 3.1, Fig. 3.1**) into NIH 3T3 cells to generate isogenic cell line models⁴¹ and selected clones in which the specific ALK fusion proteins were expressed at relatively equivalent levels (**Supplementary Fig. S3.1**). We confirmed the expression of each fusion at the expected molecular weight (**Table 3.1, Supplementary Fig. S3.1**). For FN1-ALK, we noted an additional lower molecular weight band at approximately 75kD, as previously described for this fusion variant¹⁸.



In further support of our cell line models, we performed Reverse Phase Protein Array (RPPA) in order to quantitatively evaluate changes in protein expression induced by each active ALK variant (**Fig. 3.2, Supplementary Fig. S3.2**). This extensively validated assay evaluates the relative expression of 304 cancer related proteins and phospho-proteins from each sample simultaneously (**Supplementary Table S3.1**), allowing for confident, quantitative comparisons. The most significantly changed proteins between the negative control cell lines (WT ALK, empty vector) and those expressing an ALK fusion or the constitutively active mutant, ALK F1174L are shown in **Fig. 3.2 (Supplementary Table S3.2)**. We observed a clear shift in the protein expression profile between the negative control cell lines (WT ALK, empty vector) and those expressing an ALK fusion or the constitutively active mutant, ALK F1174L (**Fig. 3.2**).

ALK Fusion Variant	Protein Length [#] (predicted MW)	Tumor Types	Reported Subcellular Localization	Reported function of the 5' fusion partner
FN1-ALK¹⁸	1799 (198kD)	IMT ⁸ , Ovarian ¹⁸	Plasma membrane, Cytoplasmic ¹⁸	Fibronectin, FN1: glycoprotein found in a dimer or multimer at the cell surface and in extracellular matrix ²³ .
KIF5B-ALK¹⁹	1481 (163kD)	NSCLC ¹⁹ , IMT*	Cytoplasmic ¹⁹	Kinesin Heavy Chain, KIF5B: microtubule-dependent motor involved in distribution of mitochondria and lysosomes and regulation of centrosome and nuclei ²⁴ .
RANBP2-ALK⁸	1427 (157kD)	IMT ⁸ , DLBCL ²⁵	Nuclear periphery, Cytoplasmic ²⁶	Ran Binding Protein, RANBP2: large scaffold and mosaic cyclophilin-related nucleoporin involved in the Ran-GTPase cycle ²⁷ .
TFG-ALK⁸	800 (88kD)	NSCLC ²⁸ , IMT ⁸ , ALCL ²⁹ , PTC ³⁰	Cytoplasmic ³¹	Tyrosine Kinase Fused Gene, TFG: localizes to the endoplasmic reticulum (ER) exit sites and modulates ER export. ³² May function at the ER/ERGIC interface to concentrate COPII-coated vesicles and link exit sites on the ER to ERGIC membranes. ³³ Inhibitory regulator of the ubiquitin-proteasome system ³⁴ .
PRKAR1A-ALK⁸	727 (80kD)	IMT ⁸	Cytoplasmic*	Protein Kinase A Type 1 Regulatory subunit, PRKAR1A: dissociated from the inactive holoenzyme upon cAMP binding ^{35,36} .
EML4-ALK V1²¹	1057 (116kD)	NSCLC ²¹	Cytoplasmic ³⁷	Echinoderm Microtubule Associated Protein-Like 4, EML4: WD-repeat protein possibly affecting microtubule formation and dynamics ³⁸ .
EML4-ALK V3b²⁰	781 (86kD)	NSCLC ²⁰	Cytoplasmic ³⁷	Echinoderm Microtubule Associated Protein-Like 4, EML4: WD-repeat protein possibly affecting microtubule formation and dynamics ³⁸ .

Table 3.2: Properties of the ALK fusion variants described in this study. This table details the ALK fusions chosen for this study, published functions of the 5' partner protein, protein length, expected molecular weight, cancers in which the fusion variant has been reported, and published subcellular localization of the ALK fusion protein. Genomic breakpoints used to synthesize the cDNA's that generate these proteins came from literature reports and are cited in the ALK Variant column. [#]Protein length represented as the number of amino acids. *Data from findings in this study.

ALK fusion partners confer differential oncogenic properties

Next, we sought to determine if the 5' partner affects transforming and proliferative properties. Using the NIH 3T3 cell line models, we first assessed focus formation, an established assay used to evaluate the ability of oncogenes to overcome contact inhibition. As expected, cells expressing empty vector and WT ALK did not support focus formation, while the ALK F1174L cells (the positive control) lost contact inhibition and were able to form foci (**Fig. 3.3A**). We observed that all seven distinct ALK fusion variants also supported focus formation, albeit to different degrees. Notably, FN1-ALK exhibited the most robust ability to form foci, more than double the amount of the positive control (ALK F1174L). Similarly, when assessed for anchorage independent growth (**Fig. 3.3B**), all fusions produced colonies in soft agar. However, only FN1-ALK produced significantly more colonies than the

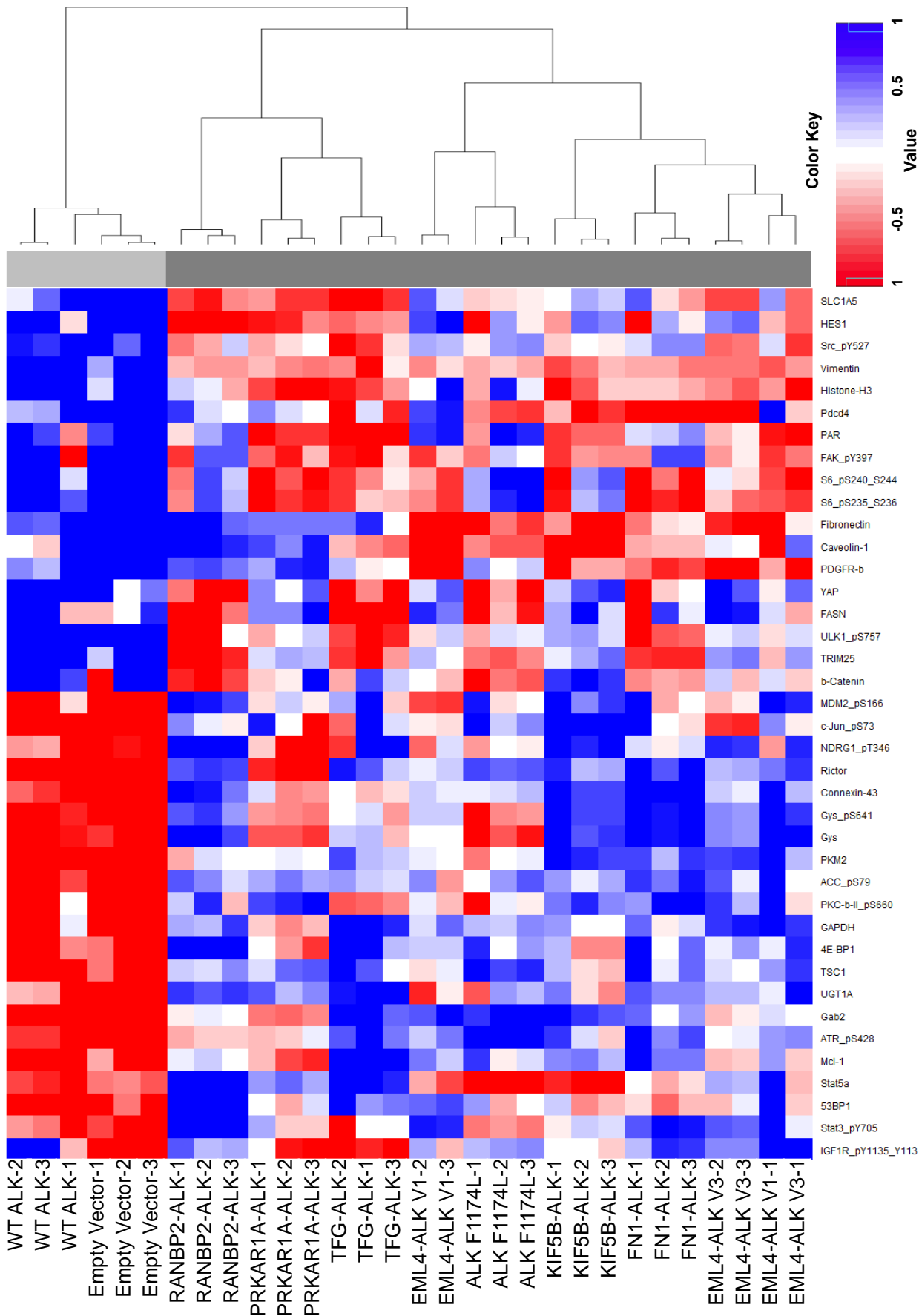
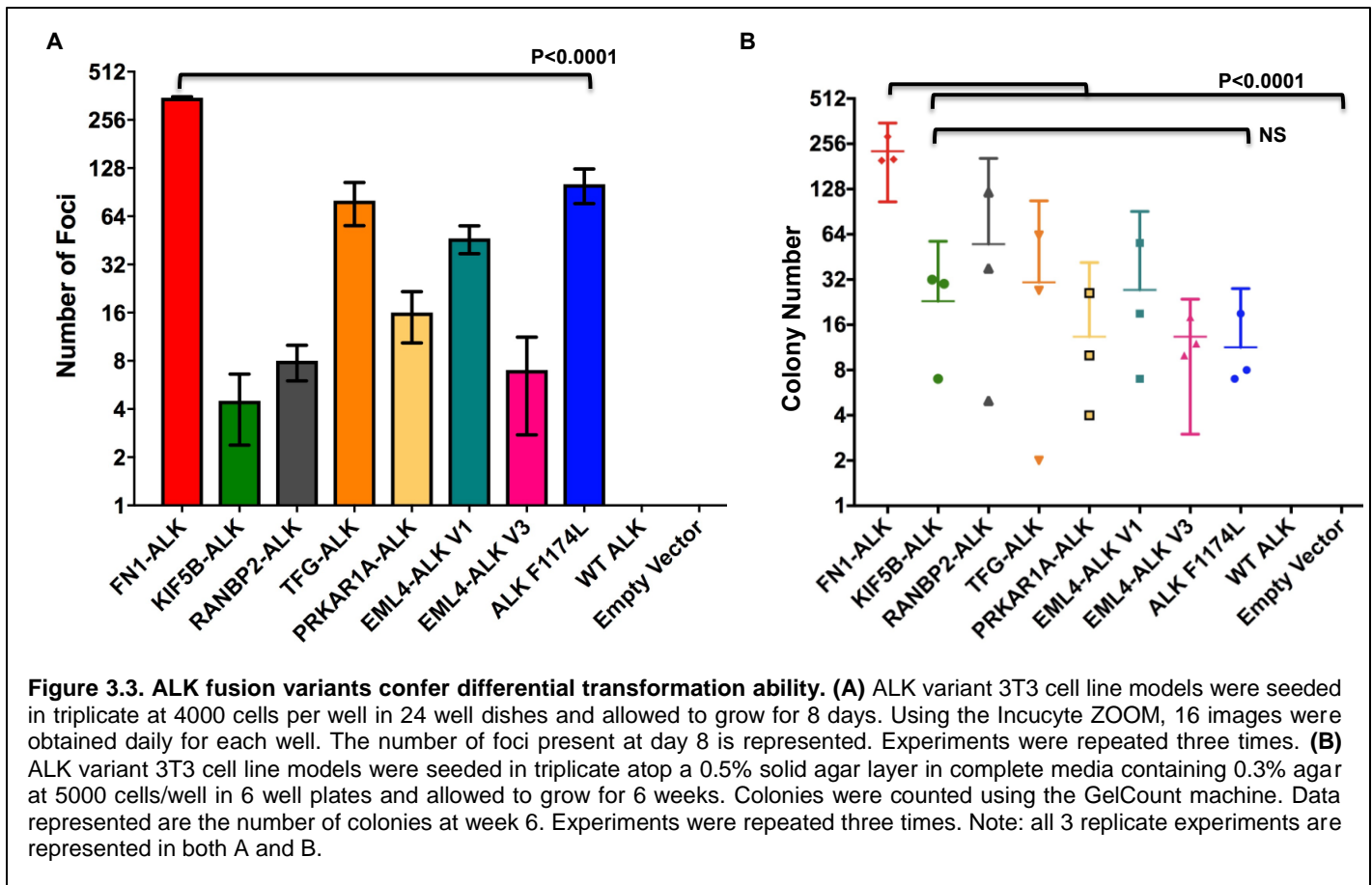


Figure 3.2. Active ALK variants change the protein expression profile of NIH 3T3 cells. This heatmap depicts those proteins that are the most significantly differentially expressed between the WT ALK and Empty Vector cell lines (group1-light gray) and all other cell lines which expression active ALK variant (group2-dark gray). This data was generated using a Reverse Phase Protein Array (RPPA). The experiment was performed using biological replicates run in triplicate. Note that blue is representative of increased expression (from the mean), and red is decreased expression (from the mean). A hierarchical tree is depicted on the right to further distinguish delineation of the cell lines. The protein names and phosphorylation sites are represented at the bottom.

other fusions (**Fig. 3.3B**). Like ALK F1174L, FN1-ALK retains the trans-membrane domain of ALK (**Fig. 3.1**) – due to an alternative breakpoint within the *ALK* locus - and exhibits similar doubling time (data not shown), yet FN1-ALK forms significantly more foci and more colonies in soft agar than ALK F1174L, suggesting that properties belonging to FN1 specifically are likely responsible for this hyperactive phenotype. These findings reaffirm the hypothesis that the 5' fusion partner plays a role in oncogenic properties.



ALK fusion variants exhibit differential response to ALK tyrosine kinase inhibition

Previous retrospective clinical studies have suggested that tumors harboring different ALK fusion variants may display different responses to ALK TKI therapy¹³⁻¹⁵. It is unclear from a biochemical perspective why this is the case since each ALK fusion protein contains the entire ALK tyrosine kinase domain, and all of the ALK TKIs in clinical use function as ATP mimetics. The efficacy of ALK TKIs against different ALK variants, differing only by the 5' fusion partner, has not been directly compared *in vitro*. We generated eight-point dose response curves to evaluate the efficacy of crizotinib, the first FDA-approved ALK TKI, against each ALK fusion variant. 3T3 cells stably

transduced with empty vector, full length WT ALK receptor, or full length ALK receptor encoding known activating mutation, F1174L, were used as controls. We observed a dose dependent decrease in cell viability across all cell lines (**Fig. 3.4A, Supplementary Fig. S3.3**). However, we noted that the efficacy of crizotinib was not uniform across all fusions tested. To verify on-target effects of crizotinib, we assessed ALK auto-phosphorylation by immunoblot (**Supplementary Fig. S3.3**). We calculated IC50 values for crizotinib against each ALK variant cell line and found a > 5 fold difference in responsiveness (**Fig. 3.4C**). Cells expressing TFG-ALK exhibited the lowest IC50 (87.36 nM), while cells expressing PRKAR1A-ALK exhibited the highest IC50 (461.8 nM) (**Fig. 3.4A, C**).

To determine if this was a drug class effect, we also investigated the efficacy of second generation ALK TKIs, alectinib and ensartinib, as well as the third generation ALK TKI, lorlatinib (**Fig. 3.4B, Supplementary Fig. S3.3, Supplementary Table S3.3**). These inhibitors are known to be more potent against the target (ALK) compared to the first generation ALK TKI, crizotinib⁴²⁻⁴⁴. Notably, there was a >10 fold difference in IC50s across the fusions for second and third generation ALK TKIs and less efficacy against the empty vector control cells. When comparing across inhibitors, PRKAR1A-ALK was consistently the least sensitive fusion (highest IC50 values) to all 4 distinct ALK TKIs evaluated. However, we noted that the order of fusion sensitivity did not remain consistent for the other 6 ALK fusion variants (**Fig. 3.4C**). For instance, KIF5B-ALK was highly sensitive to ensartinib (IC50 24.3 nM) but was one of the least sensitive fusions to crizotinib (IC50 405.1 nM) and lorlatinib (IC50 89.1 nM). These data support the hypothesis that drug efficacy is influenced by the specific fusion partner present and also highlight the importance of selecting the “right” ALK inhibitor for each fusion.

Tyrosine kinase activity of ALK fusion variants

Many structural and biochemical properties of the fusion proteins have the potential to contribute to differences in response to ALK TKIs that we observed (**Fig. 3.4**), including dimerization/oligomerization, intrinsic kinase activity, and protein stability. Previous studies have shown that the fusion partner contributes a dimerization domain to assist in auto-activation of the kinase. For example, EML4 contains a coiled-coiled domain in the N-terminus, and when this coiled-coil domain is abrogated through mutation, the resultant EML4-ALK fusion loses transforming properties²¹. Many fusion partners contain coiled-coil domains or other known dimerization domains; however, not all fusion partners have an obvious dimerization motif. Similar to other tyrosine kinases, phospho-ALK to total ALK ALK must dimerize to auto-activate and signal downstream^{21,45,46}. Pathway analysis of the RPPA data, comparing WT ALK to each ALK fusion variant or ALK F1174L individually, revealed significant up-regulation of the RAS/MAPK pathway in FN1-ALK, KIF5B-ALK,

RANBP2-ALK, and ALK F1174L (**Supplementary Table S3.4**). When assessing the auto-phosphorylation site of ALK, a surrogate for enzymatic activity of the ALK tyrosine kinase domain, we observed differences in the ratio of from one fusion to the next (**Fig. 3.5A**). Interestingly, those variants, which exhibited a pALK/ALK ratio of 0.5 or greater, also had significant RAS/MAPK pathway activation (**Supplementary Table S3.4**).

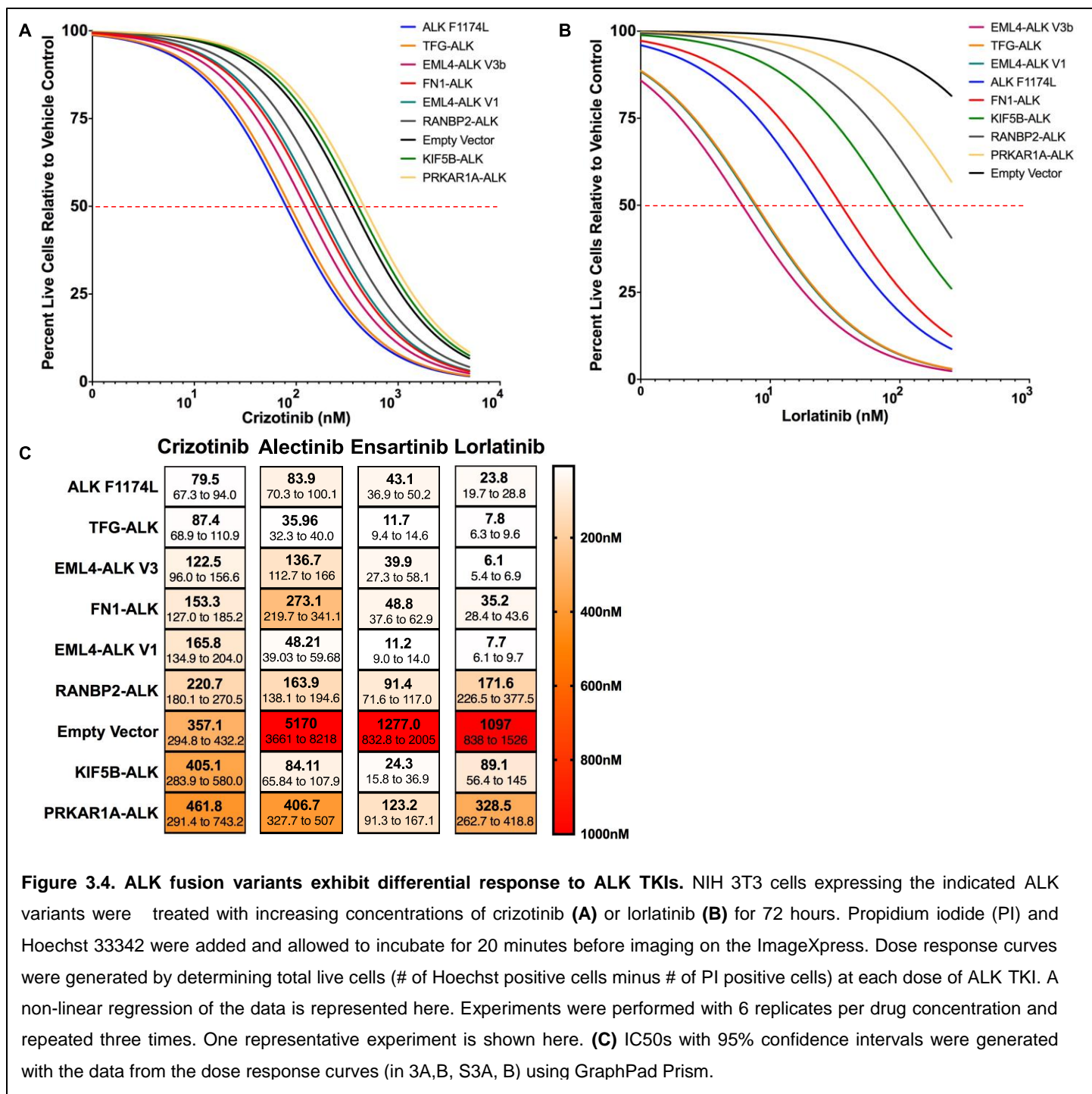


Figure 3.4. ALK fusion variants exhibit differential response to ALK TKIs. NIH 3T3 cells expressing the indicated ALK variants were treated with increasing concentrations of crizotinib (**A**) or lorlatinib (**B**) for 72 hours. Propidium iodide (PI) and Hoechst 33342 were added and allowed to incubate for 20 minutes before imaging on the ImageXpress. Dose response curves were generated by determining total live cells (# of Hoechst positive cells minus # of PI positive cells) at each dose of ALK TKI. A non-linear regression of the data is represented here. Experiments were performed with 6 replicates per drug concentration and repeated three times. One representative experiment is shown here. (**C**) IC₅₀s with 95% confidence intervals were generated with the data from the dose response curves (in 3A,B, S3A, B) using GraphPad Prism.

It is currently unknown if different fusions vary with respect to kinase activity, but structural differences between the various 5' partners, including the oligomerization domains (**Fig. 3.1**), could alter the

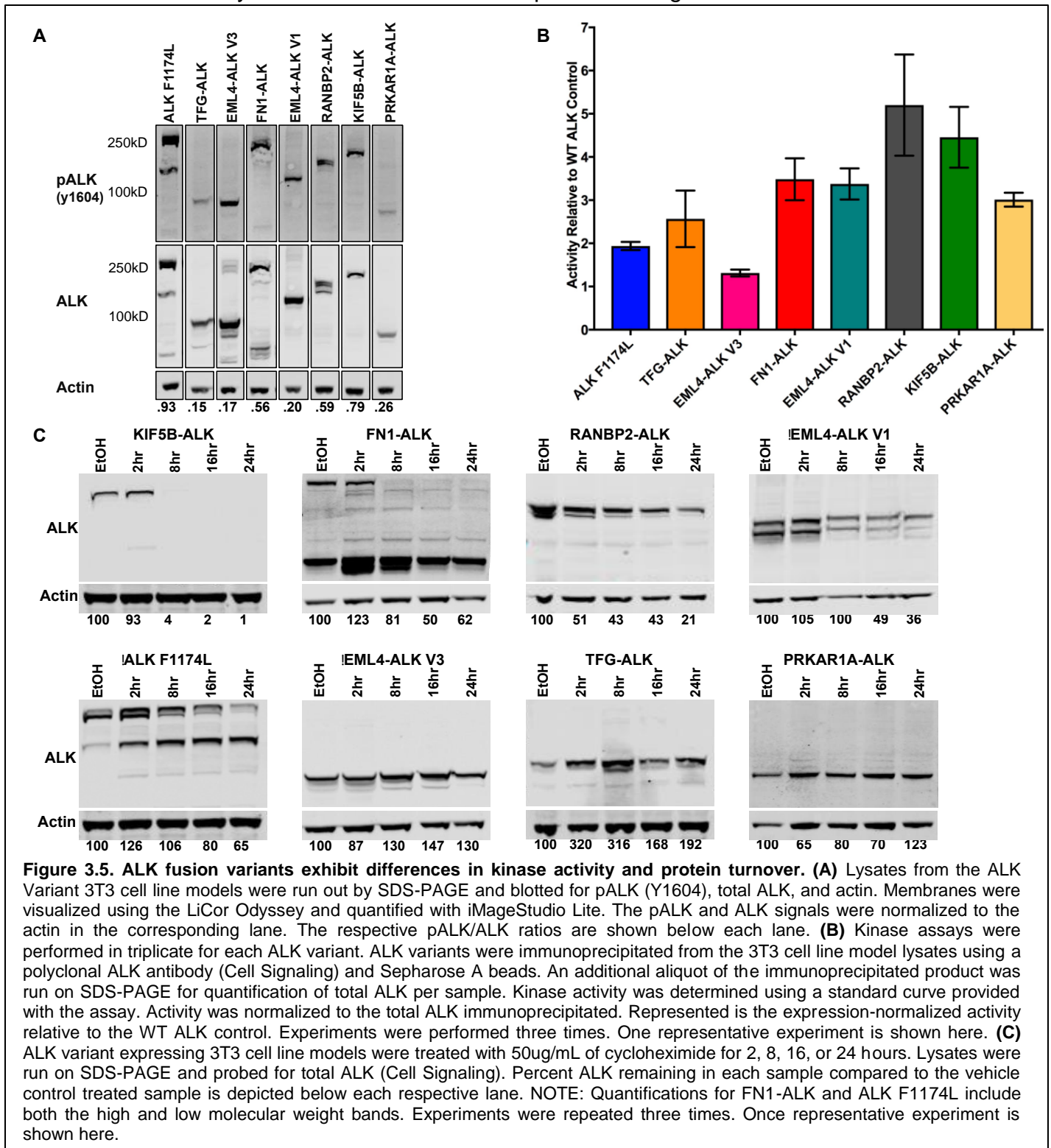
dimerization ability and therefore activation dynamics of the fusion and/or accessibility of the catalytic domain in ALK. To further evaluate relative kinase activity, we immunoprecipitated the ALK fusions and measured the enzymatic activity against an exogenous peptide substrate, using previously validated assays^{14,17}. Controlling for the amount of protein immunoprecipitated, we calculated the kinase activity for each variant relative to WT ALK (**Fig. 3.5B, Supplementary Fig. S3.4**). As expected, the known activating mutation, ALK F1174L, exhibited two-fold greater kinase activity compared to WT ALK. We also observed that all 7 ALK fusion proteins possessed *in vitro* kinase activity, albeit to different extents. RANBP2-ALK and KIF5B-ALK exhibited the highest relative activity; approximately 5 fold greater than the activity of WT ALK. These results are notable because the portion of ALK in each of these fusion proteins is the same, with the exception of FN1, which contains the ALK trans-membrane domain in addition to the ALK kinase domain and C terminus. Interestingly, kinase activity towards an exogenous substrate (**Fig. 3.5B**) did not directly correlate with ALK auto-phosphorylation ratios (**Fig. 3.5A**).

Stability of ALK fusion proteins

Next, we sought to evaluate the stability or turnover rates of the seven distinct ALK fusions⁴⁷. We performed a time course study to assess fusion protein stability by treating cells with cycloheximide to inhibit new protein synthesis. We observed marked differences in the turnover rates of the various ALK fusion proteins (**Fig. 3.5C**). Levels of EML4-ALK V3, PRKAR1A-ALK, and TFG-ALK did not decrease over the cycloheximide time course. In contrast, KIF5B-ALK, FN1-ALK, RANBP2-ALK, and EML4-ALK V1 all showed a time dependent decrease in ALK protein levels, with KIF5B-ALK having the highest turnover rate. To ensure that the rapid degradation of KIF5B-ALK and FN1-ALK was not a result of cytotoxicity, we performed the experiment with shorter time intervals. As expected, we saw a gradual decrease in total ALK over time (**Supplementary Fig. S3.5A**). Overall, we were able to confirm a negative correlation between kinase activity and protein stability at 16 hours post cycloheximide treatment (**Supplementary Fig. S3.5B**). While we observed a negatively correlative pattern between protein stability and crizotinib sensitivity, it did not reach statistical significance (**Supplementary Fig. S3.5C**).

Previous studies have shown EML4-ALK to be a client of Heat Shock Protein 90 (HSP90), and HSP90 inhibitors have been in clinical trials for ALK fusion NSCLC48. In light of our cycloheximide results, we thought the FN1-ALK, KIF5B-ALK, and RANBP2-ALK may show significant sensitivity to HSP90 inhibition. Therefore, we challenged our cell lines with ganetespib, an HSP90 inhibitor in clinical use. Surprisingly, only EML4-ALK variant 1 and 3 were sensitive to HSP90 inhibition

(Supplementary Fig. S3.5D). This suggests that the instability or turnover of FN1-ALK, KIF5B-ALK, and RANBP2-ALK may not be due to difficulties in protein folding.



3.5 Discussion

While clinical results with ALK TKIs are promising, more granular data is needed to increase

therapeutic precision. Studies involving point mutations in the ALK kinase domain that confer resistance to ALK TKI therapy have revealed that different point mutations confer varying degrees of sensitivity and resistance to the numerous ALK TKIs in clinical use⁴⁹. However, our understanding of how the 5' partner protein plays a role in therapeutic response is lacking. As next-generation sequencing technologies continue to come to the forefront of clinical diagnostics, clinicians will not only know the mutational status of the ALK kinase domain but also the 5' partner involved in the fusion. Therefore, it is vital that we lay the foundation for determining the therapeutic implications of the 5' protein.

Our studies demonstrate that the 5' partner affects the biochemical and cellular properties of the ALK fusion protein, including kinase activity, protein stability, transformative potential, and – importantly – response to ALK TKIs. For example, FN1-ALK exhibited the greatest ability to form foci and colonies in soft agar. Of note, FN1-ALK is the only ALK fusion studied which retains the ALK trans-membrane domain (**Fig. 3.1**). However, it is unlikely that the retention of this domain alone results in increased colony formation because the full length ALK receptor harboring the constitutively active F1174L mutation retains the ALK trans-membrane domain but did not display the same level of transformative ability. Two cases of EML4-ALK fusions, which retain the trans-membrane domain, have been reported^{50,51}. Given that EML4 is not a secreted protein with known extracellular interactions³⁸, it is unknown how this non-canonical breakpoint may affect transformation in this case.

In our biochemical analyses, we identified fusions which exhibited more rapid protein turnover than others, yet these data did not correlate with sensitivity to HSP90 inhibition. We noted that the fusions, which exhibited the shortest half-life (KIF5B-ALK, FN1-ALK, RANBP2-ALK, and EML4-ALK V1), also had the highest kinase activity. While ALK trafficking biology is not completely understood, especially in the context of ALK fusions, it is known that many RTKs will turnover upon activation. One study of the full length ALK receptor has shown that ALK undergoes lysosomal degradation upon activation rather than receptor recycling⁵². This could be a potential explanation for the differential degradation times of the fusions.

Furthermore, we evaluated the efficacy of four distinct and clinically relevant ALK TKIs across the seven fusion variants. At present, the assumption in the field remains that since all ALK fusion proteins retain the entire ALK kinase domain, that the efficacy of ALK TKI therapy will be analogous across all the fusions. However, our data point to a different conclusion. We consistently noted a 5-10 fold difference in the measured IC50 value for the four ALK TKIs – crizotinib, alectinib, ensartinib, and

lorlatinib – against the seven distinct fusions tested, supporting our hypothesis that the 5' partner protein can affect response to ALK inhibition (**Fig. 3.4A, Supplementary Fig. S3.3**). Indeed, IMT patients harboring the RANBP2-ALK fusion, which is less sensitive to ALK TKI in our study, also seem to exhibit more aggressive disease and less sensitivity to ALK TKI therapy^{26,53}. There are many structural and biochemical properties of the fusion proteins which have the potential to contribute to differences in response to ALK TKIs, including: 1) the type of oligomerization domains present; 2) the stoichiometry of oligomerization: dimers,⁵⁴ trimers,^{37,55} and multimers^{56,57}; 3) the intrinsic kinase activity of the fusion protein¹⁷; 4) protein-protein interactions, which may differ based on the domain structure of the fusion partner; 5) length of the 5' partner; 6) protein folding, tertiary structure, and degree of disorder; and 7) protein stability. Additional in silico modeling coupled with rigorous model systems will be necessary to dissect these critical issues.

Kinase fusions are known to localize differently based on the 5' partner (**Table 3.1**) even for those, which share the same 5' partner (and therefore promoter) but vary in the amount of that protein present in the fusion, such as the EML4-ALK variants^{37,58}. This could allow for different protein interactions. Differential pathway activation based on ALK localization has been shown in Medves et al. 2011. Additionally, in Armstrong et al. 2004 where investigators explored phenotypes of 5 different ALK fusions in NIH 3T3 cell lines, differences in pathway activation were also seen. Our RPPA data clearly shows significant changes in total and phospho-protein expression profiles when an active ALK variant is expressed (**Figure 3.2**) and also shows expression differences between the ALK variants (**Figure 3.2, Supplementary Figure S3.2, and Supplementary Tables 3.2-3.3**). The RPPA data was revisited throughout the study in search for insight into to different phenotypes assessed. The only correlation that we found was that those variants, which exhibited a pALK/ALK ratio of 0.5 or greater, also had significant RAS/MAPK pathway activation. While the RPPA panel at the MD Anderson core includes many of the most highly implicated cancer associated proteins, the panel is, by definition, limited in scope. Therefore, if there are other proteins (or additional phosphorylation sites) that are playing important roles in the phenotypes reported in this study, they may be missed. There is also a similar bias in the pathway analysis pipeline in that there are pathways defined that are mostly confined or well understood in other adult solid tumors such as breast cancer. This pathway analysis pipeline also does not include pathway definitions for other important cancer phenotypes such as migration/metastasis and metabolism.

Finally, we noted that the relative sensitivity (in terms of ranked IC50 value) of a given fusion was not consistent across all inhibitors tested. For example, while KIF5B-ALK was one of the least sensitive

fusions towards crizotinib, this same fusion was one of the most sensitive towards ensartinib (**Fig. 3.1C**), suggesting that drug selection may necessitate knowledge of the specific fusion partner in order to select the most efficacious inhibitor amongst the several FDA-approved or investigational ALK TKIs in clinical use (**Supplementary Table S3.3**). In support of these data, two recent retrospective clinical analyses have shown that NSCLC patients whose tumors harbor different *EML4-ALK* fusion variants – defined by the amount of *EML4* contained within the fusion (the *ALK* portion is the same) – experience different responses upon treatment with distinct ALK TKIs. In one study, NSCLC patients with tumors harboring *EML4-ALK* variant 3a/b exhibited longer progression free survival (PFS) when treated with lorlatinib compared to patients whose tumors harbored *EML4-ALK* variant 1. However, the reciprocal responses were observed with crizotinib¹⁵. Overall, our data, coupled with this retrospective clinical data, support the need for further prospective investigation regarding the role of the 5' fusion partner in dictating therapeutic response.

Adding to the complexity, greater than 30 different 5' fusion partners have been reported for *ALK* fusions alone (**Fig. 3.6**). Furthermore, several oncogenic tyrosine kinase fusions have been reported in solid tumor and hematologic malignancies, including *ROS1*, *RET*, and *NTRK1*, amongst others⁵, and the 5' fusion partners can overlap between these different kinases (**Fig. 3.6**). For example, *TFG* has been a reported 5' partner for *ALK*, *ROS1*, and *NTRK* fusions^{8,59}. It is not feasible to extensively study each kinase fusion in the laboratory. To approach this complex problem, a high-throughput, computational approach to assess the effect of the 5' protein on the kinase and how it may alter drug binding could be greatly beneficial. Interestingly, a recent study in *RET*-rearranged NSCLC reported that patients harboring the *KIF5B-RET* fusion exhibited significantly less sensitivity to *RET* inhibition than patients harboring other *RET* fusions⁶⁰. In our study *KIF5B-ALK* was also one of the least sensitive *ALK* fusions to crizotinib and lorlatinib. Given that *KIF5B* reduced the response for both the *ALK* and *RET* kinase fusion, we may be able to make conclusions about the effects of a 5' partner that can be applied across multiple kinase fusions such as *ALK*, *RET*, *ROS1*, and *NTRK1/2/3*.

Current limitations in this study include the lack of clinical data to corroborate our findings. This is partly due to the lack of reporting of the specific fusion variant including the 5' partner in clinical studies. Additionally, studies that may report the specific variants, such as Drilon et al. 2018, may not be statistically powered to compare outcomes based on the variant. In order to fill this gap in knowledge, the design of future clinical trials will need to include prospective analysis of the specific fusion variants.

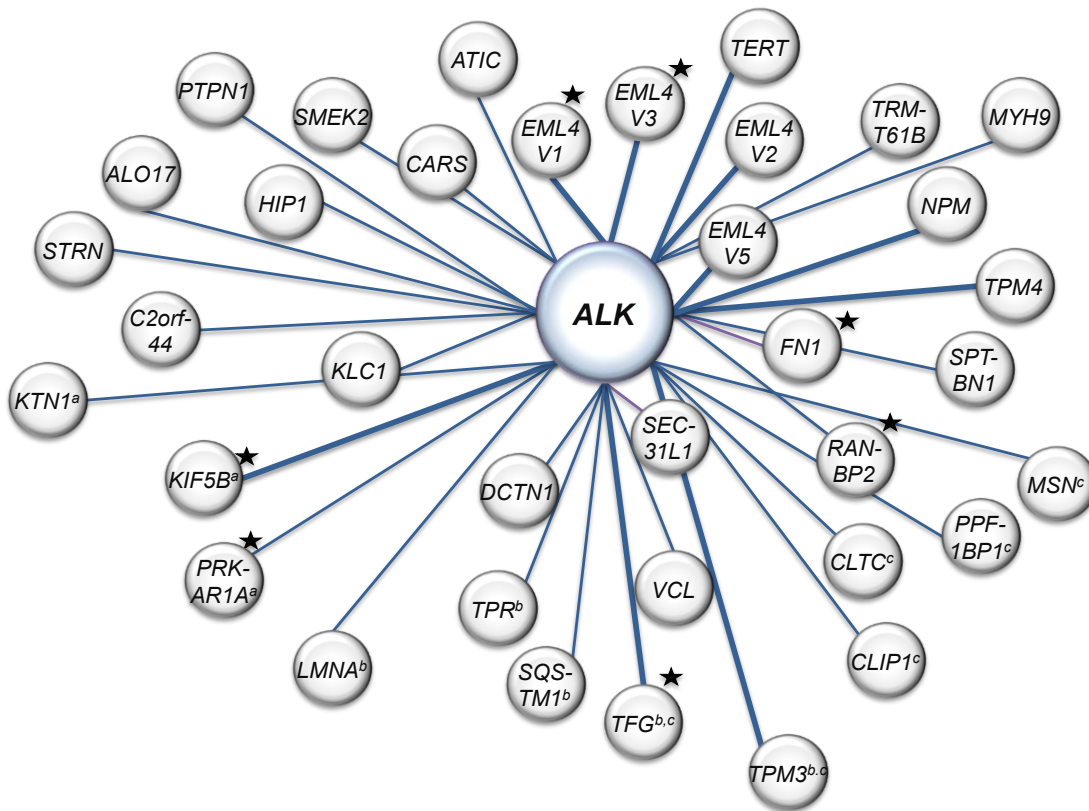


Figure 6.5. The heterogeneity of tyrosine kinase fusion partners. The ALK tyrosine kinase is represented in larger center circle. 5' fusion partners reported for ALK are depicted in gray circles. The weight of the connecting line between the kinase and 5' fusion partner reflects frequency of occurrence relative to other 5' partners reported. These fusions are not restricted to NSCLC or IMT. This figure is not a comprehensive representation of literature report¹⁻⁵. ALK fusions represented in our study are marked with a black star. 5' fusion partners which are also reported for RET, NTRK, and ROS1 fusions are marked with ^{a,b, or c} respectively.

In summary, our findings contribute to the preliminary studies in the emerging field of fusion biology, adding further evidence that the 5' partner of kinase fusion variants plays a significant role in the biological functions of the fusion that can affect cellular phenotypes and response to targeted therapies. Future clinical trials will need to prospectively consider ALK TKI efficacy in terms of the fusion partner present upon enrollment into the trial. Furthermore, given the large number of possible fusion variants that may be detected in the clinic, high-throughput computational and modeling studies will be necessary in order to support therapeutic decision-making.

3.6 References

1. Davare MA, Tognon CE: Detecting and targetting oncogenic fusion proteins in the genomic era. *Biol Cell* 107:111-29, 2015
2. Lin JJ, Shaw AT: Recent Advances in Targeting ROS1 in Lung Cancer. *J Thorac Oncol* 12:1611-1625, 2017
3. Drilon A, Laetsch TW, Kummar S, et al: Efficacy of Larotrectinib in TRK Fusion-Positive Cancers in Adults and Children. *N Engl J Med* 378:731-739, 2018
4. Pavlick D, Schrock AB, Malicki D, et al: Identification of NTRK fusions in pediatric mesenchymal tumors. *Pediatr Blood Cancer* 64, 2017
5. Shaw AT, Hsu PP, Awad MM, et al: Tyrosine kinase gene rearrangements in epithelial malignancies. *Nat Rev Cancer* 13:772-87, 2013
6. Hallberg B, Palmer RH: Mechanistic insight into ALK receptor tyrosine kinase in human cancer biology. *Nat Rev Cancer* 13:685-700, 2013
7. Kris MG, Johnson BE, Berry LD, et al: Using multiplexed assays of oncogenic drivers in lung cancers to select targeted drugs. *JAMA* 311:1998-2006, 2014
8. Lovly CM, Gupta A, Lipson D, et al: Inflammatory myofibroblastic tumors harbor multiple potentially actionable kinase fusions. *Cancer Discov* 4:889-95, 2014
9. Mosse YP, Lim MS, Voss SD, et al: Safety and activity of crizotinib for paediatric patients with refractory solid tumours or anaplastic large-cell lymphoma: a Children's Oncology Group phase 1 consortium study. *Lancet Oncol* 14:472-80, 2013
10. Drilon A, Siena S, Ou SI, et al: Safety and Antitumor Activity of the Multitargeted Pan-TRK, ROS1, and ALK Inhibitor Entrectinib: Combined Results from Two Phase I Trials (ALKA-372-001 and STARTRK-1). *Cancer Discov* 7:400-409, 2017
11. Ayeni D, Politi K, Goldberg SB: Emerging Agents and New Mutations in EGFR-Mutant Lung Cancer. *Clin Cancer Res* 21:3818-20, 2015
12. Melnick A, Licht JD: Deconstructing a disease: RARalpha, its fusion partners, and their roles in the pathogenesis of acute promyelocytic leukemia. *Blood* 93:3167-215, 1999
13. Yoshida T, Oya Y, Tanaka K, et al: Differential Crizotinib Response Duration Among ALK Fusion Variants in ALK-Positive Non-Small-Cell Lung Cancer. *J Clin Oncol* 34:3383-9, 2016
14. Woo CG, Seo S, Kim SW, et al: Differential protein stability and clinical responses of EML4-ALK fusion variants to various ALK inhibitors in advanced ALK-rearranged non-small cell lung cancer. *Ann Oncol* 28:791-797, 2017
15. Lin JJ, Zhu VW, Yoda S, et al: Impact of EML4-ALK Variant on Resistance Mechanisms and Clinical Outcomes in ALK-Positive Lung Cancer. *J Clin Oncol*:JCO2017762294, 2018

16. Red Brewer M, Yun CH, Lai D, et al: Mechanism for activation of mutated epidermal growth factor receptors in lung cancer. *Proc Natl Acad Sci U S A* 110:E3595-604, 2013
17. Wiesner T, Lee W, Obenauf AC, et al: Alternative transcription initiation leads to expression of a novel ALK isoform in cancer. *Nature* 526:453-7, 2015
18. Ren H, Tan ZP, Zhu X, et al: Identification of anaplastic lymphoma kinase as a potential therapeutic target in ovarian cancer. *Cancer Res* 72:3312-23, 2012
19. Takeuchi K, Choi YL, Togashi Y, et al: KIF5B-ALK, a novel fusion oncokinase identified by an immunohistochemistry-based diagnostic system for ALK-positive lung cancer. *Clin Cancer Res* 15:3143-9, 2009
20. Choi YL, Takeuchi K, Soda M, et al: Identification of novel isoforms of the EML4-ALK transforming gene in non-small cell lung cancer. *Cancer Res* 68:4971-6, 2008
21. Soda M, Choi YL, Enomoto M, et al: Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature* 448:561-6, 2007
22. Horvath A, Bertherat J, Groussin L, et al: Mutations and polymorphisms in the gene encoding regulatory subunit type 1-alpha of protein kinase A (PRKAR1A): an update. *Hum Mutat* 31:369-79, 2010
23. Pankov R, Yamada KM: Fibronectin at a glance. *J Cell Sci* 115:3861-3, 2002
24. Splinter D, Tanenbaum ME, Lindqvist A, et al: Bicaudal D2, dynein, and kinesin-1 associate with nuclear pore complexes and regulate centrosome and nuclear positioning during mitotic entry. *PLoS Biol* 8:e1000350, 2010
25. Lee SE, Kang SY, Takeuchi K, et al: Identification of RANBP2-ALK fusion in ALK positive diffuse large B-cell lymphoma. *Hematol Oncol* 32:221-4, 2014
26. Marino-Enriquez A, Wang WL, Roy A, et al: Epithelioid inflammatory myofibroblastic sarcoma: An aggressive intra-abdominal variant of inflammatory myofibroblastic tumor with nuclear membrane or perinuclear ALK. *Am J Surg Pathol* 35:135-44, 2011
27. Lin DH, Zimmermann S, Stuwe T, et al: Structural and functional analysis of the C-terminal domain of Nup358/RanBP2. *J Mol Biol* 425:1318-29, 2013
28. Rikova K, Guo A, Zeng Q, et al: Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell* 131:1190-203, 2007
29. Hernandez L, Bea S, Bellosillo B, et al: Diversity of genomic breakpoints in TFG-ALK translocations in anaplastic large cell lymphomas: identification of a new TFG-ALK(XL) chimeric gene with transforming activity. *Am J Pathol* 160:1487-94, 2002
30. McFadden DG, Dias-Santagata D, Sadow PM, et al: Identification of oncogenic mutations and gene fusions in the follicular variant of papillary thyroid carcinoma. *J Clin Endocrinol Metab* 99:E2457-62, 2014

31. Hernandez L, Pinyol M, Hernandez S, et al: TRK-fused gene (TFG) is a new partner of ALK in anaplastic large cell lymphoma producing two structurally different TFG-ALK translocations. *Blood* 94:3265-8, 1999
32. Beetz C, Johnson A, Schuh AL, et al: Inhibition of TFG function causes hereditary axon degeneration by impairing endoplasmic reticulum structure. *Proc Natl Acad Sci U S A* 110:5091-6, 2013
33. Johnson A, Bhattacharya N, Hanna M, et al: TFG clusters COPII-coated transport carriers and promotes early secretory pathway organization. *EMBO J* 34:811-27, 2015
34. Yagi T, Ito D, Suzuki N: Evidence of TRK-Fused Gene (TFG1) function in the ubiquitin-proteasome system. *Neurobiol Dis* 66:83-91, 2014
35. Prinz A, Diskar M, Erlbruch A, et al: Novel, isotype-specific sensors for protein kinase A subunit interaction based on bioluminescence resonance energy transfer (BRET). *Cell Signal* 18:1616-25, 2006
36. Taylor SS, Buechler JA, Yonemoto W: cAMP-dependent protein kinase: framework for a diverse family of regulatory enzymes. *Annu Rev Biochem* 59:971-1005, 1990
37. Richards MW, O'Regan L, Roth D, et al: Microtubule association of EML proteins and the EML4-ALK variant 3 oncoprotein require an N-terminal trimerization domain. *Biochem J* 467:529-36, 2015
38. Eichenmuller B, Everley P, Palange J, et al: The human EMAP-like protein-70 (ELP70) is a microtubule destabilizer that localizes to the mitotic apparatus. *J Biol Chem* 277:1301-9, 2002
39. Lee CC, Jia Y, Li N, et al: Crystal structure of the ALK (anaplastic lymphoma kinase) catalytic domain. *Biochem J* 430:425-37, 2010
40. Janoueix-Lerosey I, Lequin D, Brugieres L, et al: Somatic and germline activating mutations of the ALK kinase receptor in neuroblastoma. *Nature* 455:967-70, 2008
41. Armstrong F, Duplantier MM, Trempat P, et al: Differential effects of X-ALK fusion proteins on proliferation, transformation, and invasion properties of NIH3T3 cells. *Oncogene* 23:6071-82, 2004
42. Lovly CM, Heuckmann JM, de Stanchina E, et al: Insights into ALK-driven cancers revealed through development of novel ALK tyrosine kinase inhibitors. *Cancer Res* 71:4920-31, 2011
43. Zou HY, Friboulet L, Kodack DP, et al: PF-06463922, an ALK/ROS1 Inhibitor, Overcomes Resistance to First and Second Generation ALK Inhibitors in Preclinical Models. *Cancer Cell* 28:70-81, 2015
44. Kinoshita K, Asoh K, Furuichi N, et al: Design and synthesis of a highly selective, orally active and potent anaplastic lymphoma kinase inhibitor (CH5424802). *Bioorg Med Chem* 20:1271-80, 2012
45. Amano Y, Ishikawa R, Sakatani T, et al: Oncogenic TPM3-ALK activation requires dimerization through the coiled-coil structure of TPM3. *Biochem Biophys Res Commun* 457:457-60, 2015

46. Roskoski R, Jr.: Anaplastic lymphoma kinase (ALK): structure, oncogenic activation, and pharmacological inhibition. *Pharmacol Res* 68:68-94, 2013
47. Heuckmann JM, Balke-Want H, Malchers F, et al: Differential protein stability and ALK inhibitor sensitivity of EML4-ALK fusion variants. *Clin Cancer Res* 18:4682-90, 2012
48. Sang J, Acquaviva J, Friedland JC, et al: Targeted inhibition of the molecular chaperone Hsp90 overcomes ALK inhibitor resistance in non-small cell lung cancer. *Cancer Discov* 3:430-43, 2013
49. Gainor JF, Dardaei L, Yoda S, et al: Molecular Mechanisms of Resistance to First- and Second-Generation ALK Inhibitors in ALK-Rearranged Lung Cancer. *Cancer Discov* 6:1118-1133, 2016
50. Doebele RC, Pilling AB, Aisner DL, et al: Mechanisms of resistance to crizotinib in patients with ALK gene rearranged non-small cell lung cancer. *Clin Cancer Res* 18:1472-82, 2012
51. Penzel R, Schirmacher P, Warth A: A novel EML4-ALK variant: exon 6 of EML4 fused to exon 19 of ALK. *J Thorac Oncol* 7:1198-9, 2012
52. Mazot P, Cazes A, Dingli F, et al: Internalization and down-regulation of the ALK receptor in neuroblastoma cell lines upon monoclonal antibodies treatment. *PLoS One* 7:e33581, 2012
53. Chen ST, Lee JC: An inflammatory myofibroblastic tumor in liver with ALK and RANBP2 gene rearrangement: combination of distinct morphologic, immunohistochemical, and genetic features. *Hum Pathol* 39:1854-8, 2008
54. Baumann H, Kunapuli P, Tracy E, et al: The oncogenic fusion protein-tyrosine kinase ZNF198/fibroblast growth factor receptor-1 has signaling function comparable with interleukin-6 cytokine receptors. *J Biol Chem* 278:16198-208, 2003
55. Medves S, Noel LA, Montano-Almendras CP, et al: Multiple oligomerization domains of KANK1-PDGFRbeta are required for JAK2-independent hematopoietic cell proliferation and signaling via STAT5 and ERK. *Haematologica* 96:1406-14, 2011
56. Tognon CE, Mackereth CD, Somasiri AM, et al: Mutations in the SAM domain of the ETV6-NTRK3 chimeric tyrosine kinase block polymerization and transformation activity. *Mol Cell Biol* 24:4636-50, 2004
57. Zhao X, Ghaffari S, Lodish H, et al: Structure of the Bcr-Abl oncoprotein oligomerization domain. *Nat Struct Biol* 9:117-20, 2002
58. Richards MW, Law EW, Rennalls LP, et al: Crystal structure of EML1 reveals the basis for Hsp90 dependence of oncogenic EML4-ALK by disruption of an atypical beta-propeller domain. *Proc Natl Acad Sci U S A* 111:5195-200, 2014
59. Greco A, Mariani C, Miranda C, et al: The DNA rearrangement that generates the TRK-T3 oncogene involves a novel gene on chromosome 3 whose product has a potential coiled-coil domain. *Mol Cell Biol* 15:6118-27, 1995

60. Drilon A, Rekhtman N, Arcila M, et al: Cabozantinib in patients with advanced RET-rearranged non-small-cell lung cancer: an open-label, single-centre, phase 2, single-arm trial. *Lancet Oncol* 17:1653-1660, 2016

Chapter 4

DISCUSSION AND FUTURE DIRECTIONS

4.1 Overview

The final chapter is divided into three sections: the first is focused on Inflammatory Myofibroblastic Tumor (IMT); the second is focused on the study of oncogenic kinase fusions with an emphasis on Anaplastic Lymphoma Kinase (ALK) fusions, and the third is reserved for concluding remarks. Sections will discuss remaining questions in the field and future directions of the studies described in chapters 2 and 3.

4.2 Inflammatory Myofibroblastic Tumor

IMT Biology

There are several challenges in studying rare tumor types, and this often results in large gaps in understanding of the biology behind the disease. IMT is no exception. One obstacle that slows down research progress in IMT is the lack of an appropriate model system. To date, there are no published patient derived IMT cell lines or mouse models. This creates a challenge in understanding IMT biology as well as testing potential therapeutics pre-clinically. There are two big questions which pose as obstacles for developing a valid model: 1) What is the cell of origin for IMT? and 2) What is the role of the inflammatory infiltrate in this tumor? To begin to determine the cell of origin, we can use genomic and whole transcriptome data to give us insight into cell type and pathogenesis. Hypotheses can begin to be made regarding the cell type based on known tumor histology including positive or negative diagnostic markers and transcriptome data. Then known oncogenic drivers in IMT, such as ALK fusions, or other kinase fusions (in addition to tumor suppressor mutations such as mutant p53) can be used to initiate transformation in these cell types in mice. Similar approaches have been used to determine the cell of origin for other soft tissue sarcomas, such as rhabdomyosarcoma¹.

With an appropriate IMT mouse model, the role of the inflammatory infiltrate can then be studied. We may also approach learning about the inflammatory/tumor relationship by sorting tumor cells from the inflammatory cells in fresh tissue samples. Once sorted, RNA sequencing and/or proteomic profiling can be performed on the cohorts separately to give us a view of their distinct expression profiles. Additionally, experiments evaluating the effects of combining the cells in culture can be performed to learn more about their interactions and relationship. Since we don't currently know the role of the inflammatory cells in IMT, models systems which exclude this cell type such as cell lines and PDX, may produce questionable results for testing therapeutics. However, given the natural role of

myofibroblasts^{2,3}, it is possible that IMT is the result of wound healing gone astray. Since we know that the majority of these tumors are driven by oncogenic kinase fusions and that the tumors typically demonstrate significant regression upon treatment with a kinase inhibitor⁴, the inflammatory infiltrate may have little to do with driving tumor progression and more to do with promoting tumor initiation. In this case cell line models and PDX models may be well suited for initial pre-clinical drug testing.

Treatment Paradigms

Increasing data supports the conclusion that Inflammatory Myofibroblastic Tumor (IMT) is a kinase fusion driven disease with ~85% of cases harboring a rearrangement in kinases *ALK*, *ROS1*, *PDGFRB*, *RET*, or *NTRK3*⁵⁻⁷. Fortunately, all of these kinases are therapeutically targetable with tyrosine kinase inhibitors (TKIs) currently in clinical use. Patients with *ALK* positive IMTs have been successfully treated with the *ALK* TKI, crizotinib, in clinical trials^{4,8}. Response to crizotinib (which incidentally also targets *ROS1*) in *ROS1* positive IMTs has also been reported⁵. The positive response to TKIs in this disease highlights the importance of using next generation sequencing (NGS) diagnostic tools to identify effective therapies for IMT patients. Since there are no TKIs with FDA approval for use in IMT, patients must be granted compassionate use of the agent. The increased presence of 'basket trials' that enroll patients based on molecular profile vs. cancer type give hope for the approval of such agents in rare cohorts of patients such as these.

While the identification of kinase fusions in such a large percentage of IMTs significantly changes the repertoire of effective therapeutic options, we are still left with few rational therapies for the fusion negative cohort. Our studies have shown that IMTs, regardless of fusion status, harbor mutations affecting chromatin remodeling and the Vascular Endothelial Growth Factor (VEGF) pathway. Myofibroblasts are naturally involved in wound healing, and IMTs tend to be very large tumors^{2,3,9,10}. Both of these characteristics require increased angiogenesis, supporting the validity of these data. Fortunately, VEGF is also targetable with agents in current clinical use. For clinical utility, it would be helpful to identify histological biomarkers that indicate high VEGF activity to allow for therapy selection or patient stratification. Such a biomarker could be used to construct a clinical trial for the use of VEGF TKIs in IMT. Interestingly, the histone deacetylase inhibitor, valproic acid, has also shown anti-angiogenic properties in addition to its epigenetic regulation¹¹, making it an attractive drug candidate.

Within kinase fusion positive IMTs, additional therapeutic questions remain. As more and more patients are treated with TKIs, it is likely that we will need to address mechanisms of acquired

resistance to therapy. ALK TKI resistant IMTs have already been reported^{12,13}. Additionally, there is the question of efficacy of a TKI against different fusion variants defined by the 5' fusion partner present. Evidence in non-small cell lung cancer (NSCLC) suggests TKIs are differentially efficacious against fusions variants that differ by the 5' partner¹⁴⁻¹⁷. Kinase fusions in IMT are even more diverse than NSCLC⁵. In our own studies alone, we have identified 17 different *ALK* fusion variants that differ by the specific 5' protein (such as *TFG-ALK* vs. *RANBP2-ALK*) or by various lengths of the same 5' protein, such as the 4 *FN1-ALK* variants described in chapter 2. Addressing these gaps in knowledge can be a challenge as the study subjects of interest, fusion negative and therapy resistant, are small cohorts within an already rare disease. Success will require locking arms in collaboration with researchers and physicians in multiple institutions in order to obtain enough samples or prospectively monitor enough patients to draw confident conclusions.

4.3 ALK Fusions

Discussion and Unpublished Results

All ALK fusion proteins described to date contain the entire ALK tyrosine kinase domain, and since ALK tyrosine kinase inhibitors (TKIs) are ATP mimetics, one might anticipate the drug efficacy to be similar from one fusion to the next. However, retrospective analyses of clinical data suggest that progression free survival differs based on the ALK fusion variant present, where the variant is defined by the presence of a specific 5' partner gene. While our studies have corroborated this notion, a thorough mechanistic explanation of differential sensitivity is still lacking. There are many structural and biochemical properties of the fusion proteins which have the potential to contribute to differences in response to ALK TKIs, including: 1) the type of oligomerization domains present; 2) the stoichiometry of oligomerization: dimers,¹⁸ trimers,^{19,20} and multimers^{21,22}; 3) the intrinsic kinase activity of the fusion protein²³; 4) protein-protein interactions, which may differ based on the domain structure of the fusion partner; 5) length of the 5' partner; 6) protein folding, tertiary structure, and degree of disorder; 7) subcellular localization; and 8) protein stability.

Data from Armstrong et al. 2004²⁴ and Armstrong et al. 2007²⁵ suggest that the ALK fusion may interact with the wild-type 5' partner protein. This is certainly another factor to consider. This interaction may alter the normal processes in which the wild-type 5' partner is involved which can have numerous effects on the cell including those that may affect drug efficacy. For example, it is possible that this interaction could result in a dominant negative effect, binding up a large portion of the WT 5' partner and disabling its function. For a protein such as PRKAR1A, which was included in

our study, this may result in a disruption in PKA signaling since PRKAR1A is a part of the PKA holoenzyme and regulates its activity. Interestingly, total PKA and its activating phospho-protein sites (AMPK α pT172, AMPK- α 2 pS345), which are on the RPPA panel we performed, were not significantly changed in the PRKAR1A-ALK cell line. However, the RPPA data did reveal that the mTOR pathway was significantly down-regulated in this cell line exhibiting a fold change = |57| (**Table S3.4**), and PRKA, located upstream of mTOR, is known to have an inhibitory effect on mTOR activity²⁶. Since the RPPA pathway analysis is restricted toward particular pathways, I decided to interrogate the list of proteins showing significantly different expression between the PRKAR1A-ALK and WT ALK NIH 3T3 cell line models (**Table S3.2**) using WebGestalt²⁷ over representation enrichment analysis to identify other pathway associations. Indeed, the AMPK (PRKA) signaling pathway is in the top three enriched categories (**Table 4.1**). The significant disruption of the mTOR and AMPK pathways could explain why the PRKAR1A-ALK cell line was a consistent outlier in our study. The endogenous role of 5' partner proteins in regulation of other critical pathways will be important to consider moving forward.

Category Name	Observed # of Genes	Expected # of Genes	FDR
HIF-1 signaling pathway	8	0.43	1.72E-06
mTOR signaling pathway	7	0.64	3.50E-04
AMPK signaling pathway	6	0.52	9.53E-04
EGFR tyrosine kinase inhibitor resistance	5	0.34	1.34E-03
Longevity regulating pathway	5	0.39	2.22E-03
Endocrine resistance	5	0.41	2.24E-03
AGE-RAGE signaling pathway in diabetic complications	5	0.42	2.24E-03
Glioma	4	0.27	3.76E-03
Proteoglycans in cancer	6	0.84	3.76E-03

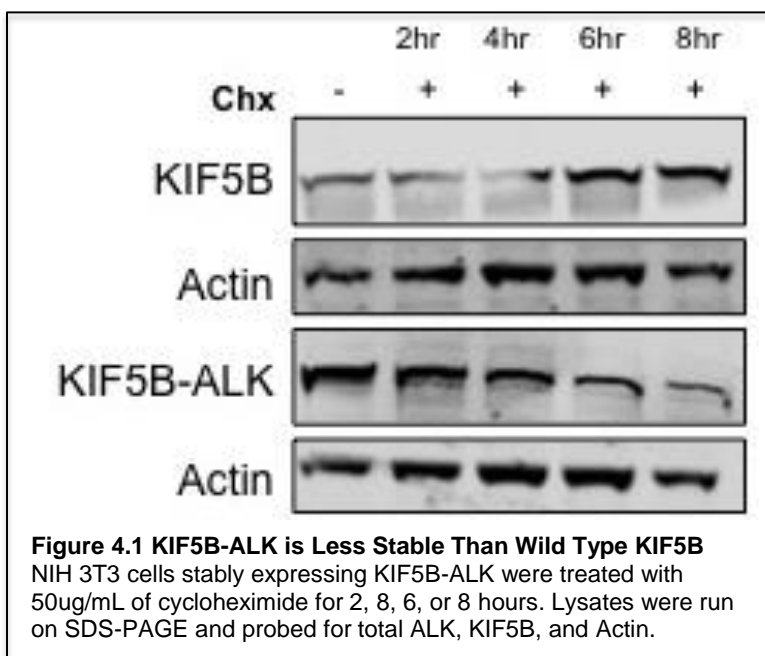
Table 4.1: Enriched Categories from the Differentially Expressed Proteins between the PRKAR1A-ALK NIH3T3 Cell Line and the WT ALK NIH3T3 Cell Line.

The table exhibits the enriched categories (determined by KEGG analysis) represented in the differentially expressed proteins between PRKAR1A-ALK expressing NIH3T3 cell lines and WT ALK expressing NIH3T3 cell lines (from table S3.2). The observed number of genes in this differential gene set that occurred in the category is listed under "Observed # of Genes", and the expected number of genes observed for any given gene set that is interrogated is listed under "Expected # of Genes". The FDR is the Benjamini-Hochberg FDR.

Armstrong et al. 2004 also demonstrated that the fusions were differentially able to activate PI3K and JAK-STAT signaling. Furthermore, the ability of the different ALK fusions to activate PI3K kinase

activity correlated with the fusion's transendothelial migration properties. However, migratory data studied in this context must be applied with care. This cellular phenotype and its relation to tumor metastasis are not directly applicable due to the high dependency of metastasis on cell/tumor type. For example, these same fusions also occur in IMT, which rarely has distant metastasis^{5,28,29}. Further, those that do experience distance metastasis are usually ALK negative^{28,29}. From our published data²⁹ it does not appear that signaling differences between the fusions in our study correlate with transformative potential (foci or colony formation), and we did not assess migration. Cellular phenotypes such as metastasis and migration would more appropriately be studied in the endogenous expression setting, ideally under the native promoter and cancer cell/tissue type.

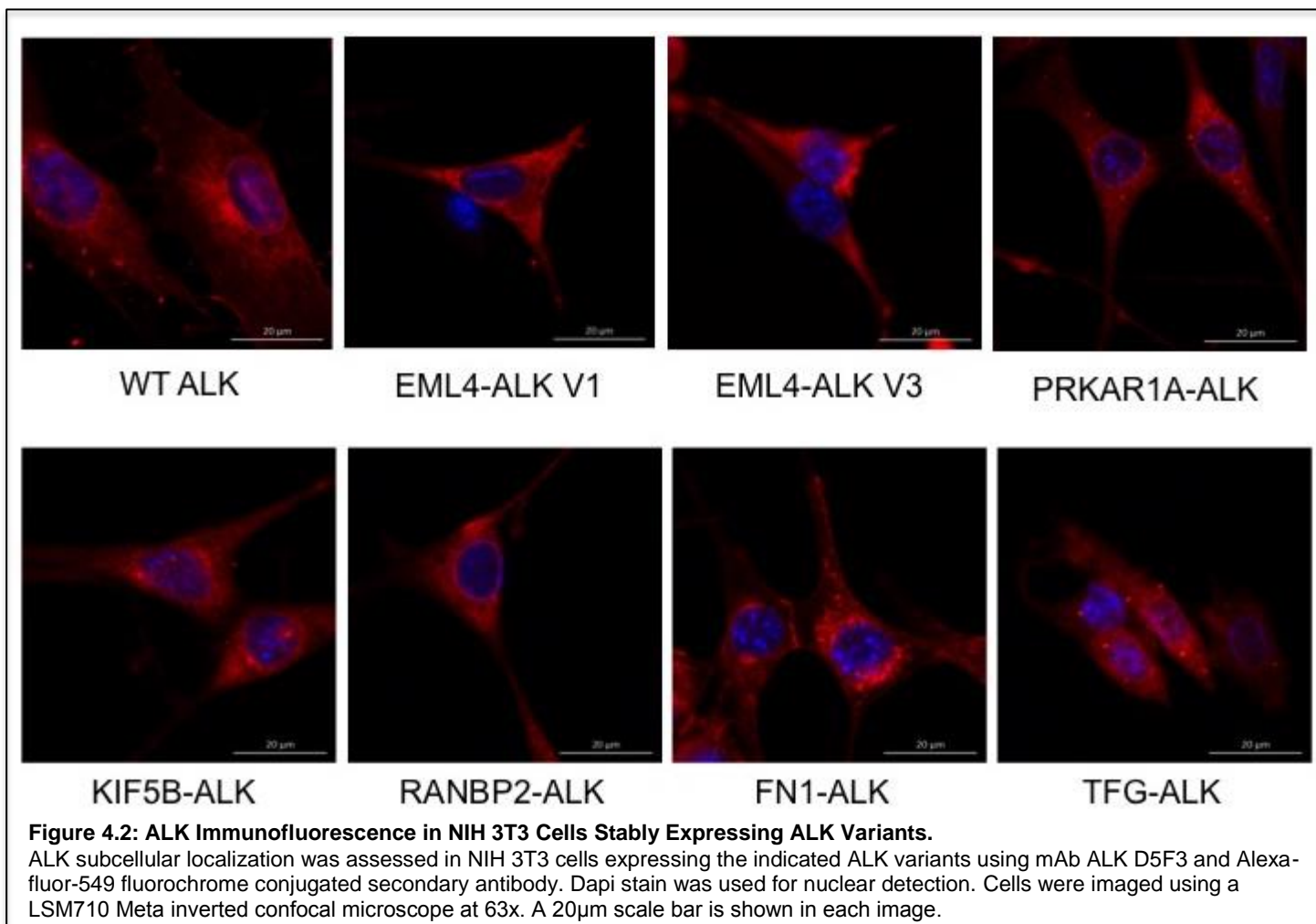
In our published work³⁰, we identified a negative correlation between fusion kinase activity and fusion protein turnover. To determine if the rapid turnover rate of KIF5B-ALK is a phenotype intrinsic to the fusion protein or conferred by the 5' partner, I performed a time course study to assess the stability of the fusion protein as well as the wild type 5' protein by treating cells with cycloheximide to inhibit new protein synthesis. In contrast to the KIF5B-ALK fusion protein, which exhibits a half-



life of approximately 2 hours, protein levels of wild type KIF5B did not decrease over the cycloheximide time course (**Figure 4.1**). Based on our data regarding protein turnover and kinase activity, in addition to the findings in previous studies of *EML4-ALK* variant 1, which is a fusion between exon 13 of *EML4* and exon 20 of *ALK*, and *EML4-ALK* variant 3, which is a fusion between exon 6 of *EML4* and exon 20 of *ALK*^{20,32}, we suspect that the explanation of differential sensitivity is related to the intrinsic structural and biochemical properties of the fusion protein itself, which contribute to kinetic activity and protein stability. These in turn may then affect the ability of a fusion to activate particular signaling pathways more than others, such as the RAS/MAPK pathway activation, which we've shown to be more highly active in cell lines harboring fusions that have greater kinetic activity and protein stability. One additional, plausible explanation for why differences in protein expression and phospho-protein profiles are seen is that these fusions may each interact with different proteins and pathways. These fusions localize differently based on the 5' partner, even for

those that share the same 5' partner but vary in the amount of that protein present in the fusion, such as the EML4-ALK^{20,32} fusion variants. This could allow for different protein interactions. Differential pathway activation based on ALK localization has also been shown in previous studies^{19,24,31}.

We have assessed the subcellular localization of the ALK fusions studied in chapter 3 by ALK immunofluorescence (IF). We noted that some fusions appeared to aggregate or form puncta, such as FN1-ALK, KIF5B-ALK, TFG-ALK, and PRKAR1A-ALK, while others were more diffusely cytoplasmic (**Figure 4.2**). Based on the pattern of fluorescence and studies reporting on wild type ALK and ALK F1174L localization³³ or localization of the wild type 5' proteins, I hypothesize that FN1-ALK is aggregating in the endoplasmic reticulum and compartments of the golgi apparatus along with the plasma membrane; and KIF5B-ALK is localizing to the centrosome. Further co-localization studies are needed to test these hypotheses and to identify any correlation to downstream pathway activation. Overall, there are many cellular and structural factors that have the potential to play a role in the response to ALK TKIs. This results in an explanation that is likely multifactorial and potentially different for each fusion.

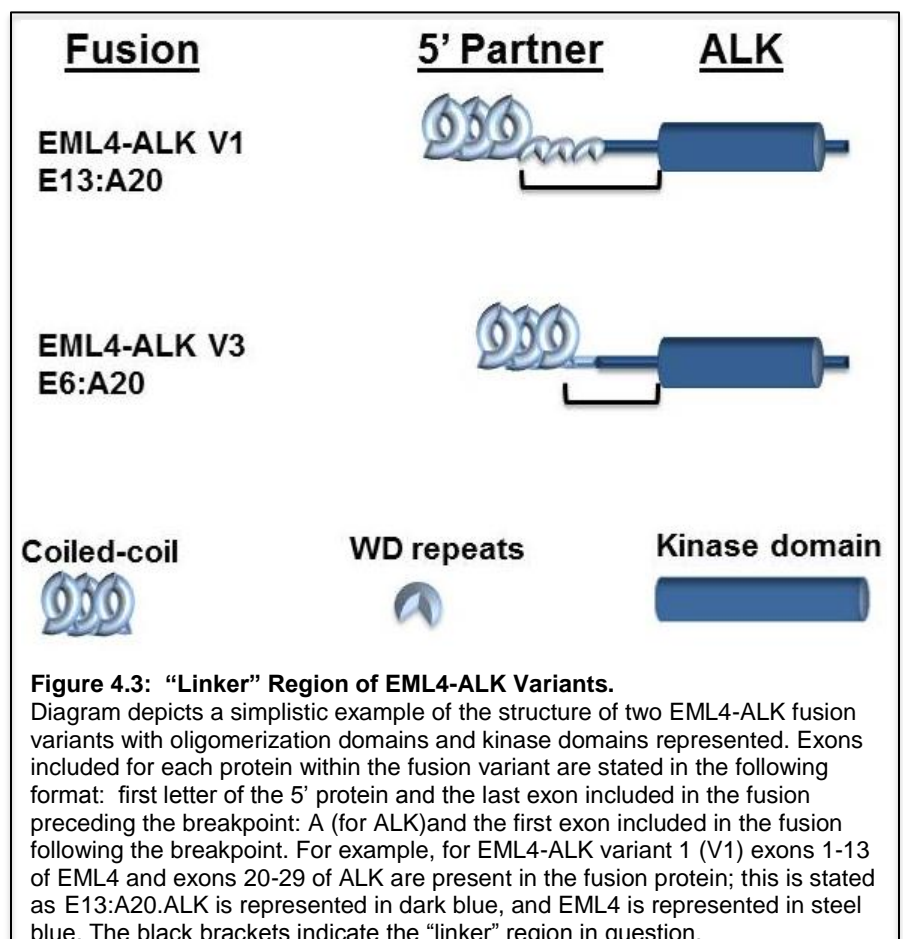


Ongoing Studies and Future Directions

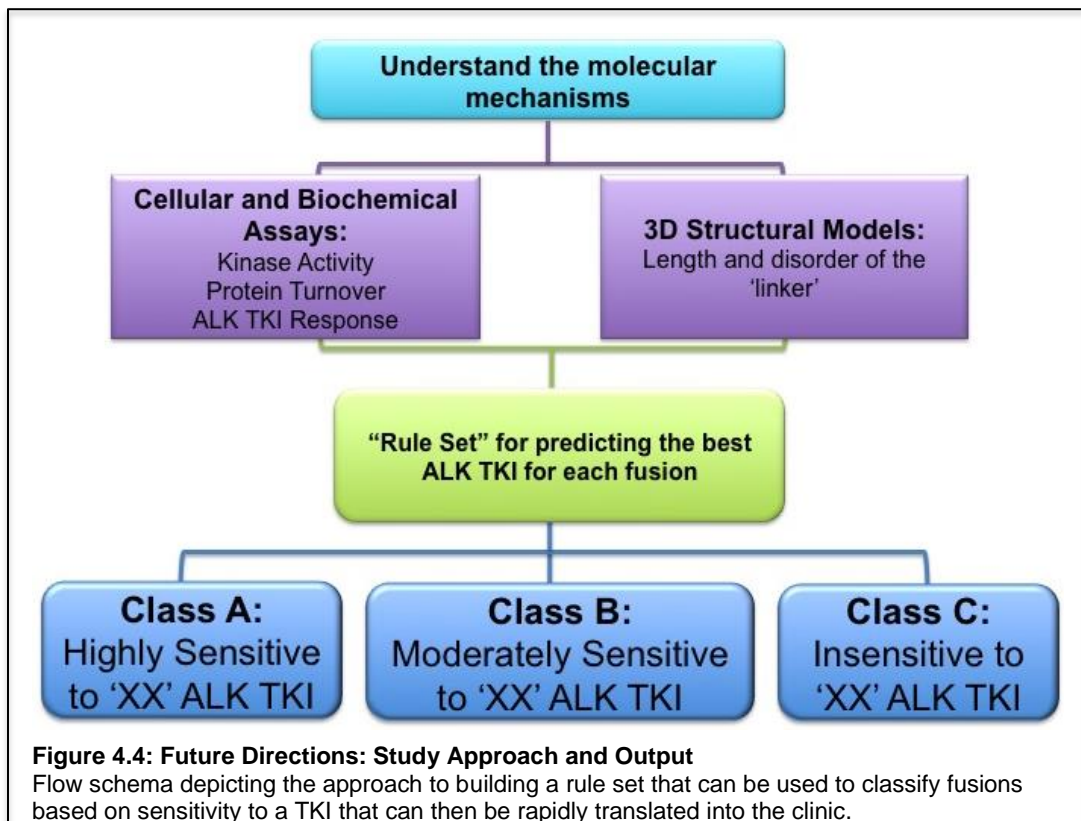
Our overall goal is to build a rule set that predicts the best TKI for each fusion. In order to do this we must determine the mechanism behind TKI efficacy and identify key factors that predict response to variable ALK TKIs. Ideally these factors could be modeled or predicted computationally to allow for high throughput evaluation. Previous studies have shown that the fusion partner contributes a dimerization domain that is sufficient and necessary for auto-activation of the kinase and transformation ability^{32,34}. Many fusion partners contain coiled-coil domains or other known dimerization domains; however, not all fusion partners have an obvious dimerization motif. Structural differences between the various 5' partners, including the oligomerization domains, additional structured domains present, and the length and/or degree of disorder of the region between the oligomerization domain of the 5' partner and the kinase domain of ALK could alter the dimerization ability -the extent in which oligomerization occurs (dimers, trimers, multimers) or the strength of the oligomerization- and therefore activation dynamics of the fusion and/or accessibility of the catalytic domain in ALK.

In our on-going work, we are collaborating with investigators within the Vanderbilt Center for Structural Biology to gain more detailed mechanistic insights into the differences in drug sensitivity observed amongst the various ALK fusion proteins, by attempting to generate computational models that can predict possible oligomerization states of each fusion, disordered regions between structured domains, as well as distances in 3D space and flexibility of the "linker" region of the fusion protein: the region

between the oligomerization domain and the kinase domain (**Figure 4.3**). Using these predictions along with wet lab experiments, we hope to generate testable hypotheses regarding how these



characteristics may be affecting kinase activity, activation potential of the fusion, and drug sensitivity. For example, we may find that fusions with a short and inflexible distance between the oligomerization domain and kinase domain all perform equally in terms of kinase activity and/or drug sensitivity. We can then make modifications to these fusions to make them longer and more flexible, or we can make longer fusions shorter and more inflexible to test this prediction. In addition, we expect that geometrical constraints imposed by a “linker” may prevent two kinase domains from



assembling into the symmetric dimer conformation necessary for activation. For example, a “linker” that is too long or too rigid may fail to bring kinase domains together into a productive assembly. We expect that this property may also be reflective of the sensitivity to ALK TKI’s. For example, the longer and more flexible this region is, the less sensitive they may be to inhibition by an ALK TKI due to steric interference by the 5’ partner and the protein associations that may come along with it. This multipronged approach will be used to build a rule set to classify fusions based on response to ALK TKIs (**Figure 4.4**).

In collaboration with the Jens Meiler laboratory, we began by constructing the predictive structures and estimating the effective distances of the linker regions (defined as the region between the putative oligomerization domain in the 5’ partner protein and the ALK kinase domain) of the 7 ALK fusions discussed throughout chapter 3 (**Table 4.2**). However, we believe that the best way to begin to test this hypothesis, is to first ask the question in a more simple system, controlling for the many cellular and biochemical properties that different 5’ proteins introduce into the experiment.


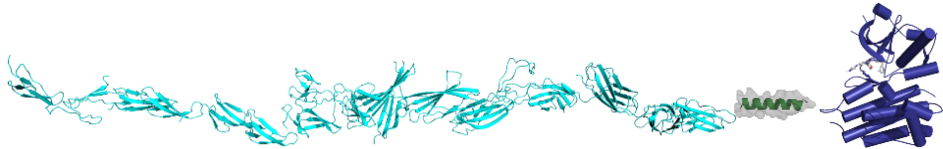
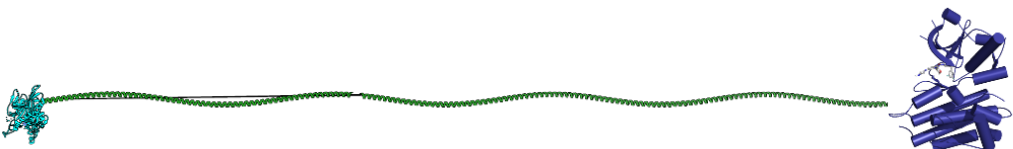
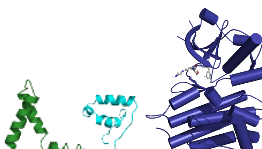
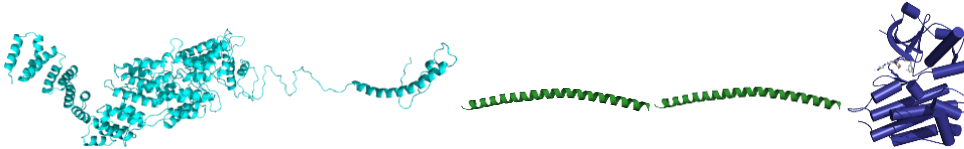
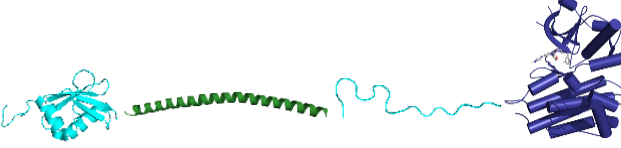
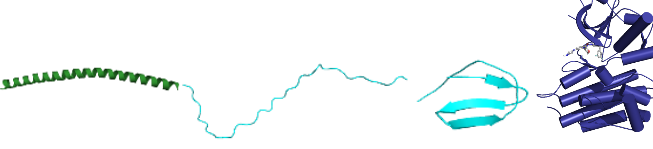
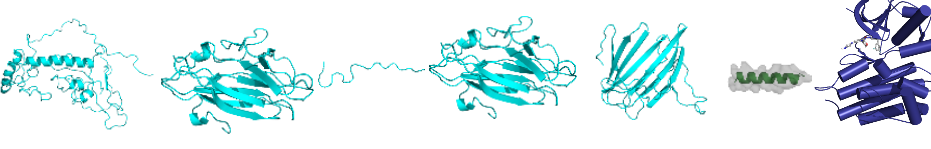
ALK Variant	Structural model	Effective Distance
EML4-ALK V1		830 Å
FN1-ALK		210 Å
KIF5B-ALK		210 Å
PRKAR1A-ALK		250 Å
RANBP2-ALK		190 Å
TFG-ALK		580 Å
EML-ALK V3		800 Å
ALK F1174L		210 Å

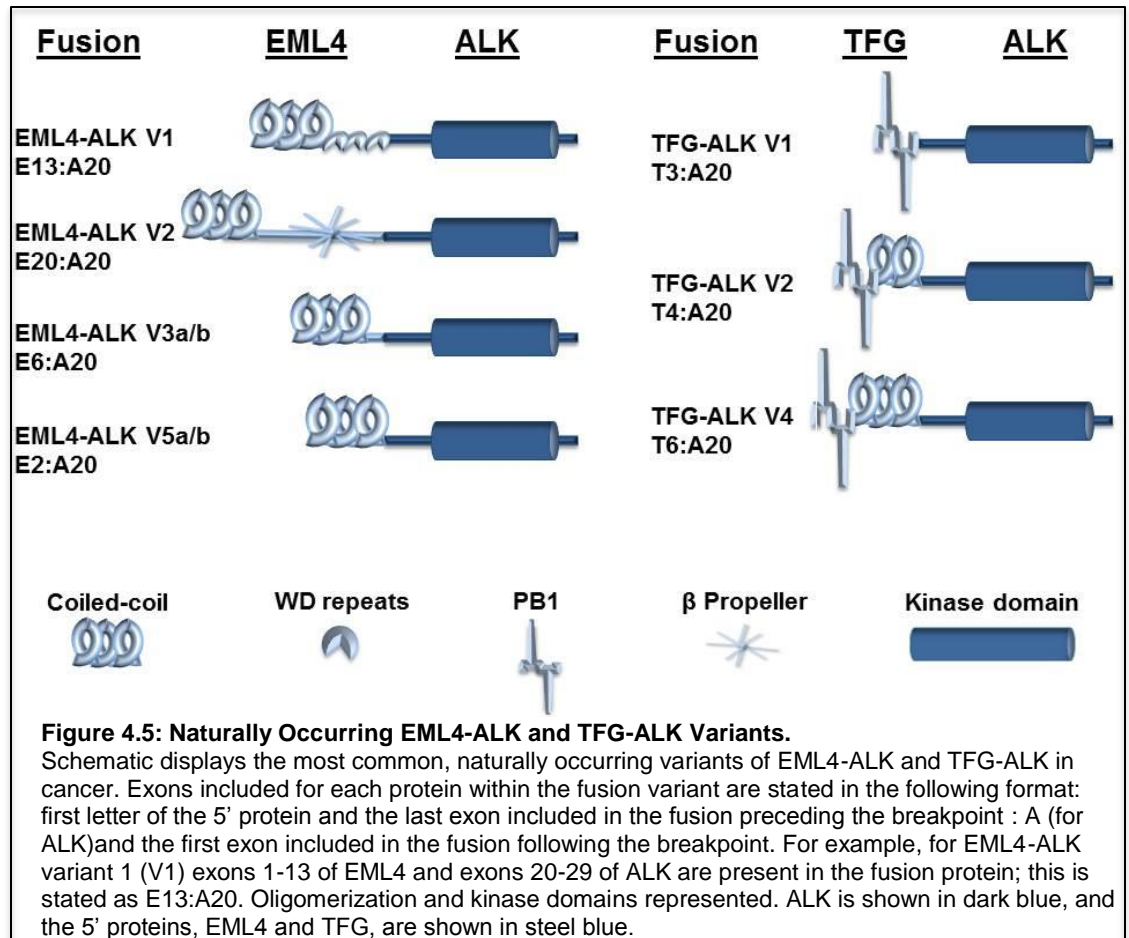
Table 4.2: Predictive Structural Models of ALK Fusion Variants and ‘Linker’ Region Lengths

Computational predictions of the structure of the ALK fusion variants were made using crystal structure data in the Protein Data Bank, RaptorX server, and PyMOL. The effective distance represents the maximum distance between the nearest putative oligomerization domain and the ALK kinase domain in 3 dimensional space. It was calculated by summing the distance between N- and C-termini of all intervening structured domains, and adding the number of unstructured residues multiplied by 3.4 Angstroms, to represent the length of a fully extended polypeptide chain.

Therefore, further studies will first focus only on naturally occurring variants in which the 5' partner remains the same, differing only by the amount of the protein present, for example, EML4-ALK variants 1, 2, 3a/b, and 5a/b or TFG-ALK variants 1, 2, and 4 (Figure 4.5). One interesting study

looking at the structural implications of fusion proteins describes different mechanisms by which disorder affects oncogenic function³⁵. The authors describe EML4-ALK in particular as being a 'special case' when compared to other ALK fusion proteins such as TFG-ALK and NPM-ALK due to

the importance of structural domains beyond the coiled-coil oligomerization domain for oncogenic function of the fusion. The various breakpoints within *EML4* often disrupt these ordered domains. The disruption of ordered domains, that are not highly repetitive, is an odd and unexpected characteristic of viable fusion proteins³⁴, and studies have shown that this disruption in EML4-ALK causes problems with protein folding and the biology of the fusion such as proper localization^{20,33}. This may explain why only the EML4-ALK variants –of all the fusions we tested in our studies- were sensitive to inhibition of the chaperone protein, Hsp90. With this in mind further studies should include both EML4-ALK and TFG-ALK variants. Since both EML4 and TFG have been reported as ALK fusion partners in NSCLC and IMT⁵, it would be interesting to evaluate these fusions in multiple cellular settings, such as AALE and NIH 3T3 cells, to see if the results are cell type dependent. The AALE cell line, immortalized normal human lung epithelium, would represent the cellular background for NSCLC while the NIH3T3's, immortalized mouse fibroblast, would represent the cellular background for IMT.



I hypothesize that there will be categories of different structural mechanisms that affect oncogenic function and different classes of fusions within those mechanistic categories. For example, studies of the EML4-ALK fusion variants may reveal one structural mechanism, which may be based on fusions that survive breakpoint interruptions of ordered domains, and within that mechanism, multiple classes of fusions are delineated that correspond with ALK TKI response. Studies of the TFG-ALK variants may reveal another structural mechanism, and within that mechanism, multiple classes of fusions corresponding with ALK TKI response. These studies may provide the first two “structural, mechanistic categories” into which additional fusions may then be assigned.

Extending these studies to the FN1-ALK variants detected in our IMT cases described in chapter 2 would be an interesting addition since these fusions retain the ALK transmembrane domain and vary in the length of FN1 as well as ALK. These would likely add another “structural, mechanistic category”. Additionally, two cases of EML4-ALK retaining the ALK transmembrane domain have been reported^{36,37}. It would be interesting to compare it determine if it behaves more like the FN1-ALK fusions or more like the other EML4-ALK variants.

Similarly, further studies may include the addition of other kinase fusions, particularly those that have common 5' partner proteins with ALK fusions, such as the TFG-ROS1 variants detected in our IMT cases in chapter 2. These TFG-ROS1 fusions retain the ROS1 transmembrane domain. Thus, it would be interesting to compare its function to the TFG-ALK variants, the naturally occurring ALK fusions that retain the ALK transmembrane domain, or a TFG-ALK fusion designed to retain the ALK transmembrane domain.

4.4 Concluding Remarks

While clinical results with targeted therapies are promising for kinase fusion positive cancers, our studies indicate that it is not sufficient to only consider the kinase when determining the best therapy. The non-kinase partner protein affects drug efficacy, and more detailed clinical data and pre-clinical investigations are warranted to increase therapeutic precision. To move the needle forward, we must address current limitations including the lack of prospective clinical data detailing the specific fusion and tissue collection. Prospective clinical data that specifies the 5' partner protein and kinase in the fusion is necessary to corroborate pre-clinical, mechanistic studies. Additionally, the design of future clinical trials will need to include prospective analysis of the specific fusion variants and will need to

prospectively consider ALK TKI efficacy in terms of the fusion partner present upon enrollment into the trial. Furthermore, to progress pre-clinical translational studies, collecting tissue to generate patient derived cell line models that harbor different fusion variants are also needed to increase accuracy of pre-clinical results. In addition, given the large number of possible fusion variants that may be detected in the clinic, it is vital that we determine the mechanisms by which the 5' proteins affect therapeutic response in a way that can be evaluated with high-throughput computational and modeling studies, which can then be rapidly translated for therapeutic decision-making.

This is an exciting time for translational cancer research. As the vision of precision medicine continues to unfold and NGS technologies continue to come to the forefront of clinical diagnostics, there will be an abundance of personalized clinical and genomic data. It is critical that the translational science keep up with this pace in order to feed rational information back into the clinical system and progress medical treatment in a timely manner. Therefore, it is imperative that basic scientists and physicians work together to keep translational science ahead of the curve, collecting tissue and clinical data prospectively, and designing rational translational studies that can rapidly re-inform clinical practice.

4.5 Methods

RPPA KEGG Analysis Proteins exhibiting significant differential expression ($FC \geq |2|$, Adj. p-value ≤ 0.05) between PRKAR1A-ALK expressing NIH3T3 cells and WT ALK expressing NIH3T3 cells were entered as the “Gene List” interrogated using the Web-based GENE SeT AnaLysis Toolkit (Web Gestalt). Parameters included: organism of interest = hsapiens; method of interest = overrepresentation enrichment analysis (ORA); functional database class = pathway; functional database name = KEGG. The reference set for enrichment analysis was ‘genome_protein-coding’.
http://www.webgestalt.org/results/Project_wg_result1536084351/Report_wg_result1536084351.html

Immunoblot and Antibodies The following primary antibodies were obtained from Cell Signaling Technology (Danvers, MA, USA): ALK mAb Mouse (3633S), ALK mAb Mouse (3791S) and ALK D5F3 mAb Rabbit (#3633). The KIF5B primary antibody (ab167429, EPR10276(B)) was obtained from ABCAM (Cambridge, MA, USA). The actin antibody (A2066) was purchased from Sigma-Aldrich (St. Louis, MO, USA). The following secondary antibodies were obtained from LiCor (Lincoln, NE, USA): IRDye 680RD Goat anti-Mouse (#92668070), and IRDye 800 Goat anti-Rabbit (#92632211). Alexa-fluor549 (A-11032) fluorochrome-conjugated secondary antibody was obtained from Invitrogen (Grand Island, NY, USA). For infrared (IR) immunoblot, cells were harvested, washed in PBS, and

lysed in KRAS buffer [50 mM Tris-HCl, 150 mM NaCl, 5 mM MgCl₂, 1% triton X-100, 0.5% Na Deoxy Cholate, 0.1% SDS, 40 nmol/L sodium fluoride, 1mM sodium orthovanadate, and complete protease inhibitors (Roche Diagnostics, Indianapolis, IN, USA)]. Signal detection was obtained using the LiCor Odyssey scanner system and quantified using iMageStudio Lite software (LiCor).

Protein Stability Assay NIH 3T3 cells stably expressing KIF5B- ALK were seeded in 6 well plates. 24 hours after seeding, cells were treated with 50µg/mL cycloheximide (Sigma-Aldrich) for 2, 4, 6, or 8 hours, harvested for lysate, and probed for total ALK protein (Cell Signaling, #3191S), KIF5B (ABCAM, ab167429, EPR10276(B)), and actin (A2066) by IR Western blot.

Immunofluorescence NIH 3T3 cell lines stably expressing ALK variants were seeded on 8 chamber slides (20,000 cells per chamber) and cultured until nearly confluent. Cells were fixed with ice cold 100% methanol for 20 minutes at 4°C and washed three times in 0.5% PBST. Cells were incubated in blocking buffer [1x PBST, 1% BSA, 0.2% Triton™ X-100] for 30 minutes at room temperature and washed 3 times with PBS. Cells incubated in 200µL of diluted ALK D5F3 primary antibody (Dilution: 1:200) overnight at 4°C. Cells were washed 3 times with PBS and incubated in an Alexa-fluor549 fluorochrome-conjugated secondary antibody (1µg/mL) dilution in the dark for 2 hours at room temperature. Cells were rinsed 3 times with PBS. Chambers were removed, Prolong® Gold Antifade Reagent with DAPI and coverslips were applied and sealed. Slides were stored horizontally at 4°C with minimal light exposure. Slides were imaged using a LSM710 Meta (Zeiss, Pleasantville, NY, USA) inverted confocal microscope.

Molecular Modeling Models of the ALK fusion proteins were generated using crystal structures from the Protein Data Bank for proteins and domains of known structure (PDB codes 5fto, 4cgb, 5ojf, 1ipg, 3im4, 3r8q, 5l73). The RaptorX server³⁸ was used to build models for loops and un-templated domains (ALK N-terminus and Gly-rich domains, RanBP2), resulting in models that are unlikely to be accurate in atomic detail, but approximately correct in size. Models of each domain were combined in an extended arrangement using PyMOL³⁹, and unstructured inter-domain regions were represented as flexible loops. The effective distance metric represents the maximum distance between the nearest putative oligomerization domain (usually a coiled-coil) and the ALK kinase domain. It was calculated by summing the distance between N- and C- termini of all intervening structured domains, and adding the number of unstructured residues multiplied by 3.4 Angstroms, to represent the length of a fully extended polypeptide chain.

4.6 References

1. Rubin BP, Nishijo K, Chen HI, et al: Evidence for an unanticipated relationship between undifferentiated pleomorphic sarcoma and embryonal rhabdomyosarcoma. *Cancer Cell* 19:177-91, 2011
2. Hinz B, Phan SH, Thannickal VJ, et al: The myofibroblast: one function, multiple origins. *Am J Pathol* 170:1807-16, 2007
3. Hinz B: Formation and function of the myofibroblast during tissue repair. *J Invest Dermatol* 127:526-37, 2007
4. Mosse YP, Voss SD, Lim MS, et al: Targeting ALK With Crizotinib in Pediatric Anaplastic Large Cell Lymphoma and Inflammatory Myofibroblastic Tumor: A Children's Oncology Group Study. *J Clin Oncol* 35:3215-3221, 2017
5. Lovly CM, Gupta A, Lipson D, et al: Inflammatory myofibroblastic tumors harbor multiple potentially actionable kinase fusions. *Cancer Discov* 4:889-95, 2014
6. Pavlick D, Schrock AB, Malicki D, et al: Identification of NTRK fusions in pediatric mesenchymal tumors. *Pediatr Blood Cancer* 64, 2017
7. Antonescu CR, Suurmeijer AJ, Zhang L, et al: Molecular characterization of inflammatory myofibroblastic tumors with frequent ALK and ROS1 gene fusions and rare novel RET rearrangement. *Am J Surg Pathol* 39:957-67, 2015
8. Schoffski P, Sufliarsky J, Gelderblom H, et al: Crizotinib in patients with advanced, inoperable inflammatory myofibroblastic tumours with and without anaplastic lymphoma kinase gene alterations (European Organisation for Research and Treatment of Cancer 90101 CREATE): a multicentre, single-drug, prospective, non-randomised phase 2 trial. *Lancet Respir Med* 6:431-441, 2018
9. Coffin CM, Watterson J, Priest JR, et al: Extrapulmonary inflammatory myofibroblastic tumor (inflammatory pseudotumor). A clinicopathologic and immunohistochemical study of 84 cases. *Am J Surg Pathol* 19:859-72, 1995
10. Meis JM, Enzinger FM: Inflammatory fibrosarcoma of the mesentery and retroperitoneum. A tumor closely simulating inflammatory pseudotumor. *Am J Surg Pathol* 15:1146-56, 1991
11. Zhao Y, You W, Zheng J, et al: Valproic acid inhibits the angiogenic potential of cervical cancer cells via HIF-1 α /VEGF signals. *Clin Transl Oncol* 18:1123-1130, 2016
12. Kimbara S, Takeda K, Fukushima H, et al: A case report of epithelioid inflammatory myofibroblastic sarcoma with RANBP2-ALK fusion gene treated with the ALK inhibitor, crizotinib. *Jpn J Clin Oncol* 44:868-71, 2014
13. Sasaki T, Okuda K, Zheng W, et al: The neuroblastoma-associated F1174L ALK mutation causes resistance to an ALK kinase inhibitor in ALK-translocated cancers. *Cancer Res* 70:10038-43, 2010

14. Yoshida T, Oya Y, Tanaka K, et al: Differential Crizotinib Response Duration Among ALK Fusion Variants in ALK-Positive Non-Small-Cell Lung Cancer. *J Clin Oncol* 34:3383-9, 2016
15. Woo CG, Seo S, Kim SW, et al: Differential protein stability and clinical responses of EML4-ALK fusion variants to various ALK inhibitors in advanced ALK-rearranged non-small cell lung cancer. *Ann Oncol* 28:791-797, 2017
16. Seo S, Woo CG, Lee DH, et al: The clinical impact of an EML4-ALK variant on survival following crizotinib treatment in patients with advanced ALK-rearranged non-small-cell lung cancer. *Ann Oncol* 28:1667-1668, 2017
17. Lin JJ, Zhu VW, Yoda S, et al: Impact of EML4-ALK Variant on Resistance Mechanisms and Clinical Outcomes in ALK-Positive Lung Cancer. *J Clin Oncol* 36:JCO2017762294, 2018
18. Baumann H, Kunapuli P, Tracy E, et al: The oncogenic fusion protein-tyrosine kinase ZNF198/fibroblast growth factor receptor-1 has signaling function comparable with interleukin-6 cytokine receptors. *J Biol Chem* 278:16198-208, 2003
19. Medves S, Noel LA, Montano-Almendras CP, et al: Multiple oligomerization domains of KANK1-PDGFRbeta are required for JAK2-independent hematopoietic cell proliferation and signaling via STAT5 and ERK. *Haematologica* 96:1406-14, 2011
20. Richards MW, O'Regan L, Roth D, et al: Microtubule association of EML proteins and the EML4-ALK variant 3 oncoprotein require an N-terminal trimerization domain. *Biochem J* 467:529-36, 2015
21. Tognon CE, Mackereth CD, Somasiri AM, et al: Mutations in the SAM domain of the ETV6-NTRK3 chimeric tyrosine kinase block polymerization and transformation activity. *Mol Cell Biol* 24:4636-50, 2004
22. Zhao X, Ghaffari S, Lodish H, et al: Structure of the Bcr-Abl oncoprotein oligomerization domain. *Nat Struct Biol* 9:117-20, 2002
23. Wiesner T, Lee W, Obenauf AC, et al: Alternative transcription initiation leads to expression of a novel ALK isoform in cancer. *Nature* 526:453-7, 2015
24. Armstrong F, Duplantier MM, Tremplat P, et al: Differential effects of X-ALK fusion proteins on proliferation, transformation, and invasion properties of NIH3T3 cells. *Oncogene* 23:6071-82, 2004
25. Armstrong F, Lamant L, Hieblot C, et al: TPM3-ALK expression induces changes in cytoskeleton organisation and confers higher metastatic capacities than other ALK fusion proteins. *Eur J Cancer* 43:640-6, 2007
26. Yang J, Nishihara R, Zhang X, et al: Energy sensing pathways: Bridging type 2 diabetes and colorectal cancer? *J Diabetes Complications* 31:1228-1236, 2017
27. Wang J, Duncan D, Shi Z, et al: WEB-based GENE SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res* 41:W77-83, 2013

28. Coffin CM, Patel A, Perkins S, et al: ALK1 and p80 expression and chromosomal rearrangements involving 2p23 in inflammatory myofibroblastic tumor. *Mod Pathol* 14:569-76, 2001
29. Coffin CM, Hornick JL, Fletcher CD: Inflammatory myofibroblastic tumor: comparison of clinicopathologic, histologic, and immunohistochemical features including ALK expression in atypical and aggressive cases. *Am J Surg Pathol* 31:509-20, 2007
30. Childress MA, Himmelberg SM, Chen H, et al: ALK Fusion Partners Impact Response to ALK Inhibition: Differential Effects on Sensitivity, Cellular Phenotypes, and Biochemical Properties. *Mol Cancer Res*, 2018
31. Richards MW, Law EW, Rennalls LP, et al: Crystal structure of EML1 reveals the basis for Hsp90 dependence of oncogenic EML4-ALK by disruption of an atypical beta-propeller domain. *Proc Natl Acad Sci U S A* 111:5195-200, 2014
32. Gouzi JY, Moog-Lutz C, Vigny M, et al: Role of the subcellular localization of ALK tyrosine kinase domain in neuronal differentiation of PC12 cells. *J Cell Sci* 118:5811-23, 2005
33. Mazot P, Cazes A, Boutterin MC, et al: The constitutive activity of the ALK mutated at positions F1174 or R1275 impairs receptor trafficking. *Oncogene* 30:2017-25, 2011
34. Soda M, Choi YL, Enomoto M, et al: Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature* 448:561-6, 2007
35. Hegyi H, Buday L, Tompa P: Intrinsic structural disorder confers cellular viability on oncogenic fusion proteins. *PLoS Comput Biol* 5:e1000552, 2009
36. Penzel R, Schirmacher P, Warth A: A novel EML4-ALK variant: exon 6 of EML4 fused to exon 19 of ALK. *J Thorac Oncol* 7:1198-9, 2012
37. Doebele RC, Pilling AB, Aisner DL, et al: Mechanisms of resistance to crizotinib in patients with ALK gene rearranged non-small cell lung cancer. *Clin Cancer Res* 18:1472-82, 2012
38. Kallberg M, Wang H, Wang S, et al: Template-based protein structure modeling using the RaptorX web server. *Nat Protoc* 7:1511-22, 2012
39. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

APPENDIX

Supplemental Tables

	Category	Term	Count	P Value	Genes	Fold Enrichment	Bonferroni	FDR
Up-regulated	KEGG PATHWAY	hsa04141:Protein processing in endoplasmic reticulum	7	1.06E-05	<i>HSPH1, HERPUD1, XBP1, CRYAB, HSPA1A, HSPA1B, DNAJB1</i>	12.44	6.02E-04	0.010
	Biological Processes	GO:1900034~regulation of cellular response to heat	6	1.16E-06	<i>HSPH1, CRYAB, BAG3, HSPA1A, HSPA1B, DNAJB1</i>	31.24	4.53E-04	0.002
	Biological Processes	GO:0010941~regulation of cell death	4	2.37E-06	<i>CRYAB, JUND, HSPA1A, HSPA1B</i>	142.00	9.23E-04	0.003
	Biological Processes	GO:1902236~negative regulation of endoplasmic reticulum stress-induced intrinsic apoptotic signaling pathway	4	1.37E-05	<i>HERPUD1, XBP1, HSPA1A, HSPA1B</i>	82.21	0.005	0.019
Down-regulated	Biological Processes	GO:0006954~inflammatory response	6	3.84E-05	<i>CYBB, MYD88, LIPA, LYN, S100A9, LGALS9</i>	13.99	0.012	0.051

Table S2.1: KEGG and GO term enrichment analysis for Differentially Expressed Genes in Pulmonary vs Extra-pulmonary IMT.

The 74 genes exhibiting significantly different expression in pulmonary tumors vs extra-pulmonary tumors ($FC \geq |2|$, $p < .05$) were used for KEGG Pathway and Gene Ontology term enrichment analysis. Up-regulated and down-regulated genes were assessed separately. Shown are the analyses results displaying an $FDR \leq .05$. The 'category' column references the type of test performed. The 'count' column displays the number of genes associated with the term that were in the set of differentially expressed genes evaluated. Statistics are calculated based on the count and the expected count.

	Category	Term	Count	P Value	Genes	Fold Enrichment	Bonferro ni	FDR
Up-regulated	Biological Processes	GO:0006954~inflammatory response	17	1.35E-06	<i>C3AR1, LIPA, S100A8, LYN, AIF1, HCK, LY86, S100A9, LYZ, ITGB2, CCL4, LGALS9, MIF, CYBA, CYBB, MYD88, PYCARD</i>	4.46	0.002	0.002
	Biological Processes	GO:0050900~leukocyte migration	9	3.11E-05	<i>C3AR1, PTPN6, LYN, GRB2, FCER1G, ITGB2, SIRPA, SLC7A7, MIF</i>	7.33	0.038	0.051
	KEGG PATHWAY	hsa04142:Lysosome	15	1.94E-09	<i>LIPA, AP1B1, PSAP, LGMN, CTSS, M6PR, DNASE2, ATP6V0C, CD68, AP1S2, LAPTM5, NAGA, CTSD, MCOLN1, ATP6V0D1</i>	8.32	3.10E-07	2.34E-06
	KEGG PATHWAY	hsa04145:Phagosome	12	1.38E-05	<i>ATP6V0C, CYBA, CYBB, CORO1A, ITGB2, CTSS, FCGR3A, ATP6V0D1, M6PR, TUBB3, TUBA1C, ATP6V1F</i>	5.26	0.002	0.017
	Cellular Components	GO:0070062~extracellular exosome	73	1.39E-16	<i>S100A8, CNDP2, FERMT3, LGMN, PGD, S100A9, CD53, RPS2, SYNGR2, YBX1, GPX1, FAM49B, CREG1, H2AFY, SERPINA1, H2AFX, CSK, ATP6V0D1, FCGR3A, TUBB3, RHOG, FTL, NT5C, LAIR1, LYN, LYZ, TCN2, SIRPA, VASP, ATP6V1F, GLUL, RENBP, DNPH1, CD320, NME1, VAMP8, CTSD, CSTA, RAB13, VSIG4, LCP1, BID, AHCY, GRB2, APOC1, ITGB2, TPM3, MIF, ATP6V0C,</i>	2.69	2.68E-14	1.44E-13

					<i>HNRNPM, TPI1, ARPC3, NAGA, IDH1, GALE, ARHGDIB, PTPN6, LIPA, PSAP, ALYREF, TOMM40, S100A10, COTL1, LGALS9, DNASE2, LSP1, HYOU1, P2RX4, CORO1A, BLVRB, H3F3A, LTA4H, SH3BGRL3</i>			
	Cellular Components	GO:0005829~cytosol	61	2.48E-07	<i>AP1B1, S100A8, CNDP2, AIF1, AP2S1, PGD, S100A9, DTYMK, RPS2, GPX1, AP1S2, MYD88, HMOX1, CSK, IMPDH1, DUS1L, RHOG, NT5C, FTL, IRAK1, LYN, EXOSC4, HMGA1, VASP, ATP6V1F, GLUL, RENBP, ARRB2, VAMP8, NME1, RNF135, PARVB, SRXN1, LCP1, BID, AHCY, GRB2, MAPKAPK3, UPP1, CYTH4, RPL38, TPM3, TPI1, ARPC3, SYN1, PYCARD, IDH1, GALE, ARHGDIB, ODC1, PTPN6, PLEK, AIMP2, HCK, ALYREF, NPL, CORO1A, PLK2, ID2, BLVRB, LTA4H</i>	1.91	5.98E-05	3.20E-04
	Cellular Components	GO:0016020~membrane	42	1.96E-05	<i>BID, ORAI1, LST1, PNKD, FERMT3, ITGB2, EHBP1L1, RPS2, HNRNPM, LPXN, CD68, FAM49B, ARPC3, HMOX1, ATP6V0D1, ARHGDIB, FTL, DDX39A, PTPN6, PLEK, MPP1, AIMP2, EFTUD2, HCLS1, ALYREF, M6PR, SIRPA, PNPLA6, CD163, ATP6V1F, LSP1, PPIF, P2RX4, CYBA, HYOU1, CORO1A, CD320, VAMP8, NME1, GRK6, NCLN, CLN6</i>	1.98	0.005	0.025
Ⓞ	KEGG	hsa04610:Complement	6	3.90E-06	<i>PLAT, CFH, SERPING1, C1R,</i>	23.11	2.11E-04	0.004

PATHWAY	and coagulation cascades			<i>C1S, PLAU</i>				
Biological Processes	GO:0030198~extracellular matrix organization	9	6.36E-08	<i>LUM, COL6A3, COL1A2, ITGB5, COL6A1, POSTN, MFAP2, DCN, CYR61</i>	16.06	2.78E-05	8.98E-05	
Biological Processes	GO:0007155~cell adhesion	9	3.61E-05	<i>WISP1, COL6A3, ITGB5, COL6A1, POSTN, MFAP4, THBS2, GAS6, CYR61</i>	6.86	0.016	0.051	
Cellular Components	GO:0031012~extracellular matrix	16	9.40E-16	<i>PLAT, LUM, EFEMP1, POSTN, DCN, MMP14, ECM1, TIMP3, HTRA1, FBLN2, COL6A3, COL1A2, COL6A1, MFAP4, THBS2, CYR61</i>	19.70	8.17E-14	9.66E-13	
Cellular Components	GO:0005576~extracellular region	24	3.03E-12	<i>PLAT, IL6ST, LUM, EFEMP1, SERPING1, C1R, C1S, DCN, ECM1, TIMP3, GAS6, CLEC11A, HTRA1, FBLN2, COL6A3, CFH, COL1A2, COL6A1, MFAP2, MFAP4, THBS2, PLAU, ANGPTL4, CYR61</i>	5.43	2.79E-10	3.31E-09	
Cellular Components	GO:0005578~proteinaceous extracellular matrix	12	1.01E-10	<i>CTHRC1, WISP1, FBLN2, LUM, COL6A3, EFEMP1, COL1A2, POSTN, ECM1, TIMP3, CYR61, ANGPTL4</i>	16.32	9.27E-09	1.10E-07	
Cellular Components	GO:0005615~extracellular space	19	4.92E-09	<i>PLAT, CTHRC1, IL6ST, LUM, EFEMP1, POSTN, SERPING1, DCN, ECM1, TIMP3, GAS6, CLEC11A, WISP1, HTRA1, COL6A3, CFH, COL1A2, PLAU, ANGPTL4</i>	5.14	4.52E-07	5.37E-06	
Cellular Components	GO:0070062~extracellular exosome	25	3.64E-08	<i>SLC44A1, IL6ST, LUM, ITGB5, GJA1, C1R, C1S, TIMP3, HTRA1, COL6A3, CFH, COL6A1, NDRG1, NT5E, PLAT, CLMP, EFEMP1, SERPING1, ECM1, GAS6,</i>	3.24	3.35E-06	3.98E-05	

					<i>FBLN2, CYBRD1, COL1A2, MFAP4, PLAU</i>			
	Cellular Components	GO:0005589~collagen type VI trimer	3	2.12E-05	COL6A3, COL6A1, DCN	364.48	0.002	0.023

Table S2.2: KEGG and GO term enrichment analysis for Differentially Expressed Genes in RANBP2-ALK Expressing IMTs vs. All Other IMTs.

The 235 genes exhibiting significantly different expression in RANBP2-ALK expressing tumors vs. all other IMTs ($FC \geq 2$, $p < .05$) were used for KEGG Pathway and Gene Ontology term enrichment analysis. Up-regulated and down-regulated genes were assessed separately. Shown are the analyses with an $FDR \leq .05$. The 'category' column references the type of test performed. The 'count' column displays the number of genes associated with the term that were in the set of differentially expressed genes evaluated. Statistics are calculated based on the count and the expected count.

	Category	Term	Count	P Value	Genes	Fold Enrichment	Bonferroni	FDR
Up-regulated	Cellular Components	GO:0005615~extracellular space	14	2.44E-06	<i>APOL1, RNASET2, FAM20C, SEMA4B, CMTM7, SERPINB4, LBP, CTSB, CFI, C2, QSOX1, GOLM1, TIMP1, CTSW</i>	4.74	1.98E-04	0.003
	Cellular Components	GO:0070062~extracellular exosome	18	2.53E-05	<i>HIST2H2AA3, HIST2H2AA4, FAM20C, COLEC12, CD74, TIMP1, ACVR1B, RAC2, RNASET2, HSPA7, SERPINB4, C2, CFI, CTSB, LBP, QSOX1, GOLM1, CD7</i>	2.92	0.002	0.027
Down-regulated	KEGG PATHWAY	ptr01130:Biosynthesis of antibiotics	10	1.52E-05	<i>ALDH1B1, ACO1, ECHS1, ACLY, PCYOX1, CAT, HADH, ACSS2, ALDH3A2, MDH1</i>	6.48	0.002	0.018
	Cellular Components	GO:0005578~proteinaceous extracellular matrix	8	3.78E-05	<i>ASPN, ADAMTS1, SPARC, COL5A3, COL16A1, ECM2, SPON2, SLIT3</i>	8.58	0.003	0.042

Table S2.3: KEGG and GO term enrichment analysis for Differentially Expressed Genes in Fusion Positive vs Fusion Negative IMT.

The 145 genes exhibiting significantly different expression in fusion positive tumors vs fusion negative tumors ($FC \geq 2$, $p < .05$) were used for KEGG Pathway and Gene Ontology term enrichment analysis. Up-regulated and down-regulated genes were assessed separately. Shown are the analyses with an $FDR \leq .05$. The 'category' column references the type of test performed. The 'count' column displays the number of genes associated with the term that were in the set of differentially expressed genes evaluated. Statistics are calculated based on the count and the expected count.

Pathway Name	FDR	Status	KEGG Link
Cell cycle	3.60E-07	Inhibited	http://www.genome.jp/dbget-bin/show_pathway?hsa04110+4171+4172+4173+4174+4176+23594+4998+64682+8697+983+701+5111+5347+4085+699+8556+2810+1111+5591+6502+9088+9232+5925+7157+991+995+8318+990+993+891+9133+890+902+898+7272+1031+1869+1871+3065
Neuroactive ligand-receptor interaction	0.001	Activated	http://www.genome.jp/dbget-bin/show_pathway?hsa04080+5645+3952+2690+2908+2904+9568+2918+7433+6344+117+2696+1394+134+7253+3973+56923+4886+2358+153+1134+1145

Table S2.4: Signaling Pathway Impact Analysis for Differentially Expressed Genes in Pulmonary vs Extra-pulmonary IMT

The 74 genes exhibiting significantly different expression in pulmonary tumors ($FC \geq |2|$, $p < .05$) were used for SPIA analysis. Shown are the analyses with an $FDR \leq .05$.

Pathway Name	FDR	Status	KEGG Link
Neuroactive ligand-receptor interaction	0.002	Inhibited	http://www.genome.jp/dbget-bin/show_pathway?hsa04080+1511+5645+2743+2903+5025+2917+6344+5745+1902+5724+2859+5028+5032+134+5737+5732+5729+3973+2149+2151+2357+2358+2359+719+185+3351+3356+1131+1134
ECM-receptor interaction	0.003	Inhibited	http://www.genome.jp/dbget-bin/show_pathway?hsa04512+284217+3908+1278+1291+1293+1311+7058+3685+3693+22801+8515+5649
Fc gamma R-mediated phagocytosis	0.009	Activated	http://www.genome.jp/dbget-bin/show_pathway?hsa04666+2214+2209+3055+4067+6850+23533+6199+10451+7409+5879+10810+10093+10094+10095+10109+10552+998+8976+1794+7408+3985+1072+55616+382+1759+653361+56848
Lysosome	0.011	Inhibited	http://www.genome.jp/dbget-bin/show_pathway?hsa04142+245972+527+533+537+9114+1509+1511+1513+1519+1520+1522+5641+2548+4668+2990+3988+1777+53+427+5660+768239+968+1497+26503+7805+1201+256471+1211+4074+51172+162+8905+8907+8943+23062+26088+57192
Fc epsilon RI signaling pathway	0.011	Activated	http://www.genome.jp/dbget-bin/show_pathway?hsa04664+2207+2206+2534+7124+5609+23533+695+6850+4067+5605+3845+4893+2885+5322+5879+5881+10451+7409
Natural killer cell mediated cytotoxicity	0.017	Activated	http://www.genome.jp/dbget-bin/show_pathway?hsa04650+5533+2534+7124+6452+6850+10870+2207+7305+5777+637+355+80329+100507436+23533+5605+369+3845+4893+2885+53358+5879+5881+10451+7409+3455+2214+3689
Regulation of actin cytoskeleton	0.024	Inhibited	http://www.genome.jp/dbget-bin/show_pathway?hsa04810+10093+10094+10095+10109+10552+5216+8976+10152+10787+3071+23191+10458+1072+3985+7414+1729+4478+5829+5499+4638+55740+5063+998+5879+5881+387+10451+7409+5605+369+23533+55970+1131+22801+3684+3685+3689+3693+8515+3845+4893+2260+3645+2247+2259+56034+2149
Chemokine signaling pathway	0.025	Activated	http://www.genome.jp/dbget-bin/show_pathway?hsa04062+1230+1234+729230+2919+388372+6348+6349+6351+6356+6357+2268+3055+4067+5829+387+23533+53358+10681+2783+2787+2793+55970+2885+2870+408+409+10451+7409+998+108+111+5906+8976+653361+5566+3845+4893+1794+5879+4793+2931
Tuberculosis	0.025	Activated	http://www.genome.jp/dbget-bin/show_pathway?hsa05152+22925+5878+245972+527+533+537+9114+841

			1+1509+808+817+56848+2209+2214+11151+5533+3689+3684+4802+4615+7099+7096+7124+637+3112+1520+387+4046+64581+6850+2207+7040+3654+3588
Calcium signaling pathway	0.041	Inhibited	http://www.genome.jp/dbget-bin/show_pathway?hsa04020+4842+7416+10105+817+5533+683+2903+5025+56848+4638+84812+5261+84876+1131+185+2149+3356+3973+5724+5737+808+292+293+108+5566+8913
Pancreatic secretion	0.041	Inhibited	http://www.genome.jp/dbget-bin/show_pathway?hsa04972+1131+6558+6344+6522+5645+1359+5322+7220+683+108+111+4218+9545+5906+387+5879

Table S2.5: Signaling Pathway Impact Analysis for Differentially Expressed Genes in RANBP2-ALK Expressing IMTs vs. All Other IMTs.

The 235 genes exhibiting significantly different expression in RANBP2-ALK expressing tumors ($FC \geq 2$, $p < .05$) were used SPIA. Shown are the analyses with an $FDR \leq .05$.

Pathway Name	FDR	Status	KEGG Link
PPAR signaling pathway	0.002	Inhibited	http://www.genome.jp/dbget-bin/show_pathway?hsa03320+364+5105+5346+2167+6258+5468+34+33+1962+4023+10580+6319+1622+4199
Cytokine-cytokine receptor interaction	0.002	Inhibited	http://www.genome.jp/dbget-bin/show_pathway?hsa04060+91+3561+3560+8809+3556+3554+657+658+3606+943+939+356+4049+1956+3586+3082+56034+2690+3559+1438+7173+1271+3590+3592+3952+1440+1234+10344+6356+10663+2833+3577+3579+9547+7133+2920
Amoebiasis	0.033	Inhibited	http://www.genome.jp/dbget-bin/show_pathway?hsa05146+10319+3554+3592+5293+5330+5613+3586+1278+50509+7414+5055+6317+6318

Table S2.6: Signaling Pathway Impact Analysis for Differentially Expressed Genes in Fusion Positive vs Fusion Negative IMT.

The 145 genes exhibiting significantly different expression in fusion positive tumors ($FC \geq |2|$, $p < .05$) were used for SPIA. Shown are the analyses with an $FDR \leq .05$.

Gene	SNV (amino acid change)	# Cases Harboring SNV
<i>RPL22</i>	44 A duplication (K15 frame shift)	18
<i>TP53</i>	C98G (P33R)	17
<i>DDX31</i>	A2395G (I799V)	17
<i>IL7R</i>	G412A (V138I)	16
<i>ZNF384</i>	1194_1196 deletion (398_399 deletion)	14
<i>SETD2</i>	C5885T (P1962L)	14
<i>ATRX</i>	2404 A duplication (R802 frame shift)	13
<i>NCOR1</i>	636 A duplication (V213 frame shift)	11
<i>DHX15</i>	1224 A duplication (Q409 frame shift)	10
<i>CROCC</i>	G5337C (E1779D)	10
<i>IGLL5</i>	C67T (R23C)	10
<i>SIRPA</i>	C300A (N100K)	7
<i>TET2</i>	T5162G (L1721W)	8
<i>RHOA</i>	A526G (R176G)	9
<i>NUP98</i>	C3424G (Q1142E)	7
<i>MTAP</i>	G166A (V56I)	7
<i>ZCCHC7</i>	1617 A duplication (R539 frame shift)	8
<i>KDR</i>	A1416T (Q472H)	7
<i>JMJD1C</i>	C5393T (P1798L)	7
<i>TSC1</i>	T812C (M271T)	5
<i>CHD4</i>	396_398 deletion (132_133 deletion)	6
<i>LRIG1</i>	C2705T (A902V)	6
<i>FLT4</i>	C1804T (H602Y)	7
<i>KMT2A</i>	9308 A duplication (Q3103 frame shift)	5
<i>OSMR</i>	G1733A (G578D)	5
<i>NOTCH1</i>	G3836A (R1279H)	6
<i>GAK</i>	A3500G (K1167R)	6
<i>KMT2C</i>	13940 A duplication (K4647 frame shift)	6
<i>MGA</i>	6953 A duplication (Q2318 frame shift)	5
<i>ZFH3</i>	G5146A (D1716N)	5

Table S2.7: Cancer Associated SNVs Shared by at Least 25% of IMTs Evaluated. All 18 IMTs undergoing RNAseq analysis were assessed for SNVs. Variants were filtered based on a list of 383 pre-defined cancer genes. The variants were maintained if they were exonic, had a damaging function, previously reported in COSMIC, and were considered pathogenic in the ClinVar Database. Variants were then excluded if they had a variant allele frequency below 0.05, depth below 10, a prevalence of greater than 10% in the 1000 Genomes Project or if no annotations were present. Variants present in at least 25% of samples are displayed here.

	Chromosome	Start	End	Reference	Alteration	Gene	Exonic Function	AA Change	Read Depth	VAF
Case 1	chr17	7676154	7676154	G	C	<i>TP53</i>	Non-synonymous SNV	TP53:NM_001126118:exon3:c.C98G:p.P33R,TP53:NM_000546:exon4:c.C215G:p.P72R,TP53:NM_001126112:exon4:c.C215G:p.P72R,TP53:NM_001126113:exon4:c.C215G:p.P72R,TP53:NM_001126114:exon4:c.C215G:p.P72R,TP53:NM_001276695:exon4:c.C98G:p.P33R,TP53:NM_001276696:exon4:c.C98G:p.P33R,TP53:NM_001276760:exon4:c.C98G:p.P33R,TP53:NM_001276761:exon4:c.C98G:p.P33R	34	1.00
	chr3	47083895	47083895	G	A	<i>SETD2</i>	Non-synonymous SNV	SETD2:NM_014159:exon12:c.C5885T:p.P1962L	38	1.00
	chr5	35860966	35860966	T	C	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon2:c.T197C:p.I66T	82	1.00
	chr5	35871088	35871088	G	A	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon4:c.G412A:p.V138I	52	1.00
	chr5	35874473	35874473	C	T	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon6:c.C731T:p.T244I	28	1.00
	chrX	77682471	77682471	C	G	<i>ATRX</i>	Non-synonymous SNV	ATRX:NM_138270:exon8:c.G2671C:p.E891Q,ATRX:NM_000489:exon9:c.G2785C:p.E929Q	12	1.00

chr5	134115992	134115992	C	A	<i>TCF7</i>	Non-synonymous SNV	TCF7:NM_213648:exon1:c.C55A:p.P19T,TCF7:NM_001134851:exon2:c.C55A:p.P19T,TCF7:NM_201632:exon2:c.C55A:p.P19T,TCF7:NM_201634:exon2:c.C55A:p.P19T,TCF7:NM_003202:exon3:c.C400A:p.P134T	11	0.73
chr19	41238005	41238005	C	T	<i>AXL</i>	Non-synonymous SNV	AXL:NM_001278599:exon4:c.C41T:p.P14L,AXL:NM_001699:exon7:c.C845T:p.P282L,AXL:NM_021913:exon7:c.C845T:p.P282L	11	0.64
chr7	92774737	92774737	C	T	<i>CDK6</i>	Non-synonymous SNV	CDK6:NM_001145306:exon3:c.G328A:p.D110N,CDK6:NM_001259:exon3:c.G328A:p.D110N	10	0.60
chr4	105275794	105275794	A	G	<i>TET2</i>	Non-synonymous SNV	TET2:NM_001127208:exon11:c.A5284G:p.I1762V	33	0.55
chr19	34466925	34466925	C	T	<i>UBA2</i>	Non-synonymous SNV	UBA2:NM_005499:exon16:c.C1652T:p.P551L	224	0.53
chr9	132911517	132911517	A	G	<i>TSC1</i>	Non-synonymous SNV	TSC1:NM_001162427:exon9:c.T812C:p.M271T,TSC1:NM_000368:exon10:c.T965C:p.M322T,TSC1:NM_001162426:exon10:c.T965C:p.M322T	21	0.52
chr3	179209622	179209622	A	G	<i>PIK3CA</i>	Non-synonymous SNV	PIK3CA:NM_006218:exon7:c.A1173G:p.I391M	14	0.50
chr1	92837557	92837557	A	G	<i>RPL5</i>	Non-synonymous SNV	RPL5:NM_000969:exon6:c.A629G:p.Y210C	3446	0.48
chr12	6667904	6667906	TGC	-	<i>ZNF384</i>	Non-frameshift deletion	ZNF384:NM_001039920:exon9:c.1194_1196del:p.398_399del,ZNF384:NM_133476:exon10:c.1359_1361del:p.453_454del,ZNF384:N	24	0.46

							M_001135734:exon11:c.1542_1544del:p.514_515del		
chr11	3702551	3702551	G	C	<i>NUP98</i>	Non-synonymous SNV	NUP98:NM_016320:exon23:c.C3424G:p.Q1142E, NUP98:NM_139132:exon23:c.C3424G:p.Q1142E	55	0.45
chr19	41253609	41253609	C	T	<i>AXL</i>	Non-synonymous SNV	AXL:NM_001278599:exon14:c.C1133T:p.T378I, AXL:NM_001699:exon16:c.C1910T:p.T637I, AXL:NM_021913:exon17:c.C1937T:p.T646I	19	0.42
chr9	132595027	132595027	T	C	<i>DDX31</i>	Non-synonymous SNV	DDX31:NM_022779:exon20:c.A2395G:p.I799V	29	0.41
chr4	105235006	105235006	G	A	<i>TET2</i>	Non-synonymous SNV	TET2:NM_001127208:exon3:c.G1064A:p.G355D, TET2:NM_017628:exon3:c.G1064A:p.G355D	10	0.40
chr2	43225479	43225479	C	T	<i>ZFP36L2</i>	Non-synonymous SNV	ZFP36L2:NM_006887:exon2:c.G325A:p.G109S	62	0.37
chr9	21816759	21816759	G	A	<i>MTAP</i>	Non-synonymous SNV	MTAP:NM_002451:exon3:c.G166A:p.V56I	26	0.35
chr20	1915319	1915319	C	A	<i>SIRPA</i>	Non-synonymous SNV	SIRPA:NM_001040023:exon2:c.C300A:p.N100K, SIRPA:NM_001040022:exon3:c.C300A:p.N100K, SIRPA:NM_080792:exon3:c.C300A:p.N100K	301	0.31
chr20	1915303	1915303	-	GT	<i>SIRPA</i>	frameshift insertion	SIRPA:NM_001040023:exon2:c.284_285insGT:p.D95fs, SIRPA:NM_001040022:exon3:c.284_285insGT:p.D95fs, SIRPA:NM_080792:exon3:c.284_285insGT:p.D95fs	266	0.31

chr20	1915305	1915306	CT	-	SIRPA	frameshift deletion	SIRPA:NM_001040023:exon2:c.286_287del:p.L96fs,SIRPA:NM_001040022:exon3:c.286_287del:p.L96fs,SIRPA:NM_080792:exon3:c.286_287del:p.L96fs	276	0.30
chr17	16146494	16146494	-	T	NCOR1	frameshift insertion	NCOR1:NM_001190438:exon7:c.636dupA:p.V213fs,NCOR1:NM_001190440:exon9:c.963dupA:p.V322fs,NCOR1:NM_006311:exon10:c.963dupA:p.V322fs	39	0.26
chr20	1915338	1915338	C	A	SIRPA	Non-synonymous SNV	SIRPA:NM_001040023:exon2:c.C319A:p.R107S,SIRPA:NM_001040022:exon3:c.C319A:p.R107S,SIRPA:NM_080792:exon3:c.C319A:p.R107S	294	0.23
chr1	6197724	6197724	-	T	RPL22	frameshift insertion	RPL22:NM_000983:exon2:c.44dupA:p.K15fs	3177	0.22
chr20	1915406	1915408	CGA	-	SIRPA	Non-frameshift deletion	SIRPA:NM_001040023:exon2:c.387_389del:p.129_130del,SIRPA:NM_001040022:exon3:c.387_389del:p.129_130del,SIRPA:NM_080792:exon3:c.387_389del:p.129_130del	255	0.20
chr3	49360949	49360949	T	C	RHOA	Non-synonymous SNV	RHOA:NM_001313943:exon5:c.A526G:p.R176G	15	0.20
chr4	24548878	24548878	-	T	DHX15	frameshift insertion	DHX15:NM_001358:exon6:c.1224dupA:p.Q409fs	218	0.20
chr20	1915344	1915344	G	A	SIRPA	Non-synonymous SNV	SIRPA:NM_001040023:exon2:c.G325A:p.G109S,SIRPA:NM_001040022:exon3:c.G325A:p.G109S,SIRPA:NM_080792:exon3:c.G325A:p.G109S	293	0.19

Table S2.8: All Cancer Associated SNVs Called for Case 1. Variants were filtered based on a list of 383 pre-defined cancer genes. The variants were maintained if they were exonic, had a damaging function, previously reported in COSMIC, and were considered pathogenic in the ClinVar Database. Variants were then excluded if they had a variant allele frequency below 0.05, depth below 10, a prevalence of greater than 10% in the 1000 Genomes Project, or if no annotations were present. SNVs are shaded dark orange if the variant allele frequency (VAF) =1, and lighter shades of orange for VAFs $\geq .5$ and < 1 and VAFs $< .5$ and $> .4$. The 'AA Change' column lists the ref.seq. number, the base pair alteration, and the amino acid alteration.

	Chromosome	Start	End	Reference	Alteration	Gene	Exonic Function	AA Change	Read Depth	VAF
Case 2	chr12	6667904	6667906	TGC	-	ZNF384	Non-frameshift deletion	ZNF384:NM_001039920:exon9:c.1194_1196del:p.398_399del,ZNF384:NM_133476:exon10:c.1359_1361del:p.453_454del,ZNF384:NM_001135734:exon11:c.1542_1544del:p.514_515del	19	1.00
	chr2	29193500	29193500	G	C	ALK	Non-synonymous SNV	ALK:NM_004304:exon29:c.C4587G:p.D1529E	59	1.00
	chr2	29193615	29193615	T	C	ALK	Non-synonymous SNV	ALK:NM_004304:exon29:c.A4472G:p.K1491R	52	1.00
	chr2	29222392	29222392	C	T	ALK	Non-synonymous SNV	ALK:NM_004304:exon22:c.G3467A:p.C1156Y	21	1.00
	chr20	1915303	1915303	-	GT	SIRPA	frameshift insertion	SIRPA:NM_001040023:exon2:c.284_285insGT:p.D95fs,SIRPA:NM_001040022:exon3:c.284_285insGT:p.D95fs,SIRPA:NM_080792:exon3:c.284_285insGT:p.D95fs	11	1.00
	chr20	1915305	1915306	CT	-	SIRPA	frameshift deletion	SIRPA:NM_001040023:exon2:c.286_287del:p.L96fs,SIRPA:NM_001040022:exon3:c.286_287del:p.L96fs,SIRPA:NM_080792:exon3:c.286_287del:p.L96fs	11	1.00
	chr20	1915319	1915319	C	A	SIRPA	Non-synonymous SNV	SIRPA:NM_001040023:exon2:c.C300A:p.N100K,SIRPA:NM_001040022:exon3:c.C300A:p.N100K,SIRPA:NM_080792:exon3:c.C300A:p.N100K	11	1.00

chr20	1915406	1915408	CGA	-	SIRPA	Non-frameshift deletion	SIRPA:NM_001040023:exon2:c.387_389del:p.129_130del,SIRPA:NM_001040022:exon3:c.387_389del:p.129_130del,SIRPA:NM_080792:exon3:c.387_389del:p.129_130del	11	1.00
chr7	22731420	22731420	T	A	IL6	Non-synonymous SNV	IL6:NM_000600:exon5:c.T486A:p.D162E	22	0.59
chr17	7676154	7676154	G	C	TP53	Non-synonymous SNV	TP53:NM_001126118:exon3:c.C98G:p.P33R,TP53:NM_000546:exon4:c.C215G:p.P72R,TP53:NM_001126112:exon4:c.C215G:p.P72R,TP53:NM_001126113:exon4:c.C215G:p.P72R,TP53:NM_001126114:exon4:c.C215G:p.P72R,TP53:NM_001276695:exon4:c.C98G:p.P33R,TP53:NM_001276696:exon4:c.C98G:p.P33R,TP53:NM_001276760:exon4:c.C98G:p.P33R,TP53:NM_001276761:exon4:c.C98G:p.P33R	25	0.52
chr6	44250093	44250093	G	A	HSP90AB1	Non-synonymous SNV	HSP90AB1:NM_001271969:exon5:c.G587A:p.R196K,HSP90AB1:NM_001271970:exon5:c.G587A:p.R196K,HSP90AB1:NM_001271971:exon5:c.G443A:p.R148K,HSP90AB1:NM_001271972:exon5:c.G557A:p.R186K,HSP90AB1:NM_007355:exon5:c.G587A:p.R196K	666	0.48
chr3	128486108	128486108	C	T	GATA2	Non-synonymous SNV	GATA2:NM_001145662:exon3:c.G490A:p.A164T,GATA2:NM_032638:exon3:c.G490A:p.A164T,GATA2:NM	13	0.46

							_001145661:exon4:c.G490 A:p.A164T		
chr5	35876347	35876347	C	T	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon8:c.C1241T:p.T414M	16	0.44
chr5	150123172	150123172	G	A	<i>PDGFRB</i>	Non-synonymous SNV	PDGFRB:NM_002609:exon15:c.C2053T:p.R685C	21	0.43
chr9	132595027	132595027	T	C	<i>DDX31</i>	Nonsynonymous SNV	DDX31:NM_022779:exon20:c.A2395G:p.I799V	39	0.33
chr5	35871088	35871088	G	A	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon4:c.G412A:p.V138I	13	0.31
chr5	38883969	38883969	T	G	<i>OSMR</i>	Non-synonymous SNV	OSMR:NM_001168355:exon5:c.T561G:p.H187Q,OSMR:NM_003999:exon5:c.T561G:p.H187Q	23	0.30
chr1	16946292	16946292	A	G	<i>CROCC</i>	Non-synonymous SNV	CROCC:NM_014675:exon16:c.A2170G:p.M724V	10	0.30
chr22	22893780	22893780	G	C	<i>IGLL5</i>	Non-synonymous SNV	IGLL5:NM_001178126:exon2:c.G287C:p.C96S	65	0.28
chrX	147937529	147937529	-	A	<i>FMR1</i>	frameshift insertion	FMR1:NM_001185075:exon11:c.1055dupA:p.E352fs, FMR1:NM_001185076:exon11:c.1055dupA:p.E352fs, FMR1:NM_001185081:exon11:c.1055dupA:p.E352fs, FMR1:NM_001185082:exon11:c.1055dupA:p.E352fs, FMR1:NM_002024:exon11:c.1055dupA:p.E352fs	16	0.25
chr2	144399760	144399760	-	T	<i>ZEB2</i>	frameshift insertion	ZEB2:NM_001171653:exon7:c.1354dupA:p.M452fs,ZEB2:NM_014795:exon8:c.1426dupA:p.M476fs	17	0.24

chr18	12836849	12836849	-	A	<i>PTPN2</i>	frameshift insertion	PTPN2:NM_001308287:exon2:c.115dupT:p.Y39fs,PTPN2:NM_001207013:exon3:c.202dupT:p.Y68fs,PTPN2:NM_002828:exon3:c.202dupT:p.Y68fs,PTPN2:NM_080422:exon3:c.202dupT:p.Y68fs,PTPN2:NM_080423:exon3:c.202dupT:p.Y68fs	98	0.21
chr1	6197724	6197724	-	T	<i>RPL22</i>	frameshift insertion	RPL22:NM_000983:exon2:c.44dupA:p.K15fs	3185	0.20
chr16	4260224	4260224	T	G	<i>TFAP4</i>	Non-synonymous SNV	TFAP4:NM_003223:exon6:c.A688C:p.T230P	20	0.20
chr3	49360990	49360990	T	C	<i>RHOA</i>	Non-synonymous SNV	RHOA:NM_001313943:exon5:c.A485G:p.K162R	10	0.20
chr4	24548878	24548878	-	T	<i>DHX15</i>	frameshift insertion	DHX15:NM_001358:exon6:c.1224dupA:p.Q409fs	69	0.13

Table S2.9: All Cancer Associated SNVs Called for Case 2. Variants were filtered based on a list of 383 pre-defined cancer genes. The variants were maintained if they were exonic, had a damaging function, previously reported in COSMIC, and were considered pathogenic in the ClinVar Database. Variants were then excluded if they had a variant allele frequency below 0.05, depth below 10, a prevalence of greater than 10% in the 1000 Genomes Project, or if no annotations were present. SNVs are shaded dark orange if the variant allele frequency (VAF) =1, and lighter shades of orange for VAFs>=.5 and<1 and VAFs <.5 and>.4. The 'AA Change' column lists the ref.seq. number, the base pair alteration, and the amino acid alteration.

	Chromosome	Start	End	Reference	Alteration	Gene	Exonic Function	AA Change	Read Depth	VAF
Case 3	chr12	6667904	6667906	TGC	-	<i>ZNF384</i>	Non-frameshift deletion	ZNF384:NM_001039920:exon9:c.1194_1196del:p.398_399del,ZNF384:NM_133476:exon10:c.1359_1361del:p.453_454del,ZNF384:NM_001135734:exon11:c.1542_1544del:p.514_515del	14	1.00
	chr3	47083895	47083895	G	A	<i>SETD2</i>	Non-synonymous SNV	SETD2:NM_014159:exon12:c.C5885T:p.P1962L	26	1.00
	chr16	3717472	3717472	G	A	<i>TRAP1</i>	Non-synonymous SNV	TRAP1:NM_001272049:exon1:c.C37T:p.R13C,TRAP1:NM_016292:exon1:c.C37T:p.R13C	17	0.65
	chr5	38921762	38921762	G	A	<i>OSMR</i>	Non-synonymous SNV	OSMR:NM_003999:exon12:c.G1733A:p.G578D	42	0.62
	chr3	179209622	179209622	A	G	<i>PIK3CA</i>	Non-synonymous SNV	PIK3CA:NM_006218:exon7:c.A1173G:p.I391M	13	0.62
	chr11	3679646	3679646	C	T	<i>NUP98</i>	Non-synonymous SNV	NUP98:NM_139132:exon30:c.G4759A:p.E1587K,NUP98:NM_016320:exon31:c.G4981A:p.E1661K	36	0.56
	chr10	63193075	63193075	G	A	<i>JMJD1C</i>	Non-synonymous SNV	JMJD1C:NM_001282948:exon15:c.C5393T:p.P1798L,JMJD1C:NM_032776:exon16:c.C5939T:p.P1980L	19	0.53

chr17	7676154	7676154	G	C	TP53	Non-synonymous SNV	TP53:NM_001126118:exon3:c.C98G:p.P33R,TP53:NM_000546:exon4:c.C215G:p.P72R,TP53:NM_001126112:exon4:c.C215G:p.P72R,TP53:NM_001126113:exon4:c.C215G:p.P72R,TP53:NM_001126114:exon4:c.C215G:p.P72R,TP53:NM_001276695:exon4:c.C98G:p.P33R,TP53:NM_001276696:exon4:c.C98G:p.P33R,TP53:NM_001276760:exon4:c.C98G:p.P33R,TP53:NM_001276761:exon4:c.C98G:p.P33R	68	0.50
chr12	14456621	14456621	A	T	ATF7IP	Non-synonymous SNV	ATF7IP:NM_001286514:exon7:c.A2053T:p.N685Y,ATF7IP:NM_001286515:exon7:c.A2053T:p.N685Y,ATF7IP:NM_018179:exon7:c.A2056T:p.N686Y,ATF7IP:NM_0181352:exon7:c.A2080T:p.N694Y	19	0.42
chr4	55106807	55106807	T	A	KDR	Non-synonymous SNV	KDR:NM_002253:exon11:c.A1416T:p.Q472H	27	0.33
chr1	16945542	16945542	G	A	CROCC	Non-synonymous SNV	CROCC:NM_014675:exon15:c.G2072A:p.R691H	16	0.25
chr2	112830485	112830485	T	C	IL1B	Non-synonymous SNV	IL1B:NM_000576:exon7:c.A686G:p.E229G	13	0.23
chr1	6197724	6197724	-	T	RPL22	frameshift insertion	RPL22:NM_000983:exon2:c.44dupA:p.K15fs	2370	0.22
chr1	16953387	16953387	T	A	CROCC	Non-synonymous SNV	CROCC:NM_014675:exon21:c.T3092A:p.L1031Q	16	0.19
chr1	16953389	16953389	G	A	CROCC	Non-synonymous	CROCC:NM_014675:exon21:c.G3094A:p.E1032K	16	0.19

						SNV			
chr1	16953404	16953404	G	C	<i>CROCC</i>	Non-synonymous SNV	CROCC:NM_014675:exon21:c.G3109C:p.E1037Q	16	0.19
chr17	59946728	59946728	-	A	<i>RPS6KB1</i>	frameshift insertion	RPS6KB1:NM_001272042:exon14:c.1450dupA:p.Y483fs,RPS6KB1:NM_001272060:exon15:c.1450dupA:p.Y483fs,RPS6KB1:NM_003161:exon15:c.1519dupA:p.Y506fs,RPS6KB1:NM_001272044:exon16:c.1360dupA:p.Y453fs	22	0.18
chr22	22893804	22893804	A	G	<i>IGLL5</i>	Non-synonymous SNV	IGLL5:NM_001178126:exon2:c.A311G:p.K104R	257	0.18
chr4	24548878	24548878	-	T	<i>DHX15</i>	frameshift insertion	DHX15:NM_001358:exon6:c.1224dupA:p.Q409fs	66	0.14
chr22	22893795	22893795	C	G	<i>IGLL5</i>	Non-synonymous SNV	IGLL5:NM_001178126:exon2:c.C302G:p.T101S	244	0.12

Table S2.10: All Cancer Associated SNVs Called for Case 3. Variants were filtered based on a list of 383 pre-defined cancer genes. The variants were maintained if they were exonic, had a damaging function, previously reported in COSMIC, and were considered pathogenic in the ClinVar Database. Variants were then excluded if they had a variant allele frequency below 0.05, depth below 10, a prevalence of greater than 10% in the 1000 Genomes Project, or if no annotations were present. SNVs are shaded dark orange if the variant allele frequency (VAF) =1, and lighter shades of orange for VAFs >=0.5 and <1 and VAFs <0.5 and >0.4. The 'AA Change' column lists the ref.seq. number, the base pair alteration, and the amino acid alteration.

	Chromosome	Start	End	Reference	Alteration	Gene	Exonic Function	AA Change	Read Depth	VAF
Case 4	chr11	3702551	3702551	G	C	<i>NUP98</i>	Non-synonymous SNV	NUP98:NM_016320:exon23:c.C3424G:p.Q1142E, NUP98:NM_139132:exon23:c.C3424G:p.Q1142E	13	1.00
	chr12	6667904	6667906	TGC	-	<i>ZNF384</i>	Non-frameshift deletion	ZNF384:NM_001039920:exon9:c.1194_1196del:p.398_399del, ZNF384:NM_133476:exon10:c.1359_1361del:p.453_454del, ZNF384:NM_001135734:exon11:c.1542_1544del:p.514_515del	11	1.00
	chr5	35860966	35860966	T	C	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon2:c.T197C:p.I66T	37	1.00
	chr5	35871088	35871088	G	A	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon4:c.G412A:p.V138I	40	1.00
	chrX	77682471	77682471	C	G	<i>ATRX</i>	Non-synonymous SNV	ATRX:NM_138270:exon8:c.G2671C:p.E891Q, ATRX:NM_000489:exon9:c.G2785C:p.E929Q	11	1.00
	chr20	32434666	32434666	G	A	<i>ASXL1</i>	Non-synonymous SNV	ASXL1:NM_015338:exon12:c.G1954A:p.G652S	14	0.71
	chr18	62979049	62979049	G	A	<i>PHLPP1</i>	Non-synonymous SNV	PHLPP1:NM_194449:exon17:c.G4772A:p.S1591N	11	0.64
	chr9	37126685	37126685	G	A	<i>ZCCHC7</i>	Non-synonymous SNV	ZCCHC7:NM_001289119:exon2:c.G353A:p.G118D, ZCCHC7:NM_001289120:exon2:c.G353A:p.G118D, ZCCHC7:NM_001289121:exon2:c.G353A:p.G118D, ZCCHC7:NM_032226:exon2:c.G353A:p.G118D	26	0.58

chr9	132595027	132595027	T	C	<i>DDX31</i>	Non-synonymous SNV	DDX31:NM_022779:exon20:c.A2395G:p.I799V	21	0.57
chr4	55113391	55113391	C	T	<i>KDR</i>	Non-synonymous SNV	KDR:NM_002253:exon7:c.G889A:p.V297I	18	0.56
chr4	849932	849932	T	C	<i>GAK</i>	Non-synonymous SNV	GAK:NM_001286833:exon24:c.A3500G:p.K1167R, GAK:NM_005255:exon27:c.A3794G:p.K1265R	194	0.52
chr5	35876149	35876149	A	C	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon8:c.A1043C:p.N348T	18	0.50
chr5	150070264	150070264	C	T	<i>CSF1R</i>	Non-synonymous SNV	CSF1R:NM_001288705:exon8:c.G1237A:p.G413S, CSF1R:NM_005211:exon9:c.G1237A:p.G413S	72	0.49
chr17	7676154	7676154	G	C	<i>TP53</i>	Non-synonymous SNV	TP53:NM_001126118:exon3:c.C98G:p.P33R, TP53:NM_000546:exon4:c.C215G:p.P72R, TP53:NM_001126112:exon4:c.C215G:p.P72R, TP53:NM_001126113:exon4:c.C215G:p.P72R, TP53:NM_001126114:exon4:c.C215G:p.P72R, TP53:NM_001276695:exon4:c.C98G:p.P33R, TP53:NM_001276696:exon4:c.C98G:p.P33R, TP53:NM_001276760:exon4:c.C98G:p.P33R, TP53:NM_001276761:exon4:c.C98G:p.P33R	33	0.48
chr2	208243593	208243593	C	T	<i>IDH1</i>	Non-synonymous SNV	IDH1:NM_001282386:exon6:c.G532A:p.V178I, IDH1:NM_001282387:exon6:c.G532A:p.V178I, IDH1:NM_005896:exon6:c.G532A:p.V178I	128	0.48
chr3	12612044	12612044	T	C	<i>RAF1</i>	Non-	RAF1:NM_002880:exon3:c	52	0.46

						synonymous SNV	.A226G:p.M76V		
chr15	41568909	41568909	A	G	<i>TYRO3</i>	Non-synonymous SNV	TYRO3:NM_006293:exon9:c.A1139G:p.N380S	10	0.40
chr5	35874473	35874473	C	T	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon6:c.C731T:p.T244I	20	0.40
chr1	81951044	81951044	T	C	<i>ADGRL2</i>	Non-synonymous SNV	ADGRL2:NM_001297705:exon7:c.T1519C:p.S507P,ADGRL2:NM_001297706:exon7:c.T1519C:p.S507P,ADGRL2:NM_012302:exon7:c.T1519C:p.S507P,ADGRL2:NM_001297704:exon8:c.T1519C:p.S507P	26	0.38
chr9	136500769	136500769	G	A	<i>NOTCH1</i>	Non-synonymous SNV	NOTCH1:NM_017617:exon31:c.C5717T:p.A1906V	26	0.35
chr4	55106807	55106807	T	A	<i>KDR</i>	Non-synonymous SNV	KDR:NM_002253:exon11:c.A1416T:p.Q472H	13	0.31
chr4	55106779	55106779	A	G	<i>KDR</i>	Non-synonymous SNV	KDR:NM_002253:exon11:c.T1444C:p.C482R	19	0.26
chr1	6197724	6197724	-	T	<i>RPL22</i>	frameshift insertion	RPL22:NM_000983:exon2:c.44dupA:p.K15fs	2818	0.23
chr9	37357252	37357252	-	A	<i>ZCCHC7</i>	frameshift insertion	ZCCHC7:NM_001289119:exon9:c.1617dupA:p.R539fs,ZCCHC7:NM_001289120:exon9:c.1617dupA:p.R539fs,ZCCHC7:NM_001289121:exon9:c.1617dupA:p.R539fs,ZCCHC7:NM_032226:exon9:c.1617dupA:p.R539fs	40	0.20
chr2	197416764	197416764	-	T	<i>SF3B1</i>	frameshift insertion	SF3B1:NM_012433:exon6:c.642dupA:p.L215fs	62	0.15

Table S2.11: All Cancer Associated SNVs Called for Case 4. Variants were filtered based on a list of 383 pre-defined cancer genes. The variants were maintained if they were exonic, had a damaging function, previously reported in COSMIC, and were considered pathogenic in the ClinVar Database. Variants were then excluded if they had a variant allele

frequency below 0.05, depth below 10, a prevalence of greater than 10% in the 1000 Genomes Project, or if no annotations were present. SNVs are shaded dark orange if the variant allele frequency (VAF) =1, and lighter shades of orange for VAFs ≥ 0.5 and < 1 and VAFs < 0.5 and > 0.4 . The 'AA Change' column lists the ref.seq. number, the base pair alteration, and the amino acid alteration.

	Chromosome	Start	End	Reference	Alteration	Gene	Exonic Function	AA Change	Read Depth	VAF
Case 5	chr3	47083895	47083895	G	A	SETD2	Non-synonymous SNV	SETD2:NM_014159:exon12:c.C5885T:p.P1962L	45	1.00
	chr5	68292320	68292320	G	A	PIK3R1	Non-synonymous SNV	PIK3R1:NM_181504:exon2:c.G168A:p.M56I,PIK3R1:NM_181524:exon2:c.G78A:p.M26I,PIK3R1:NM_181523:exon8:c.G978A:p.M326I	28	1.00
	chrX	77682471	77682471	C	G	ATRX	Non-synonymous SNV	ATRX:NM_138270:exon8:c.G2671C:p.E891Q,ATRX:NM_000489:exon9:c.G2785C:p.E929Q	10	1.00
	chr17	7676154	7676154	G	C	TP53	Non-synonymous SNV	TP53:NM_001126118:exon3:c.C98G:p.P33R,TP53:NM_000546:exon4:c.C215G:p.P72R,TP53:NM_001126112:exon4:c.C215G:p.P72R,TP53:NM_001126113:exon4:c.C215G:p.P72R,TP53:NM_001126114:exon4:c.C215G:p.P72R,TP53:NM_001276695:exon4:c.C98G:p.P33R,TP53:NM_001276696:exon4:c.C98G:p.P33R,TP53:NM_001276760:exon4:c.C98G:p.P33R,TP53:NM_001276761:exon4:c.C98G:p.P33R	77	0.99
	chr4	105275672	105275672	T	G	TET2	Non-synonymous SNV	TET2:NM_001127208:exon11:c.T5162G:p.L1721W	11	0.73
	chr1	16138117	16138117	G	T	EPHA2	Non-synonymous SNV	EPHA2:NM_004431:exon5:c.C1048A:p.P350T	46	0.67
	chr9	132595027	132595027	T	C	DDX31	Non-synonymous SNV	DDX31:NM_022779:exon20:c.A2395G:p.I799V	38	0.61

chr5	35860966	35860966	T	C	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon2:c.T197C:p.I66T	10	0.60
chr1	81952012	81952012	A	G	<i>ADGRL2</i>	Non-synonymous SNV	ADGRL2:NM_001297705:exon8:c.A1652G:p.K551R,ADGRL2:NM_001297706:exon8:c.A1652G:p.K551R,ADGRL2:NM_012302:exon8:c.A1652G:p.K551R,ADGRL2:NM_001297704:exon9:c.A1652G:p.K551R	44	0.59
chr3	66380763	66380763	C	T	<i>LRIG1</i>	Non-synonymous SNV	LRIG1:NM_015541:exon18:c.G2869A:p.A957T	16	0.56
chr12	6667904	6667906	TGC	-	<i>ZNF384</i>	Non-frameshift deletion	ZNF384:NM_001039920:exon9:c.1194_1196del:p.398_399del,ZNF384:NM_133476:exon10:c.1359_1361del:p.453_454del,ZNF384:NM_001135734:exon11:c.1542_1544del:p.514_515del	20	0.55
chr19	15179425	15179425	G	T	<i>NOTCH3</i>	Non-synonymous SNV	NOTCH3:NM_000435:exon21:c.C3399A:p.H1133Q	144	0.50
chr10	63168077	63168077	C	T	<i>JMJD1C</i>	Non-synonymous SNV	JMJD1C:NM_001282948:exon25:c.G7045A:p.E2349K,JMJD1C:NM_032776:exon26:c.G7591A:p.E2531K	109	0.49
chr8	127738294	127738294	A	G	<i>MYC</i>	Non-synonymous SNV	MYC:NM_002467:exon2:c.A77G:p.N26S	182	0.48
chr19	1627366	1627366	A	G	<i>TCF3</i>	Non-synonymous SNV	TCF3:NM_001136139:exon5:c.T359C:p.L120P,TCF3:NM_003200:exon6:c.T359C:p.L120P	47	0.47
chr4	54264951	54264951	C	T	<i>PDGFRA</i>	Non-synonymous SNV	PDGFRA:NM_006206:exon5:c.C661T:p.L221F	129	0.46

chr3	66380453	66380453	G	C	<i>LRIG1</i>	Non-synonymous SNV	LRIG1:NM_015541:exon19:c.C3092G:p.P1031R	22	0.45
chr8	105802428	105802428	G	C	<i>ZFPM2</i>	Non-synonymous SNV	ZFPM2:NM_012082:exon8:c.G2346C:p.E782D	33	0.45
chr19	41352971	41352971	C	G	<i>TGFB1</i>	Non-synonymous SNV	TGFB1:NM_000660:exon1:c.G74C:p.R25P	102	0.44
chr13	28311623	28311623	G	A	<i>FLT1</i>	Non-synonymous SNV	FLT1:NM_002019:exon27:c.C3602T:p.P1201L	29	0.38
chr15	41762197	41762197	-	A	<i>MGA</i>	frameshift insertion	MGA:NM_001080541:exon21:c.6953dupA:p.Q2318fs, MGA:NM_001164273:exon22:c.7580dupA:p.Q2527fs	20	0.30
chr17	16146494	16146494	-	T	<i>NCOR1</i>	frameshift insertion	NCOR1:NM_001190438:exon7:c.636dupA:p.V213fs, NCOR1:NM_001190440:exon9:c.963dupA:p.V322fs, NCOR1:NM_006311:exon10:c.963dupA:p.V322fs	22	0.27
chr9	37357252	37357252	-	A	<i>ZCCHC7</i>	frameshift insertion	ZCCHC7:NM_001289119:exon9:c.1617dupA:p.R539fs, ZCCHC7:NM_001289120:exon9:c.1617dupA:p.R539fs, ZCCHC7:NM_001289121:exon9:c.1617dupA:p.R539fs, ZCCHC7:NM_032226:exon9:c.1617dupA:p.R539fs	50	0.24
chr12	6601979	6601981	TCC	-	<i>CHD4</i>	Non-frameshift deletion	CHD4:NM_001297553:exon3:c.396_398del:p.132_133del, CHD4:NM_001273:exon4:c.417_419del:p.139_140del	46	0.22
chr19	10160049	10160049	-	T	<i>DNMT1</i>	frameshift insertion	DNMT1:NM_001379:exon14:c.1009dupA:p.M337fs, DNMT1:NM_001130823:exon15:c.1057dupA:p.M353fs	19	0.21

	chr7	152146689	152146689	-	T	<i>KMT2C</i>	frameshift insertion	KMT2C:NM_170606:exon53:c.13940dupA:p.K4647fs	29	0.21
	chr22	20092882	20092882	-	A	<i>DGCR8</i>	frameshift insertion	DGCR8:NM_022720:exon8:c.1681dupA:p.S560fs	54	0.20
	chr1	6197724	6197724	-	T	<i>RPL22</i>	frameshift insertion	RPL22:NM_000983:exon2:c.44dupA:p.K15fs	3085	0.20
	chr4	24548878	24548878	-	T	<i>DHX15</i>	frameshift insertion	DHX15:NM_001358:exon6:c.1224dupA:p.Q409fs	118	0.19
	chr2	197416764	197416764	-	T	<i>SF3B1</i>	frameshift insertion	SF3B1:NM_012433:exon6:c.642dupA:p.L215fs	89	0.12

Table S2.12: All Cancer Associated SNVs Called for Case 5. Variants were filtered based on a list of 383 pre-defined cancer genes. The variants were maintained if they were exonic, had a damaging function, previously reported in COSMIC, and were considered pathogenic in the ClinVar Database. Variants were then excluded if they had a variant allele frequency below 0.05, depth below 10, a prevalence of greater than 10% in the 1000 Genomes Project, or if no annotations were present. SNVs are shaded dark orange if the variant allele frequency (VAF) =1, and lighter shades of orange for VAFs>=.5 and<1 and VAFs <.5 and>.4. The 'AA Change' column lists the ref.seq. number, the base pair alteration, and the amino acid alteration.

	Chromosome	Start	End	Reference	Alteration	Gene	Exonic Function	AA Change	Read Depth	VAF
	Case 6	chr17	7676154	7676154	G	C	<i>TP53</i>	Non-synonymous SNV	TP53:NM_001126118:exon3:c.C98G:p.P33R,TP53:NM_000546:exon4:c.C215G:p.P72R,TP53:NM_001126112:exon4:c.C215G:p.P72R,TP53:NM_001126113:exon4:c.C215G:p.P72R,TP53:NM_001126114:exon4:c.C215G:p.P72R,TP53:NM_001276695:exon4:c.C98G:p.P33R,TP53:NM_001276696:exon4:c.C98G:p.P33R,TP53:NM_001276760:exon4:c.C98G:p.P33R,TP53:NM_001276761:exon4:c.C98G:p.P33R	68
chr5		68292320	68292320	G	A	<i>PIK3R1</i>	Non-synonymous SNV	PIK3R1:NM_181504:exon2:c.G168A:p.M56I,PIK3R1:NM_181524:exon2:c.G78A:p.M26I,PIK3R1:NM_181523:exon8:c.G978A:p.M326I	45	1.00
chr8		105802747	105802747	C	G	<i>ZFPM2</i>	Non-synonymous SNV	ZFPM2:NM_012082:exon8:c.C2665G:p.Q889E	20	1.00
chr9		132595027	132595027	T	C	<i>DDX31</i>	Non-synonymous SNV	DDX31:NM_022779:exon20:c.A2395G:p.I799V	26	1.00
chrX		77682471	77682471	C	G	<i>ATRX</i>	Non-synonymous SNV	ATRX:NM_138270:exon8:c.G2671C:p.E891Q,ATRX:NM_000489:exon9:c.G2785C:p.E929Q	19	1.00
chr5		35874473	35874473	C	T	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon6:c.C731T:p.T244I	18	0.83
chr9		136506781	136506781	C	T	<i>NOTCH1</i>	Non-synonymous SNV	NOTCH1:NM_017617:exon23:c.G3836A:p.R1279H	16	0.75

chr7	55161562	55161562	G	A	<i>EGFR</i>	Non-synonymous SNV	EGFR:NM_005228:exon13:c.G1562A:p.R521K,EGFR:NM_201282:exon13:c.G1562A:p.R521K,EGFR:NM_201284:exon13:c.G1562A:p.R521K	11	0.73
chr3	47121396	47121396	C	T	<i>SETD2</i>	Non-synonymous SNV	SETD2:NM_014159:exon3:c.G3240A:p.M1080I	13	0.62
chr5	180621758	180621758	G	A	<i>FLT4</i>	Non-synonymous SNV	FLT4:NM_002020:exon13:c.C1804T:p.H602Y,FLT4:NM_182925:exon13:c.C1804T:p.H602Y	15	0.60
chr14	34761774	34761774	C	T	<i>BAZ1A</i>	Non-synonymous SNV	BAZ1A:NM_182648:exon23:c.G4130A:p.R1377K,BAZ1A:NM_013448:exon24:c.G4226A:p.R1409K	78	0.59
chr3	66380453	66380453	G	C	<i>LRIG1</i>	Non-synonymous SNV	LRIG1:NM_015541:exon19:c.C3092G:p.P1031R	58	0.57
chr14	34761902	34761902	A	T	<i>BAZ1A</i>	Non-synonymous SNV	BAZ1A:NM_182648:exon23:c.T4002A:p.N1334K,BAZ1A:NM_013448:exon24:c.T4098A:p.N1366K	93	0.56
chr3	47083895	47083895	G	A	<i>SETD2</i>	Non-synonymous SNV	SETD2:NM_014159:exon12:c.C5885T:p.P1962L	45	0.56
chr3	66380763	66380763	C	T	<i>LRIG1</i>	Non-synonymous SNV	LRIG1:NM_015541:exon18:c.G2869A:p.A957T	30	0.50
chr9	21816759	21816759	G	A	<i>MTAP</i>	Non-synonymous SNV	MTAP:NM_002451:exon3:c.G166A:p.V56I	25	0.48
chr7	22731420	22731420	T	A	<i>IL6</i>	Non-synonymous SNV	IL6:NM_000600:exon5:c.T486A:p.D162E	103	0.46
chr4	108048727	108048727	G	A	<i>LEF1</i>	Non-synonymous SNV	LEF1:NM_001130714:exon10:c.C1106T:p.T369M	144	0.44

chr9	8485834	8485834	G	A	<i>PTPRD</i>	Non-synonymous SNV	PTPRD:NM_002839:exon28:c.C2983T:p.R995C	18	0.44
chr5	35860966	35860966	T	C	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon2:c.T197C:p.I66T	80	0.44
chr9	37126781	37126781	G	A	<i>ZCCHC7</i>	Non-synonymous SNV	ZCCHC7:NM_001289119:exon2:c.G449A:p.R150Q,ZCCHC7:NM_001289120:exon2:c.G449A:p.R150Q,ZCCHC7:NM_001289121:exon2:c.G449A:p.R150Q,ZCCHC7:NM_032226:exon2:c.G449A:p.R150Q	36	0.42
chr5	35871088	35871088	G	A	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon4:c.G412A:p.V138I	66	0.41
chr1	16137994	16137994	C	T	<i>EPHA2</i>	Non-synonymous SNV	EPHA2:NM_004431:exon5:c.G1171A:p.G391R	20	0.40
chr8	127738294	127738294	A	G	<i>MYC</i>	Non-synonymous SNV	MYC:NM_002467:exon2:c.A77G:p.N26S	49	0.39
chr12	6667904	6667915	TGCTGCT GCTGC	-	<i>ZNF384</i>	Non-frameshift deletion	ZNF384:NM_001039920:exon9:c.1185_1196del:p.395_399del,ZNF384:NM_133476:exon10:c.1350_1361del:p.450_454del,ZNF384:NM_001135734:exon11:c.1533_1544del:p.511_515del	20	0.35
chr1	6197724	6197724	-	T	<i>RPL22</i>	frameshift insertion	RPL22:NM_000983:exon2:c.44dupA:p.K15fs	2067	0.23
chrX	71122774	71122776	TGC	-	<i>MED12</i>	Non-frameshift deletion	MED12:NM_005120:exon10:c.1385_1387del:p.462_463del	10	0.20

chr17	16146494	16146494	-	T	<i>NCOR1</i>	frameshift insertion	NCOR1:NM_001190438:exon7:c.636dupA:p.V213fs,NCOR1:NM_001190440:exon9:c.963dupA:p.V322fs,NCOR1:NM_006311:exon10:c.963dupA:p.V322fs	39	0.18
chr9	37357252	37357252	-	A	<i>ZCCHC7</i>	frameshift insertion	ZCCHC7:NM_001289119:exon9:c.1617dupA:p.R539fs,ZCCHC7:NM_001289120:exon9:c.1617dupA:p.R539fs,ZCCHC7:NM_001289121:exon9:c.1617dupA:p.R539fs,ZCCHC7:NM_032226:exon9:c.1617dupA:p.R539fs	55	0.16
chr22	22893804	22893804	A	G	<i>IGLL5</i>	Non-synonymous SNV	IGLL5:NM_001178126:exon2:c.A311G:p.K104R	40	0.15
chr1	64873427	64873427	-	T	<i>JAK1</i>	frameshift insertion	JAK1:NM_002227:exon5:c.425dupA:p.K142fs	97	0.13
chr12	6601979	6601981	TCC	-	<i>CHD4</i>	Non-frameshift deletion	CHD4:NM_001297553:exon3:c.396_398del:p.132_133del,CHD4:NM_001273:exon4:c.417_419del:p.139_140del	52	0.12

Table S2.13: All Cancer Associated SNVs Called for Case 6. Variants were filtered based on a list of 383 pre-defined cancer genes. The variants were maintained if they were exonic, had a damaging function, previously reported in COSMIC, and were considered pathogenic in the ClinVar Database. Variants were then excluded if they had a variant allele frequency below 0.05, depth below 10, a prevalence of greater than 10% in the 1000 Genomes Project, or if no annotations were present. SNVs are shaded dark orange if the variant allele frequency (VAF) =1, and lighter shades of orange for VAFs >=0.5 and <1 and VAFs <0.5 and >0.4. The 'AA Change' column lists the ref.seq. number, the base pair alteration, and the amino acid alteration.

	Chromosome	Start	End	Reference	Alteration	Gene	Exonic Function	AA Change	Read Depth	VAF
	Case 7	chr17	7676154	7676154	G	C	<i>TP53</i>	Non-synonymous SNV	TP53:NM_001126118:exon3:c.C98G:p.P33R,TP53:NM_000546:exon4:c.C215G:p.P72R,TP53:NM_001126112:exon4:c.C215G:p.P72R,TP53:NM_001126113:exon4:c.C215G:p.P72R,TP53:NM_001126114:exon4:c.C215G:p.P72R,TP53:NM_001276695:exon4:c.C98G:p.P33R,TP53:NM_001276696:exon4:c.C98G:p.P33R,TP53:NM_001276760:exon4:c.C98G:p.P33R,TP53:NM_001276761:exon4:c.C98G:p.P33R	35
chr9		132911517	132911517	A	G	<i>TSC1</i>	Non-synonymous SNV	TSC1:NM_001162427:exon9:c.T812C:p.M271T,TSC1:NM_000368:exon10:c.T965C:p.M322T,TSC1:NM_001162426:exon10:c.T965C:p.M322T	22	0.64
chr9		21816759	21816759	G	A	<i>MTAP</i>	Non-synonymous SNV	MTAP:NM_002451:exon3:c.G166A:p.V56I	25	0.60
chr5		38921762	38921762	G	A	<i>OSMR</i>	Non-synonymous SNV	OSMR:NM_003999:exon12:c.G1733A:p.G578D	78	0.56
chr16		3781229	3781229	G	T	<i>CREBBP</i>	Non-synonymous SNV	CREBBP:NM_001079846:exon6:c.C1537A:p.L513I,CREBBP:NM_004380:exon7:c.C1651A:p.L551I	10	0.50
chr9		132595027	132595027	T	C	<i>DDX31</i>	Non-synonymous SNV	DDX31:NM_022779:exon20:c.A2395G:p.I799V	36	0.50

chr5	38884097	38884097	T	C	<i>OSMR</i>	Non-synonymous SNV	OSMR:NM_001168355:exon5:c.T689C:p.V230A,OSMR:NM_003999:exon5:c.T689C:p.V230A	75	0.43
chr2	112832730	112832730	-	T	<i>IL1B</i>	frameshift insertion	IL1B:NM_000576:exon5:c.397dupA:p.S133fs	10	0.40
chr1	16130268	16130268	C	T	<i>EPHA2</i>	Non-synonymous SNV	EPHA2:NM_004431:exon15:c.G2627A:p.R876H	28	0.39
chr5	35860966	35860966	T	C	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon2:c.T197C:p.I66T	28	0.39
chr5	35871088	35871088	G	A	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon4:c.G412A:p.V138I	23	0.39
chr12	6601979	6601981	TCC	-	<i>CHD4</i>	Non-frameshift deletion	CHD4:NM_001297553:exon3:c.396_398del:p.132_133del,CHD4:NM_001273:exon4:c.417_419del:p.139_140del	29	0.28
chr17	16146494	16146494	-	T	<i>NCOR1</i>	frameshift insertion	NCOR1:NM_001190438:exon7:c.636dupA:p.V213fs,NCOR1:NM_001190440:exon9:c.963dupA:p.V322fs,NCOR1:NM_006311:exon10:c.963dupA:p.V322fs	12	0.25
chr20	1915303	1915303	-	GT	<i>SIRPA</i>	frameshift insertion	SIRPA:NM_001040023:exon2:c.284_285insGT:p.D95fs,SIRPA:NM_001040022:exon3:c.284_285insGT:p.D95fs,SIRPA:NM_080792:exon3:c.284_285insGT:p.D95fs	33	0.24
chr20	1915305	1915306	CT	-	<i>SIRPA</i>	frameshift deletion	SIRPA:NM_001040023:exon2:c.286_287del:p.L96fs,SIRPA:NM_001040022:exon3:c.286_287del:p.L96fs,SIRPA:NM_080792:exon3:c.286_287del:p.L96fs	33	0.24

chr5	35867438	35867438	-	A	<i>IL7R</i>	frameshift insertion	IL7R:NM_002185:exon3:c.355dupA:p.C118fs	26	0.23
chr1	6197724	6197724	-	T	<i>RPL22</i>	frameshift insertion	RPL22:NM_000983:exon2:c.44dupA:p.K15fs	2470	0.22
chr12	6667904	6667906	TGC	-	<i>ZNF384</i>	Non-frameshift deletion	ZNF384:NM_001039920:exon9:c.1194_1196del:p.398_399del,ZNF384:NM_133476:exon10:c.1359_1361del:p.453_454del,ZNF384:NM_001135734:exon11:c.1542_1544del:p.514_515del	14	0.21
chr13	28389939	28389939	-	T	<i>FLT1</i>	frameshift insertion	FLT1:NM_001159920:exon13:c.1825dupA:p.M609fs,FLT1:NM_001160030:exon13:c.1825dupA:p.M609fs,FLT1:NM_002019:exon13:c.1825dupA:p.M609fs	42	0.17
chr20	1915406	1915408	CGA	-	<i>SIRPA</i>	Non-frameshift deletion	SIRPA:NM_001040023:exon2:c.387_389del:p.129_130del,SIRPA:NM_001040022:exon3:c.387_389del:p.129_130del,SIRPA:NM_080792:exon3:c.387_389del:p.129_130del	54	0.17
chr20	1915319	1915319	C	A	<i>SIRPA</i>	Non-synonymous SNV	SIRPA:NM_001040023:exon2:c.C300A:p.N100K,SIRPA:NM_001040022:exon3:c.C300A:p.N100K,SIRPA:NM_080792:exon3:c.C300A:p.N100K	37	0.16

Table S2.14: All Cancer Associated SNVs Called for Case 7. Variants were filtered based on a list of 383 pre-defined cancer genes. The variants were maintained if they were exonic, had a damaging function, previously reported in COSMIC, and were considered pathogenic in the ClinVar Database. Variants were then excluded if they had a variant allele frequency below 0.05, depth below 10, a prevalence of greater than 10% in the 1000 Genomes Project, or if no annotations were present. SNVs are shaded dark orange if the variant allele frequency (VAF) =1, and lighter shades of orange for VAFs >=0.5 and <1 and VAFs <0.5 and >0.4. The 'AA Change' column lists the ref.seq. number, the base pair alteration, and the amino acid alteration.

	Chromosome	Start	End	Reference	Alteration	Gene	Exonic Function	AA Change	Read Depth	VAF
Case 9	chr9	132595027	132595027	T	C	<i>DDX31</i>	Non-synonymous SNV	DDX31:NM_022779:exon20:c.A2395G:p.I799V	22	1.00
	chr4	55106807	55106807	T	A	<i>KDR</i>	Non-synonymous SNV	KDR:NM_002253:exon11:c.A1416T:p.Q472H	10	0.80
	chr5	35860966	35860966	T	C	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon2:c.T197C:p.I66T	25	0.64
	chr5	38883969	38883969	T	G	<i>OSMR</i>	Non-synonymous SNV	OSMR:NM_001168355:exon5:c.T561G:p.H187Q,OSMR:NM_003999:exon5:c.T561G:p.H187Q	75	0.56
	chr19	1627366	1627366	A	G	<i>TCF3</i>	Non-synonymous SNV	TCF3:NM_001136139:exon5:c.T359C:p.L120P,TCF3:NM_003200:exon6:c.T359C:p.L120P	38	0.55
	chr5	35871088	35871088	G	A	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon4:c.G412A:p.V138I	41	0.54
	chr10	63186253	63186253	T	C	<i>JMJD1C</i>	Non-synonymous SNV	JMJD1C:NM_001282948:exon18:c.A6155G:p.N2052S,JMJD1C:NM_032776:exon19:c.A6701G:p.N2234S	59	0.53
	chr12	103938446	103938446	C	T	<i>HSP90B1</i>	Non-synonymous SNV	HSP90B1:NM_003299:exon7:c.C962T:p.P321L	969	0.50
	chr22	22888120	22888120	C	T	<i>IGLL5</i>	Non-synonymous SNV	IGLL5:NM_001178126:exon1:c.C67T:p.R23C	102	0.49
	chr19	1627423	1627423	T	C	<i>TCF3</i>	Non-synonymous SNV	TCF3:NM_001136139:exon5:c.A302G:p.K101R,TCF3:NM_003200:exon6:c.A302G:p.K101R	32	0.47
	chr5	180624003	180624003	T	C	<i>FLT4</i>	Non-synonymous	FLT4:NM_002020:exon11:c.A1480G:p.T494A,FLT4:N	15	0.47

						SNV	M_182925:exon11:c.A1480G:p.T494A		
chr9	136506781	136506781	C	T	<i>NOTCH1</i>	Non-synonymous SNV	NOTCH1:NM_017617:exon23:c.G3836A:p.R1279H	13	0.46
chr17	16057554	16057554	T	C	<i>NCOR1</i>	Non-synonymous SNV	NCOR1:NM_001190440:exon39:c.A6043G:p.R2015G, NCOR1:NM_006311:exon40:c.A6352G:p.R2118G	22	0.45
chr1	8014274	8014274	C	T	<i>ERFFI1</i>	Non-synonymous SNV	ERFFI1:NM_018948:exon4:c.G325A:p.D109N	242	0.40
chr9	132911517	132911517	A	G	<i>TSC1</i>	Non-synonymous SNV	TSC1:NM_001162427:exon9:c.T812C:p.M271T, TSC1:NM_000368:exon10:c.T965C:p.M322T, TSC1:NM_001162426:exon10:c.T965C:p.M322T	10	0.40
chr9	21816759	21816759	G	A	<i>MTAP</i>	Non-synonymous SNV	MTAP:NM_002451:exon3:c.G166A:p.V56I	23	0.39
chr2	208243593	208243593	C	T	<i>IDH1</i>	Non-synonymous SNV	IDH1:NM_001282386:exon6:c.G532A:p.V178I, IDH1:NM_001282387:exon6:c.G532A:p.V178I, IDH1:NM_005896:exon6:c.G532A:p.V178I	52	0.38
chr22	23314009	23314009	C	G	<i>BCR</i>	Non-synonymous SNV	BCR:NM_021574:exon20:c.C3367G:p.L1123V, BCR:NM_004327:exon21:c.C3499G:p.L1167V	118	0.36

chr17	7676154	7676154	G	C	<i>TP53</i>	Non-synonymous SNV	TP53:NM_001126118:exon3:c.C98G:p.P33R,TP53:NM_000546:exon4:c.C215G:p.P72R,TP53:NM_001126112:exon4:c.C215G:p.P72R,TP53:NM_001126113:exon4:c.C215G:p.P72R,TP53:NM_001126114:exon4:c.C215G:p.P72R,TP53:NM_001276695:exon4:c.C98G:p.P33R,TP53:NM_001276696:exon4:c.C98G:p.P33R,TP53:NM_001276760:exon4:c.C98G:p.P33R,TP53:NM_001276761:exon4:c.C98G:p.P33R	21	0.33
chr1	6197724	6197724	-	T	<i>RPL22</i>	frameshift insertion	RPL22:NM_000983:exon2:c.44dupA:p.K15fs	2632	0.22
chr9	37357252	37357252	-	A	<i>ZCCHC7</i>	frameshift insertion	ZCCHC7:NM_001289119:exon9:c.1617dupA:p.R539fs,ZCCHC7:NM_001289120:exon9:c.1617dupA:p.R539fs,ZCCHC7:NM_001289121:exon9:c.1617dupA:p.R539fs,ZCCHC7:NM_032226:exon9:c.1617dupA:p.R539fs	28	0.18
chr3	30650379	30650379	-	A	<i>TGFBR2</i>	frameshift insertion	TGFBR2:NM_003242:exon3:c.374dupA:p.E125fs,TGFBR2:NM_001024847:exon4:c.449dupA:p.E150fs	70	0.17
chr3	49360949	49360949	T	C	<i>RHOA</i>	Non-synonymous SNV	RHOA:NM_001313943:exon5:c.A526G:p.R176G	18	0.17
chr9	21818093	21818093	G	T	<i>MTAP</i>	Non-synonymous SNV	MTAP:NM_002451:exon4:c.G238T:p.A80S	24	0.13

Table S2.15: All Cancer Associated SNVs Called for Case 9. Variants were filtered based on a list of 383 pre-defined cancer genes. The variants were maintained if they were exonic, had a damaging function, previously reported in COSMIC, and were considered pathogenic in the ClinVar Database. Variants were then excluded if they had a variant allele frequency below 0.05, depth below 10, a prevalence of greater than 10% in the 1000 Genomes Project, or if no annotations were present. SNVs are shaded dark orange if the variant allele frequency (VAF) =1, and lighter shades of orange for VAFs >=0.5 and <1 and VAFs <0.5 and >0.4. The 'AA Change' column lists the ref.seq. number, the base pair alteration, and the amino acid alteration.

	Chromosome	Start	End	Reference	Alteration	Gene	Exonic Function	AA Change	Read Depth	VAF
Case 10	chr17	7676154	7676154	G	C	<i>TP53</i>	Non-synonymous SNV	TP53:NM_001126118:exon3:c.C98G:p.P33R,TP53:NM_000546:exon4:c.C215G:p.P72R,TP53:NM_001126112:exon4:c.C215G:p.P72R,TP53:NM_001126113:exon4:c.C215G:p.P72R,TP53:NM_001126114:exon4:c.C215G:p.P72R,TP53:NM_001276695:exon4:c.C98G:p.P33R,TP53:NM_001276696:exon4:c.C98G:p.P33R,TP53:NM_001276760:exon4:c.C98G:p.P33R,TP53:NM_001276761:exon4:c.C98G:p.P33R	20	1.00
	chr2	29193500	29193500	G	C	<i>ALK</i>	Non-synonymous SNV	ALK:NM_004304:exon29:c.C4587G:p.D1529E	56	1.00
	chr3	66381544	66381544	G	A	<i>LRIG1</i>	Non-synonymous SNV	LRIG1:NM_015541:exon17:c.C2705T:p.A902V	104	1.00
	chr9	132595027	132595027	T	C	<i>DDX31</i>	Non-synonymous SNV	DDX31:NM_022779:exon20:c.A2395G:p.I799V	15	1.00
	chrX	77682471	77682471	C	G	<i>ATRX</i>	Non-synonymous SNV	ATRX:NM_138270:exon8:c.G2671C:p.E891Q,ATRX:NM_000489:exon9:c.G2785C:p.E929Q	10	1.00
	chrX	154400814	154400814	G	A	<i>RPL10</i>	Non-synonymous SNV	RPL10:NM_001256577:exon6:c.G442A:p.V148I,RPL10:NM_001256580:exon6:c.G497A:p.S166N,RPL10:NM_001303624:exon6:c.G605A:p.S202N,RPL10:NM_001303625:exon7:c.G605A:	3084	1.00

							p.S202N,RPL10:NM_006013:exon7:c.G605A:p.S202N		
chr2	29193615	29193615	T	C	ALK	Non-synonymous SNV	ALK:NM_004304:exon29:c.A4472G:p.K1491R	40	0.98
chr12	6667904	6667906	TGC	-	ZNF384	Non-frameshift deletion	ZNF384:NM_001039920:exon9:c.1194_1196del:p.398_399del,ZNF384:NM_133476:exon10:c.1359_1361del:p.453_454del,ZNF384:NM_001135734:exon11:c.1542_1544del:p.514_515del	13	0.92
chr5	35874473	35874473	C	T	IL7R	Non-synonymous SNV	IL7R:NM_002185:exon6:c.C731T:p.T244I	14	0.79
chr1	120069346	120069346	C	T	NOTCH2	Non-synonymous SNV	NOTCH2:NM_001200001:exon1:c.G61A:p.A21T,NOTCH2:NM_024408:exon1:c.G61A:p.A21T	27	0.70
chr22	41177749	41177749	C	T	EP300	Non-synonymous SNV	EP300:NM_001429:exon31:c.C6038T:p.P2013L	20	0.55
chr4	849720	849720	C	T	GAK	Non-synonymous SNV	GAK:NM_001286833:exon25:c.G3595A:p.D1199N,GAK:NM_005255:exon28:c.G3889A:p.D1297N	153	0.54
chr4	108048727	108048727	G	A	LEF1	Non-synonymous SNV	LEF1:NM_001130714:exon10:c.C1106T:p.T369M	152	0.49
chr5	35860966	35860966	T	C	IL7R	Non-synonymous SNV	IL7R:NM_002185:exon2:c.T197C:p.I66T	43	0.49
chr5	150135834	150135834	T	A	PDGFRB	Non-synonymous	PDGFRB:NM_002609:exon3:c.A85T:p.I29F	83	0.48

						SNV			
chr3	47083895	47083895	G	A	<i>SETD2</i>	Non-synonymous SNV	SETD2:NM_014159:exon12:c.C5885T:p.P1962L	23	0.43
chr20	32071703	32071703	A	G	<i>HCK</i>	Non-synonymous SNV	HCK:NM_001172129:exon2:c.A41G:p.N14S,HCK:NM_001172130:exon2:c.A104G:p.N35S,HCK:NM_001172131:exon2:c.A41G:p.N14S,HCK:NM_002110:exon2:c.A104G:p.N35S,HCK:NM_001172132:exon3:c.A44G:p.N15S,HCK:NM_001172133:exon3:c.A41G:p.N14S	28	0.43
chr5	150077330	150077330	C	T	<i>CSF1R</i>	Non-synonymous SNV	CSF1R:NM_001288705:exon5:c.G835A:p.V279M,CSF1R:NM_005211:exon6:c.G835A:p.V279M	33	0.42
chr5	35871088	35871088	G	A	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon4:c.G412A:p.V138I	43	0.35
chr1	6197724	6197724	-	T	<i>RPL22</i>	frameshift insertion	RPL22:NM_000983:exon2:c.44dupA:p.K15fs	2278	0.22
chr9	37357252	37357252	-	A	<i>ZCCHC7</i>	frameshift insertion	ZCCHC7:NM_001289119:exon9:c.1617dupA:p.R539fs,ZCCHC7:NM_001289120:exon9:c.1617dupA:p.R539fs,ZCCHC7:NM_001289121:exon9:c.1617dupA:p.R539fs,ZCCHC7:NM_032226:exon9:c.1617dupA:p.R539fs	53	0.17
chr12	14478321	14478321	-	A	<i>ATF7IP</i>	frameshift insertion	ATF7IP:NM_001286514:exon12:c.2944dupA:p.P981fs,ATF7IP:NM_001286515:exon12:c.2944dupA:p.P981fs,ATF7IP:NM_018179:exon12:c.2947dupA:p.P982fs,ATF7IP:NM_181352:exon12:c.2971dupA:p.P990fs	46	0.15

	chr4	24548878	24548878	-	T	<i>DHX15</i>	frameshift insertion	DHX15:NM_001358:exon6: c.1224dupA:p.Q409fs	96	0.13
--	------	----------	----------	---	---	--------------	-------------------------	-----------------------------------------------	----	------

Table S2.16: All Cancer Associated SNVs Called for Case 10. Variants were filtered based on a list of 383 pre-defined cancer genes. The variants were maintained if they were exonic, had a damaging function, previously reported in COSMIC, and were considered pathogenic in the ClinVar Database. Variants were then excluded if they had a variant allele frequency below 0.05, depth below 10, a prevalence of greater than 10% in the 1000 Genomes Project, or if no annotations were present. SNVs are shaded dark orange if the variant allele frequency (VAF) =1, and lighter shades of orange for VAFs >=.5 and <1 and VAFs <.5 and >.4. The 'AA Change' column lists the ref.seq. number, the base pair alteration, and the amino acid alteration.

Case 11	Chromosome	Start	End	Reference	Alteration	Gene	Exonic Function	AA Change	Read Depth	VAF
	chr12	6667904	6667906	TGC	-	<i>ZNF384</i>	Non-frameshift deletion	ZNF384:NM_001039920:exon9:c.1194_1196del:p.398_399del,ZNF384:NM_133476:exon10:c.1359_1361del:p.453_454del,ZNF384:NM_001135734:exon11:c.1542_1544del:p.514_515del	15	1.00
	chr17	7676154	7676154	G	C	<i>TP53</i>	Non-synonymous SNV	TP53:NM_001126118:exon3:c.C98G:p.P33R,TP53:NM_000546:exon4:c.C215G:p.P72R,TP53:NM_001126112:exon4:c.C215G:p.P72R,TP53:NM_001126113:exon4:c.C215G:p.P72R,TP53:NM_001126114:exon4:c.C215G:p.P72R,TP53:NM_001276695:exon4:c.C98G:p.P33R,TP53:NM_001276696:exon4:c.C98G:p.P33R,TP53:NM_001276760:exon4:c.C98G:p.P33R,TP53:NM_001276761:exon4:c.C98G:p.P33R	97	1.00
	chr2	29193500	29193500	G	C	<i>ALK</i>	Non-synonymous SNV	ALK:NM_004304:exon29:c.C4587G:p.D1529E	196	1.00
	chr5	35860966	35860966	T	C	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon2:c.T197C:p.I66T	19	1.00
	chr5	35871088	35871088	G	A	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon4:c.G412A:p.V138I	23	1.00
chr8	105802428	105802428	G	C	<i>ZFPM2</i>	Non-synonymous SNV	ZFPM2:NM_012082:exon8:c.G2346C:p.E782D	11	1.00	

chr9	132595027	132595027	T	C	<i>DDX31</i>	Non-synonymous SNV	DDX31:NM_022779:exon20:c.A2395G:p.I799V	24	1.00
chrX	77682471	77682471	C	G	<i>ATRX</i>	Non-synonymous SNV	ATRX:NM_138270:exon8:c.G2671C:p.E891Q,ATRX:NM_000489:exon9:c.G2785C:p.E929Q	23	1.00
chr7	148828812	148828812	C	G	<i>EZH2</i>	Non-synonymous SNV	EZH2:NM_152998:exon5:c.G436C:p.D146H,EZH2:NM_001203247:exon6:c.G553C:p.D185H,EZH2:NM_001203248:exon6:c.G526C:p.D176H,EZH2:NM_001203249:exon6:c.G526C:p.D176H,EZH2:NM_004456:exon6:c.G553C:p.D185H	24	0.75
chr8	32756465	32756465	T	C	<i>NRG1</i>	Non-synonymous SNV	NRG1:NM_001159996:exon5:c.T404C:p.M135T,NRG1:NM_001160004:exon9:c.T857C:p.M286T,NRG1:NM_001160008:exon9:c.T857C:p.M286T,NRG1:NM_013957:exon9:c.T857C:p.M286T,NRG1:NM_013960:exon9:c.T866C:p.M289T,NRG1:NM_013964:exon9:c.T866C:p.M289T,NRG1:NM_013956:exon10:c.T881C:p.M294T	12	0.75
chr11	108304735	108304735	G	A	<i>ATM</i>	Non-synonymous SNV	ATM:NM_000051:exon37:c.G5557A:p.D1853N	26	0.62
chr5	180618911	180618911	G	A	<i>FLT4</i>	Non-synonymous SNV	FLT4:NM_002020:exon21:c.C2860T:p.P954S,FLT4:NM_182925:exon21:c.C2860T:p.P954S	49	0.59

chr17	43071077	43071077	T	C	<i>BRCA1</i>	Non-synonymous SNV	BRCA1:NM_007297:exon14:c.A4696G:p.S1566G, BRCA1:NM_007298:exon14:c.A1525G:p.S509G, BRCA1:NM_007294:exon15:c.A4837G:p.S1613G, BRCA1:NM_007299:exon15:c.A1525G:p.S509G, BRCA1:NM_007300:exon16:c.A4900G:p.S1634G	11	0.55
chr3	47083895	47083895	G	A	<i>SETD2</i>	Non-synonymous SNV	SETD2:NM_014159:exon12:c.C5885T:p.P1962L	67	0.54
chr9	134048405	134048405	G	A	<i>BRD3</i>	Non-synonymous SNV	BRD3:NM_007371:exon6:c.C764T:p.A255V	19	0.53
chr17	43092418	43092418	T	C	<i>BRCA1</i>	Non-synonymous SNV	BRCA1:NM_007297:exon9:c.A2972G:p.E991G, BRCA1:NM_007294:exon10:c.A3113G:p.E1038G, BRCA1:NM_007300:exon10:c.A3113G:p.E1038G	10	0.50
chr9	21816759	21816759	G	A	<i>MTAP</i>	Non-synonymous SNV	MTAP:NM_002451:exon3:c.G166A:p.V56I	54	0.50
chr4	55106807	55106807	T	A	<i>KDR</i>	Non-synonymous SNV	KDR:NM_002253:exon11:c.A1416T:p.Q472H	92	0.49
chr9	130885205	130885205	C	T	<i>ABL1</i>	Non-synonymous SNV	ABL1:NM_005157:exon11:c.C2915T:p.S972L, ABL1:NM_007313:exon11:c.C2972T:p.S991L	68	0.49
chr1	36472277	36472277	C	T	<i>CSF3R</i>	Non-synonymous SNV	CSF3R:NM_000760:exon8:c.G958A:p.D320N, CSF3R:NM_156039:exon8:c.G958A:p.D320N, CSF3R:NM_172313:exon8:c.G958A:p.D320N	31	0.48

chr9	131144699	131144699	A	G	<i>NUP214</i>	Non-synonymous SNV	NUP214:NM_005085:exon12:c.A1714G:p.M572V	15	0.47
chr10	43114671	43114671	G	A	<i>RET</i>	Non-synonymous SNV	RET:NM_020630:exon11:c.G2071A:p.G691S,RET:NM_020975:exon11:c.G2071A:p.G691S	12	0.42
chr7	129212026	129212026	C	T	<i>SMO</i>	Non-synonymous SNV	SMO:NM_005631:exon12:c.C1939T:p.P647S	24	0.42
chr9	136505728	136505728	G	T	<i>NOTCH1</i>	Non-synonymous SNV	NOTCH1:NM_017617:exon25:c.C4168A:p.P1390T	37	0.41
chr3	49360990	49360990	T	C	<i>RHOA</i>	Non-synonymous SNV	RHOA:NM_001313943:exon5:c.A485G:p.K162R	14	0.36
chr12	49039800	49039800	G	A	<i>KMT2D</i>	Non-synonymous SNV	KMT2D:NM_003482:exon31:c.C7970T:p.A2657V	15	0.33
chr17	16146494	16146494	-	T	<i>NCOR1</i>	frameshift insertion	NCOR1:NM_001190438:exon7:c.636dupA:p.V213fs,NCOR1:NM_001190440:exon9:c.963dupA:p.V322fs,NCOR1:NM_006311:exon10:c.963dupA:p.V322fs	26	0.27
chr1	16945542	16945542	G	A	<i>CROCC</i>	Non-synonymous SNV	CROCC:NM_014675:exon15:c.G2072A:p.R691H	17	0.24
chr1	6197724	6197724	-	T	<i>RPL22</i>	frameshift insertion	RPL22:NM_000983:exon2:c.44dupA:p.K15fs	2919	0.22
chr9	37357252	37357252	-	A	<i>ZCCHC7</i>	frameshift insertion	ZCCHC7:NM_001289119:exon9:c.1617dupA:p.R539fs,ZCCHC7:NM_001289120:exon9:c.1617dupA:p.R539fs,ZCCHC7:NM_001289121:exon9:c.1617dupA:p.R539fs,ZCCHC7:NM_032226:exon9:c.1617dupA:p.R539fs	34	0.18

	chr11	3699158	3699158	-	T	<i>NUP98</i>	frameshift insertion	NUP98:NM_016320:exon 25:c.3932dupA:p.N1311fs, NUP98:NM_139132:exon 25:c.3932dupA:p.N1311fs	66	0.14
--	-------	---------	---------	---	---	--------------	-------------------------	-----------------------------------------------------------------------------------------------------	----	------

Table S2.17: All Cancer Associated SNVs Called for Case 11. Variants were filtered based on a list of 383 pre-defined cancer genes. The variants were maintained if they were exonic, had a damaging function, previously reported in COSMIC, and were considered pathogenic in the ClinVar Database. Variants were then excluded if they had a variant allele frequency below 0.05, depth below 10, a prevalence of greater than 10% in the 1000 Genomes Project, or if no annotations were present. SNVs are shaded dark orange if the variant allele frequency (VAF) =1, and lighter shades of orange for VAFs >=0.5 and <1 and VAFs <0.5 and >0.4. The 'AA Change' column lists the ref.seq. number, the base pair alteration, and the amino acid alteration.

Case 12	Chromosome	Start	End	Reference	Alteration	Gene	Exonic Function	AA Change	Read Depth	VAF
	chr12	6667904	6667906	TGC	-	<i>ZNF384</i>	Non-frameshift deletion	ZNF384:NM_001039920:exon9:c.1194_1196del:p.398_399del,ZNF384:NM_133476:exon10:c.1359_1361del:p.453_454del,ZNF384:NM_001135734:exon11:c.1542_1544del:p.514_515del	13	1.00
	chr17	7676154	7676154	G	C	<i>TP53</i>	Non-synonymous SNV	TP53:NM_001126118:exon3:c.C98G:p.P33R,TP53:NM_000546:exon4:c.C215G:p.P72R,TP53:NM_001126112:exon4:c.C215G:p.P72R,TP53:NM_001126113:exon4:c.C215G:p.P72R,TP53:NM_001126114:exon4:c.C215G:p.P72R,TP53:NM_001276695:exon4:c.C98G:p.P33R,TP53:NM_001276696:exon4:c.C98G:p.P33R,TP53:NM_001276760:exon4:c.C98G:p.P33R,TP53:NM_001276761:exon4:c.C98G:p.P33R	12	1.00
	chr5	68292320	68292320	G	A	<i>PIK3R1</i>	Non-synonymous SNV	PIK3R1:NM_181504:exon2:c.G168A:p.M56I,PIK3R1:NM_181524:exon2:c.G78A:p.M26I,PIK3R1:NM_181523:exon8:c.G978A:p.M326I	53	1.00
	chr9	132595027	132595027	T	C	<i>DDX31</i>	Non-synonymous SNV	DDX31:NM_022779:exon20:c.A2395G:p.I799V	14	1.00
	chr5	35860966	35860966	T	C	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon2:c.T197C:p.I66T	12	0.75

chr19	15185631	15185631	C	T	<i>NOTCH3</i>	Non-synonymous SNV	NOTCH3:NM_000435:exon13:c.G2000A:p.G667D	34	0.62
chr15	41578151	41578151	A	T	<i>TYRO3</i>	Non-synonymous SNV	TYRO3:NM_006293:exon19:c.A2548T:p.I850L	57	0.60
chr5	35871088	35871088	G	A	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon4:c.G412A:p.V138I	16	0.56
chr7	152162598	152162598	G	A	<i>KMT2C</i>	Non-synonymous SNV	KMT2C:NM_170606:exon43:c.C10979T:p.S3660L	13	0.54
chr4	55106807	55106807	T	A	<i>KDR</i>	Non-synonymous SNV	KDR:NM_002253:exon11:c.A1416T:p.Q472H	28	0.54
chr5	180603313	180603313	C	A	<i>FLT4</i>	Non-synonymous SNV	FLT4:NM_182925:exon30:c.G3971T:p.R1324L	19	0.53
chr3	47083895	47083895	G	A	<i>SETD2</i>	Non-synonymous SNV	SETD2:NM_014159:exon12:c.C5885T:p.P1962L	29	0.52
chr16	67611435	67611435	-	A	<i>CTCF</i>	frameshift insertion	CTCF:NM_006565:exon3:c.604dupA:p.A201fs	10	0.50
chr17	59213482	59213482	A	C	<i>SMG8</i>	Non-synonymous SNV	SMG8:NM_018149:exon3:c.A2659C:p.M887L	30	0.50
chr1	150578851	150578851	G	A	<i>MCL1</i>	Non-synonymous SNV	MCL1:NM_021960:exon1:c.C680T:p.A227V,MCL1:NM_182763:exon1:c.C680T:p.A227V,MCL1:NM_001197320:exon2:c.C221T:p.A74V	444	0.49
chr16	8900578	8900578	T	C	<i>USP7</i>	Non-synonymous SNV	USP7:NM_001286457:exon21:c.A2213G:p.D738G, USP7:NM_003470:exon21:c.A2261G:p.D754G, USP7:NM_001286458:exon22:c.A1964G:p.D655G	57	0.47

chr8	105802428	105802428	G	C	ZFPM2	Non-synonymous SNV	ZFPM2:NM_012082:exon8:c.G2346C:p.E782D	13	0.46
chr5	180624003	180624003	T	C	FLT4	Non-synonymous SNV	FLT4:NM_002020:exon11:c.A1480G:p.T494A,FLT4:NM_182925:exon11:c.A1480G:p.T494A	11	0.45
chr4	849932	849932	T	C	GAK	Non-synonymous SNV	GAK:NM_001286833:exon24:c.A3500G:p.K1167R,GAK:NM_005255:exon27:c.A3794G:p.K1265R	110	0.45
chr20	53576098	53576098	T	C	ZNF217	Non-synonymous SNV	ZNF217:NM_006526:exon3:c.A2666G:p.D889G	25	0.44
chr17	21015763	21015763	C	T	USP22	Non-synonymous SNV	USP22:NM_015276:exon6:c.G827A:p.R276Q	30	0.43
chr22	22893804	22893804	A	G	IGLL5	Non-synonymous SNV	IGLL5:NM_001178126:exon2:c.A311G:p.K104R	682	0.39
chr22	22893795	22893795	C	G	IGLL5	Non-synonymous SNV	IGLL5:NM_001178126:exon2:c.C302G:p.T101S	568	0.38
chr1	6197724	6197724	-	T	RPL22	frameshift insertion	RPL22:NM_000983:exon2:c.44dupA:p.K15fs	2186	0.22
chr17	16146494	16146494	-	T	NCOR1	frameshift insertion	NCOR1:NM_001190438:exon7:c.636dupA:p.V213fs,NCOR1:NM_001190440:exon9:c.963dupA:p.V322fs,NCOR1:NM_006311:exon10:c.963dupA:p.V322fs	21	0.19
chr12	14424946	14424946	A	T	ATF7IP	Non-synonymous SNV	ATF7IP:NM_001286514:exon2:c.A1031T:p.N344I,ATF7IP:NM_001286515:exon2:c.A1031T:p.N344I,ATF7IP:NM_018179:exon2:c.A1031T:p.N344I,ATF7IP:NM_181352:exon2:c.A1055T:p.N352I	11	0.18

chr12	14424947	14424947	T	A	<i>ATF7IP</i>	Non-synonymous SNV	ATF7IP:NM_001286514:exon2:c.T1032A:p.N344K,ATF7IP:NM_001286515:exon2:c.T1032A:p.N344K,ATF7IP:NM_018179:exon2:c.T1032A:p.N344K,ATF7IP:NM_181352:exon2:c.T1056A:p.N352K	11	0.18
chr19	15257070	15257070	G	A	<i>BRD4</i>	Non-synonymous SNV	BRD4:NM_014299:exon8:c.C1445T:p.P482L,BRD4:NM_058243:exon8:c.C1445T:p.P482L	11	0.18
chr20	1915319	1915319	C	A	<i>SIRPA</i>	Non-synonymous SNV	SIRPA:NM_001040023:exon2:c.C300A:p.N100K,SIRPA:NM_001040022:exon3:c.C300A:p.N100K,SIRPA:NM_080792:exon3:c.C300A:p.N100K	17	0.18
chr20	1915338	1915338	C	A	<i>SIRPA</i>	Non-synonymous SNV	SIRPA:NM_001040023:exon2:c.C319A:p.R107S,SIRPA:NM_001040022:exon3:c.C319A:p.R107S,SIRPA:NM_080792:exon3:c.C319A:p.R107S	18	0.17
chr2	38986101	38986101	-	T	<i>SOS1</i>	frameshift insertion	SOS1:NM_005633:exon23:c.3724dupA:p.S1242fs	46	0.15
chr20	1915344	1915344	G	A	<i>SIRPA</i>	Non-synonymous SNV	SIRPA:NM_001040023:exon2:c.G325A:p.G109S,SIRPA:NM_001040022:exon3:c.G325A:p.G109S,SIRPA:NM_080792:exon3:c.G325A:p.G109S	17	0.12
chr2	197401450	197401450	-	A	<i>SF3B1</i>	Stop-gain	SF3B1:NM_012433:exon17:c.2445dupT:p.K816_H817delinsX	227	0.11
chr22	22893794	22893794	A	G	<i>IGLL5</i>	Non-synonymous SNV	IGLL5:NM_001178126:exon2:c.A301G:p.T101A	566	0.11

	chr22	22893783	22893783	A	T	<i>IGLL5</i>	Non-synonymous SNV	IGLL5:NM_001178126:exon2:c.A290T:p.Y97F	421	0.09
--	-------	----------	----------	---	---	--------------	--------------------	-----------------------------------------	-----	------

Table S2.18: All Cancer Associated SNVs Called for Case 12. Variants were filtered based on a list of 383 pre-defined cancer genes. The variants were maintained if they were exonic, had a damaging function, previously reported in COSMIC, and were considered pathogenic in the ClinVar Database. Variants were then excluded if they had a variant allele frequency below 0.05, depth below 10, a prevalence of greater than 10% in the 1000 Genomes Project, or if no annotations were present. SNVs are shaded dark orange if the variant allele frequency (VAF) =1, and lighter shades of orange for VAFs>=.5 and<1 and VAFs <.5 and>.4. The 'AA Change' column lists the ref.seq. number, the base pair alteration, and the amino acid alteration.

Case 13	Chromosome	Start	End	Reference	Alteration	Gene	Exonic Function	AA Change	Read Depth	VAF
	chr3	47083895	47083895	G	A	<i>SETD2</i>	Non-synonymous SNV	SETD2:NM_014159:exon12:c.C5885T:p.P1962L	30	1.00
	chr5	35860966	35860966	T	C	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon2:c.T197C:p.I66T	68	1.00
	chr5	35871088	35871088	G	A	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon4:c.G412A:p.V138I	77	1.00
	chrX	77682471	77682471	C	G	<i>ATRX</i>	Non-synonymous SNV	ATRX:NM_138270:exon8:c.G2671C:p.E891Q,ATRX:NM_000489:exon9:c.G2785C:p.E929Q	96	1.00
	chr10	63214620	63214620	G	C	<i>JMJD1C</i>	Non-synonymous SNV	JMJD1C:NM_001282948:exon7:c.C1001G:p.T334S, JMJD1C:NM_032776:exon8:c.C1547G:p.T516S	49	0.63
	chr1	169594799	169594799	C	T	<i>SELP</i>	Non-synonymous SNV	SELP:NM_003005:exon13:c.G2180A:p.G727E	13	0.62
	chr4	55106807	55106807	T	A	<i>KDR</i>	Non-synonymous SNV	KDR:NM_002253:exon11:c.A1416T:p.Q472H	109	0.57
	chr4	24576653	24576653	G	A	<i>DHX15</i>	Non-synonymous SNV	DHX15:NM_001358:exon2:c.C97T:p.R33C	207	0.52
	chr3	66383278	66383278	G	A	<i>LRIG1</i>	Non-synonymous SNV	LRIG1:NM_015541:exon15:c.C2195T:p.P732L	31	0.52

chr17	7676154	7676154	G	C	TP53	Non-synonymous SNV	TP53:NM_001126118:exon3:c.C98G:p.P33R,TP53:NM_000546:exon4:c.C215G:p.P72R,TP53:NM_001126112:exon4:c.C215G:p.P72R,TP53:NM_001126113:exon4:c.C215G:p.P72R,TP53:NM_001126114:exon4:c.C215G:p.P72R,TP53:NM_001276695:exon4:c.C98G:p.P33R,TP53:NM_001276696:exon4:c.C98G:p.P33R,TP53:NM_001276760:exon4:c.C98G:p.P33R,TP53:NM_001276761:exon4:c.C98G:p.P33R	32	0.47
chrX	154400814	154400814	G	A	RPL10	Non-synonymous SNV	RPL10:NM_001256577:exon6:c.G442A:p.V148I,RPL10:NM_001256580:exon6:c.G497A:p.S166N,RPL10:NM_001303624:exon6:c.G605A:p.S202N,RPL10:NM_001303625:exon7:c.G605A:p.S202N,RPL10:NM_006013:exon7:c.G605A:p.S202N	858	0.45
chr12	14424747	14424747	G	A	ATF7IP	Non-synonymous SNV	ATF7IP:NM_001286514:exon2:c.G832A:p.E278K,ATF7IP:NM_001286515:exon2:c.G832A:p.E278K,ATF7IP:NM_018179:exon2:c.G832A:p.E278K,ATF7IP:NM_181352:exon2:c.G856A:p.E286K	104	0.44
chr9	132595027	132595027	T	C	DDX31	Non-synonymous SNV	DDX31:NM_022779:exon20:c.A2395G:p.I799V	12	0.42

chr11	118505199	118505199	-	A	<i>KMT2A</i>	frameshift insertion	KMT2A:NM_001197104:exon27:c.9308dupA:p.Q3103fs,KMT2A:NM_005933:exon27:c.9299dupA:p.Q3100fs	15	0.40
chr10	63208191	63208196	CTAAAC	-	<i>JMJD1C</i>	Non-frameshift deletion	JMJD1C:NM_001282948:exon9:c.2927_2932del:p.976_978del,JMJD1C:NM_032776:exon10:c.3473_3478del:p.1158_1160del	61	0.39
chr1	16946274	16946274	G	A	<i>CROCC</i>	Non-synonymous SNV	CROCC:NM_014675:exon16:c.G2152A:p.V718M	11	0.27
chr1	16946292	16946292	A	G	<i>CROCC</i>	Non-synonymous SNV	CROCC:NM_014675:exon16:c.A2170G:p.M724V	11	0.27
chr19	4054601	4054601	G	T	<i>ZBTB7A</i>	Non-synonymous SNV	ZBTB7A:NM_015898:exon2:c.C632A:p.A211E	11	0.27
chr19	4054604	4054604	G	T	<i>ZBTB7A</i>	Non-synonymous SNV	ZBTB7A:NM_015898:exon2:c.C629A:p.A210D	11	0.27
chr20	56383154	56383154	A	G	<i>AURKA</i>	Non-synonymous SNV	AURKA:NM_003600:exon5:c.T397C:p.F133L,AURKA:NM_198435:exon5:c.T397C:p.F133L,AURKA:NM_198437:exon5:c.T397C:p.F133L,AURKA:NM_198434:exon6:c.T397C:p.F133L,AURKA:NM_198436:exon6:c.T397C:p.F133L,AURKA:NM_198433:exon7:c.T397C:p.F133L	15	0.27
chr4	105275672	105275672	T	G	<i>TET2</i>	Non-synonymous SNV	TET2:NM_001127208:exon11:c.T5162G:p.L1721W	27	0.26
chr15	41762197	41762197	-	A	<i>MGA</i>	frameshift insertion	MGA:NM_001080541:exon21:c.6953dupA:p.Q2318fs,MGA:NM_001164273:exon22:c.7580dupA:p.Q252	21	0.24

							7fs		
chr20	56383135	56383135	A	G	<i>AURKA</i>	Non-synonymous SNV	AURKA:NM_003600:exon5:c.T416C:p.L139P,AURKA:NM_198435:exon5:c.T416C:p.L139P,AURKA:NM_198437:exon5:c.T416C:p.L139P,AURKA:NM_198434:exon6:c.T416C:p.L139P,AURKA:NM_198436:exon6:c.T416C:p.L139P,AURKA:NM_198433:exon7:c.T416C:p.L139P	14	0.21
chrX	77681846	77681846	C	A	<i>ATRX</i>	Non-synonymous SNV	ATRX:NM_138270:exon8:c.G3296T:p.R1099I,ATRX:NM_000489:exon9:c.G3410T:p.R1137I	48	0.21
chr19	4054607	4054607	A	C	<i>ZBTB7A</i>	Non-synonymous SNV	ZBTB7A:NM_015898:exon2:c.T626G:p.V209G	10	0.20
chr1	6197724	6197724	-	T	<i>RPL22</i>	frameshift insertion	RPL22:NM_000983:exon2:c.44dupA:p.K15fs	573	0.19
chr22	20092882	20092882	-	A	<i>DGCR8</i>	frameshift insertion	DGCR8:NM_022720:exon8:c.1681dupA:p.S560fs	31	0.16
chr22	22893783	22893783	A	T	<i>IGLL5</i>	Non-synonymous SNV	IGLL5:NM_001178126:exon2:c.A290T:p.Y97F	504	0.15
chr4	24548878	24548878	-	T	<i>DHX15</i>	frameshift insertion	DHX15:NM_001358:exon6:c.1224dupA:p.Q409fs	144	0.13
chr10	21673692	21673692	A	G	<i>MLLT10</i>	Non-synonymous SNV	MLLT10:NM_001195626:exon10:c.A1394G:p.E465G,MLLT10:NM_004641:exon11:c.A1394G:p.E465G	23	0.13
chr11	108293436	108293436	-	A	<i>ATM</i>	frameshift insertion	ATM:NM_000051:exon31:c.4736dupA:p.Q1579fs	90	0.12

	chr10	21673696	21673696	G	T	<i>MLLT10</i>	Non-synonymous SNV	MLLT10:NM_001195626:exon10:c.G1398T:p.K466N,MLLT10:NM_004641:exon11:c.G1398T:p.K466N	26	0.12
--	-------	----------	----------	---	---	---------------	--------------------	--------------------------------------------------------------------------------------	----	------

Table S2.19: All Cancer Associated SNVs Called for Case 13. Variants were filtered based on a list of 383 pre-defined cancer genes. The variants were maintained if they were exonic, had a damaging function, previously reported in COSMIC, and were considered pathogenic in the ClinVar Database. Variants were then excluded if they had a variant allele frequency below 0.05, depth below 10, a prevalence of greater than 10% in the 1000 Genomes Project, or if no annotations were present. SNVs are shaded dark orange if the variant allele frequency (VAF) =1, and lighter shades of orange for VAFs >=0.5 and <1 and VAFs <0.5 and >0.4. The 'AA Change' column lists the ref.seq. number, the base pair alteration, and the amino acid alteration.

Case 14	Chromosome	Start	End	Reference	Alteration	Gene	Exonic Function	AA Change	Read Depth	VAF
	chr17	7676154	7676154	G	C	TP53	Non-synonymous SNV	TP53:NM_001126118:exon3:c.C98G:p.P33R,TP53:NM_000546:exon4:c.C215G:p.P72R,TP53:NM_001126112:exon4:c.C215G:p.P72R,TP53:NM_001126113:exon4:c.C215G:p.P72R,TP53:NM_001126114:exon4:c.C215G:p.P72R,TP53:NM_001276695:exon4:c.C98G:p.P33R,TP53:NM_001276696:exon4:c.C98G:p.P33R,TP53:NM_001276760:exon4:c.C98G:p.P33R,TP53:NM_001276761:exon4:c.C98G:p.P33R	22	1.00
	chr20	1915406	1915408	CGA	-	SIRPA	Non-frameshift deletion	SIRPA:NM_001040023:exon2:c.387_389del:p.129_130del,SIRPA:NM_001040022:exon3:c.387_389del:p.129_130del,SIRPA:NM_080792:exon3:c.387_389del:p.129_130del	26	1.00
	chrX	40074864	40074864	G	A	BCOR	Non-synonymous SNV	BCOR:NM_001123383:exon4:c.C482T:p.A161V,BCOR:NM_001123384:exon4:c.C482T:p.A161V,BCOR:NM_001123385:exon4:c.C482T:p.A161V,BCOR:NM_017745:exon4:c.C482T:p.A161V	14	1.00
chr20	1915303	1915303	-	GT	SIRPA	frameshift insertion	SIRPA:NM_001040023:exon2:c.284_285insGT:p.D95fs,SIRPA:NM_001040022:exon3:c.284_285insGT:p.D95fs,SIRPA:NM_080792:exon3:c.284_285insGT	56	0.95	

							:p.D95fs		
chr20	1915305	1915306	CT	-	SIRPA	frameshift deletion	SIRPA:NM_001040023:exon2:c.286_287del:p.L96fs, SIRPA:NM_001040022:exon3:c.286_287del:p.L96fs, SIRPA:NM_080792:exon3:c.286_287del:p.L96fs	56	0.95
chr20	1915319	1915319	C	A	SIRPA	Non-synonymous SNV	SIRPA:NM_001040023:exon2:c.C300A:p.N100K, SIRPA:NM_001040022:exon3:c.C300A:p.N100K, SIRPA:NM_080792:exon3:c.C300A:p.N100K	50	0.94
chr20	1915338	1915338	C	A	SIRPA	Non-synonymous SNV	SIRPA:NM_001040023:exon2:c.C319A:p.R107S, SIRPA:NM_001040022:exon3:c.C319A:p.R107S, SIRPA:NM_080792:exon3:c.C319A:p.R107S	32	0.91
chr7	55161562	55161562	G	A	EGFR	Non-synonymous SNV	EGFR:NM_005228:exon13:c.G1562A:p.R521K, EGFR:NM_201282:exon13:c.G1562A:p.R521K, EGFR:NM_201284:exon13:c.G1562A:p.R521K	31	0.90
chr20	1915344	1915344	G	A	SIRPA	Non-synonymous SNV	SIRPA:NM_001040023:exon2:c.G325A:p.G109S, SIRPA:NM_001040022:exon3:c.G325A:p.G109S, SIRPA:NM_080792:exon3:c.G325A:p.G109S	25	0.88
chr9	132595027	132595027	T	C	DDX31	Non-synonymous SNV	DDX31:NM_022779:exon20:c.A2395G:p.I799V	10	0.70

chr9	136505767	136505767	G	A	<i>NOTCH1</i>	Non-synonymous SNV	NOTCH1:NM_017617:exon25:c.C4129T:p.P1377S	74	0.68
chr3	142453222	142453222	G	C	<i>ATR</i>	Non-synonymous SNV	ATR:NM_001184:exon46:c.C7667G:p.T2556S	57	0.63
chr3	47121396	47121396	C	T	<i>SETD2</i>	Non-synonymous SNV	SETD2:NM_014159:exon3:c.G3240A:p.M1080I	83	0.53
chr19	41238025	41238025	G	A	<i>AXL</i>	Non-synonymous SNV	AXL:NM_001278599:exon4:c.G61A:p.V21M,AXL:NM_001699:exon7:c.G865A:p.V289M,AXL:NM_021913:exon7:c.G865A:p.V289M	132	0.52
chr3	128486108	128486108	C	T	<i>GATA2</i>	Non-synonymous SNV	GATA2:NM_001145662:exon3:c.G490A:p.A164T,GATA2:NM_032638:exon3:c.G490A:p.A164T,GATA2:NM_001145661:exon4:c.G490A:p.A164T	10	0.50
chr4	105234042	105234042	C	T	<i>TET2</i>	Non-synonymous SNV	TET2:NM_001127208:exon3:c.C100T:p.L34F,TET2:NM_017628:exon3:c.C100T:p.L34F	42	0.50
chr5	150129960	150129960	C	T	<i>PDGFRB</i>	Non-synonymous SNV	PDGFRB:NM_002609:exon10:c.G1376A:p.R459H	380	0.50
chr5	180624003	180624003	T	C	<i>FLT4</i>	Non-synonymous SNV	FLT4:NM_002020:exon11:c.A1480G:p.T494A,FLT4:NM_182925:exon11:c.A1480G:p.T494A	72	0.47
chr1	119929089	119929089	C	T	<i>NOTCH2</i>	Non-synonymous SNV	NOTCH2:NM_024408:exon23:c.G3779A:p.R1260H	228	0.46
chr8	127740499	127740499	C	A	<i>MYC</i>	Non-synonymous SNV	MYC:NM_002467:exon3:c.C906A:p.H302Q	123	0.46

chr1	179108013	179108013	C	T	<i>ABL2</i>	Non-synonymous SNV	ABL2:NM_001168236:exon11:c.G3191A:p.S1064N, ABL2:NM_001168238:exon12:c.G2882A:p.S961N, ABL2:NM_001168239:exon12:c.G2837A:p.S946N, ABL2:NM_005158:exon12:c.G3209A:p.S1070N, ABL2:NM_007314:exon12:c.G3254A:p.S1085N, ABL2:NM_001136000:exon13:c.G2900A:p.S967N, ABL2:NM_001168237:exon13:c.G2945A:p.S982N	53	0.43
chr4	105275843	105275843	A	G	<i>TET2</i>	Non-synonymous SNV	TET2:NM_001127208:exon11:c.A5333G:p.H1778R	21	0.43
chr16	72950696	72950696	C	A	<i>ZFHX3</i>	Non-synonymous SNV	ZFHX3:NM_001164766:exon2:c.G247T:p.A83S, ZFHX3:NM_006885:exon3:c.G2989T:p.A997S	77	0.40
chr3	49360949	49360949	T	C	<i>RHOA</i>	Non-synonymous SNV	RHOA:NM_001313943:exon5:c.A526G:p.R176G	30	0.40
chr19	15254239	15254239	C	T	<i>BRD4</i>	Non-synonymous SNV	BRD4:NM_014299:exon11:c.G2071A:p.G691S, BRD4:NM_058243:exon11:c.G2071A:p.G691S	58	0.40
chr12	111434918	111434918	T	C	<i>SH2B3</i>	Non-synonymous SNV	SH2B3:NM_001291424:exon1:c.T17C:p.L6P	14	0.36
chr16	3664410	3664410	-	CGG	<i>TRAP1</i>	Non-frameshift insertion	TRAP1:NM_001272049:exon12:c.1273_1274insCCG:p.G425delinsAG, TRAP1:NM_016292:exon13:c.1432_1433insCCG:p.G478delinsAG	38	0.34
chr3	49360990	49360990	T	C	<i>RHOA</i>	Non-synonymous SNV	RHOA:NM_001313943:exon5:c.A485G:p.K162R	22	0.32

chr13	32332592	32332592	A	C	<i>BRCA2</i>	Non-synonymous SNV	BRCA2:NM_000059:exon10:c.A1114C:p.N372H	26	0.31
chr13	32379885	32379885	-	A	<i>BRCA2</i>	frameshift insertion	BRCA2:NM_000059:exon23:c.9090dupA:p.T3030fs	17	0.29
chr3	49360951	49360951	T	C	<i>RHOA</i>	Non-synonymous SNV	RHOA:NM_001313943:exon5:c.A524G:p.Y175C	30	0.27
chr11	118505199	118505199	-	A	<i>KMT2A</i>	frameshift insertion	KMT2A:NM_001197104:exon27:c.9308dupA:p.Q3103fs,KMT2A:NM_005933:exon27:c.9299dupA:p.Q3100fs	17	0.24
chr1	6197724	6197724	-	T	<i>RPL22</i>	frameshift insertion	RPL22:NM_000983:exon2:c.44dupA:p.K15fs	365	0.21
chr22	20092882	20092882	-	A	<i>DGCR8</i>	frameshift insertion	DGCR8:NM_022720:exon8:c.1681dupA:p.S560fs	19	0.21
chr3	49360997	49360997	T	C	<i>RHOA</i>	Non-synonymous SNV	RHOA:NM_001313943:exon5:c.A478G:p.R160G	24	0.21
chr6	113970875	113970875	-	T	<i>HDAC2</i>	frameshift insertion	HDAC2:NM_001527:exon1:c.33dupA:p.V12fs	58	0.21
chr12	6601979	6601981	TCC	-	<i>CHD4</i>	Non-frameshift deletion	CHD4:NM_001297553:exon3:c.396_398del:p.132_133del,CHD4:NM_001273:exon4:c.417_419del:p.139_140del	150	0.19
chr10	63214489	63214489	-	T	<i>JMJD1C</i>	frameshift insertion	JMJD1C:NM_001282948:exon7:c.1131dupA:p.D378fs,JMJD1C:NM_032776:exon8:c.1677dupA:p.D560fs	79	0.19
chrX	77681846	77681846	C	A	<i>ATRX</i>	Non-synonymous SNV	ATRX:NM_138270:exon8:c.G3296T:p.R1099I,ATRX:NM_000489:exon9:c.G3410T:p.R1137I	76	0.18

chr18	63318650	63318650	C	G	<i>BCL2</i>	Non-synonymous SNV	BCL2:NM_000633:exon2:c.G17C:p.R6T,BCL2:NM_000657:exon2:c.G17C:p.R6T	12	0.17
chr17	16146494	16146494	-	T	<i>NCOR1</i>	frameshift insertion	NCOR1:NM_001190438:exon7:c.636dupA:p.V213fs,NCOR1:NM_001190440:exon9:c.963dupA:p.V322fs,NCOR1:NM_006311:exon10:c.963dupA:p.V322fs	98	0.15
chr19	15178881	15178881	T	C	<i>NOTCH3</i>	Non-synonymous SNV	NOTCH3:NM_000435:exon23:c.A3779G:p.Q1260R	111	0.14
chr1	16946292	16946292	A	G	<i>CROCC</i>	Non-synonymous SNV	CROCC:NM_014675:exon16:c.A2170G:p.M724V	23	0.13
chr1	16946274	16946274	G	A	<i>CROCC</i>	Non-synonymous SNV	CROCC:NM_014675:exon16:c.G2152A:p.V718M	26	0.12
chr22	22893804	22893804	A	G	<i>IGLL5</i>	Non-synonymous SNV	IGLL5:NM_001178126:exon2:c.A311G:p.K104R	873	0.08

Table S2.20: All Cancer Associated SNVs Called for Case 14. Variants were filtered based on a list of 383 pre-defined cancer genes. The variants were maintained if they were exonic, had a damaging function, previously reported in COSMIC, and were considered pathogenic in the ClinVar Database. Variants were then excluded if they had a variant allele frequency below 0.05, depth below 10, a prevalence of greater than 10% in the 1000 Genomes Project, or if no annotations were present. SNVs are shaded dark orange if the variant allele frequency (VAF) =1, and lighter shades of orange for VAFs >=0.5 and <1 and VAFs <0.5 and >0.4. The 'AA Change' column lists the ref.seq. number, the base pair alteration, and the amino acid alteration.

Case 16	Chromosome	Start	End	Reference	Alteration	Gene	Exonic Function	AA Change	Read Depth	VAF
	chr10	19204380	19204380	G	A	<i>MALRD1</i>	Non-synonymous SNV	MALRD1:NM_001142308:exon16:c.G2177A:p.S726N	12	1.00
	chr17	7676154	7676154	G	C	<i>TP53</i>	Non-synonymous SNV	TP53:NM_001126118:exon3:c.C98G:p.P33R,TP53:NM_000546:exon4:c.C215G:p.P72R,TP53:NM_001126112:exon4:c.C215G:p.P72R,TP53:NM_001126113:exon4:c.C215G:p.P72R,TP53:NM_001126114:exon4:c.C215G:p.P72R,TP53:NM_001276695:exon4:c.C98G:p.P33R,TP53:NM_001276696:exon4:c.C98G:p.P33R,TP53:NM_001276760:exon4:c.C98G:p.P33R,TP53:NM_001276761:exon4:c.C98G:p.P33R	44	1.00
	chr5	35860966	35860966	T	C	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon2:c.T197C:p.I66T	142	1.00
	chr5	35871088	35871088	G	A	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon4:c.G412A:p.V138I	125	1.00
	chrX	77682471	77682471	C	G	<i>ATRX</i>	Non-synonymous SNV	ATRX:NM_138270:exon8:c.G2671C:p.E891Q,ATRX:NM_000489:exon9:c.G2785C:p.E929Q	116	1.00
	chrX	154400814	154400814	G	A	<i>RPL10</i>	Non-synonymous SNV	RPL10:NM_001256577:exon6:c.G442A:p.V148I,RPL10:NM_001256580:exon6:c.G497A:p.S166N,RPL10:NM_001303624:exon6:c.G605A:p.S202N,RPL10:NM_001303625:exon7:c.G605A:p.S202N,RPL10:NM_001303625:exon7:c.G605A:p.S202N,RPL10:NM_001303625:exon7:c.G605A:p.S202N	1071	0.69

							006013:exon7:c.G605A:p.S202N		
chr1	179143010	179143010	C	G	<i>ABL2</i>	Non-synonymous SNV	ABL2:NM_001136000:exon1:c.G35C:p.S12T,ABL2:NM_001168239:exon1:c.G35C:p.S12T,ABL2:NM_005158:exon1:c.G35C:p.S12T	12	0.67
chr1	16960806	16960806	A	G	<i>CROCC</i>	Non-synonymous SNV	CROCC:NM_014675:exon27:c.A4081G:p.T1361A	16	0.63
chr1	16969269	16969269	A	G	<i>CROCC</i>	Non-synonymous SNV	CROCC:NM_014675:exon32:c.A5230G:p.S1744G	36	0.61
chr16	72950696	72950696	C	A	<i>ZFH3</i>	Non-synonymous SNV	ZFH3:NM_001164766:exon2:c.G247T:p.A83S,ZFH3:NM_006885:exon3:c.G2989T:p.A997S	56	0.59
chr16	71648927	71648927	C	T	<i>PHLPP2</i>	Non-synonymous SNV	PHLPP2:NM_001289003:exon18:c.G3734A:p.R1245Q,PHLPP2:NM_015020:exon19:c.G3935A:p.R1312Q	12	0.58
chr20	53576098	53576098	T	C	<i>ZNF217</i>	Non-synonymous SNV	ZNF217:NM_006526:exon3:c.A2666G:p.D889G	57	0.58
chr5	35874473	35874473	C	T	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon6:c.C731T:p.T244I	114	0.57
chr7	152162598	152162598	G	A	<i>KMT2C</i>	Non-synonymous SNV	KMT2C:NM_170606:exon43:c.C10979T:p.S3660L	73	0.55
chr9	8485834	8485834	G	A	<i>PTPRD</i>	Non-synonymous	PTPRD:NM_002839:exon28:c.C2983T:p.R995C	11	0.55

						SNV			
chr7	55161562	55161562	G	A	<i>EGFR</i>	Non-synonymous SNV	EGFR:NM_005228:exon13:c.G1562A:p.R521K,EGFR:NM_201282:exon13:c.G1562A:p.R521K,EGFR:NM_201284:exon13:c.G1562A:p.R521K	127	0.54
chr5	68292320	68292320	G	A	<i>PIK3R1</i>	Non-synonymous SNV	PIK3R1:NM_181504:exon2:c.G168A:p.M56I,PIK3R1:NM_181524:exon2:c.G78A:p.M26I,PIK3R1:NM_181523:exon8:c.G978A:p.M326I	193	0.52
chr3	66380763	66380763	C	T	<i>LRIG1</i>	Non-synonymous SNV	LRIG1:NM_015541:exon18:c.G2869A:p.A957T	105	0.50
chr4	54727298	54727298	A	C	<i>KIT</i>	Non-synonymous SNV	KIT:NM_000222:exon10:c.A1621C:p.M541L,KIT:NM_001093772:exon10:c.A1609C:p.M537L	93	0.48
chr3	47083895	47083895	G	A	<i>SETD2</i>	Non-synonymous SNV	SETD2:NM_014159:exon12:c.C5885T:p.P1962L	81	0.48
chr9	21816759	21816759	G	A	<i>MTAP</i>	Non-synonymous SNV	MTAP:NM_002451:exon3:c.G166A:p.V56I	113	0.48
chr12	111434918	111434918	T	C	<i>SH2B3</i>	Non-synonymous SNV	SH2B3:NM_001291424:exon1:c.T17C:p.L6P	30	0.47
chr10	63177741	63177741	G	C	<i>JMJD1C</i>	Non-synonymous SNV	JMJD1C:NM_001282948:exon22:c.C6654G:p.D2218E,JMJD1C:NM_032776:exon23:c.C7200G:p.D2400E	233	0.46
chr3	66380453	66380453	G	C	<i>LRIG1</i>	Non-synonymous SNV	LRIG1:NM_015541:exon19:c.C3092G:p.P1031R	153	0.46

chr19	15238894	15238894	-	GCT	<i>BRD4</i>	Non-frameshift insertion	BRD4:NM_058243:exon19:c.3868_3869insAGC:p.R1290delinsQR	33	0.45
chr9	132595027	132595027	T	C	<i>DDX31</i>	Non-synonymous SNV	DDX31:NM_022779:exon20:c.A2395G:p.I799V	21	0.43
chr4	849720	849720	C	T	<i>GAK</i>	Non-synonymous SNV	GAK:NM_001286833:exon25:c.G3595A:p.D1199N, GAK:NM_005255:exon28:c.G3889A:p.D1297N	81	0.41
chr15	41729314	41729314	T	C	<i>MGA</i>	Non-synonymous SNV	MGA:NM_001080541:exon11:c.T3808C:p.C1270R, MGA:NM_001164273:exon11:c.T3808C:p.C1270R	47	0.40
chr1	16954783	16954783	G	A	<i>CROCC</i>	Non-synonymous SNV	CROCC:NM_014675:exon23:c.G3371A:p.R1124Q	13	0.38
chr9	37357252	37357252	-	A	<i>ZCCHC7</i>	frameshift insertion	ZCCHC7:NM_001289119:exon9:c.1617dupA:p.R539fs,ZCCHC7:NM_001289120:exon9:c.1617dupA:p.R539fs,ZCCHC7:NM_001289121:exon9:c.1617dupA:p.R539fs,ZCCHC7:NM_032226:exon9:c.1617dupA:p.R539fs	29	0.28
chr3	49360949	49360949	T	C	<i>RHOA</i>	Non-synonymous SNV	RHOA:NM_001313943:exon5:c.A526G:p.R176G	26	0.27
chr1	16954740	16954740	G	A	<i>CROCC</i>	Non-synonymous SNV	CROCC:NM_014675:exon23:c.G3328A:p.V1110M	15	0.27
chr1	16946274	16946274	G	A	<i>CROCC</i>	Non-synonymous SNV	CROCC:NM_014675:exon16:c.G2152A:p.V718M	40	0.25
chr1	6197724	6197724	-	T	<i>RPL22</i>	frameshift insertion	RPL22:NM_000983:exon2:c.44dupA:p.K15fs	763	0.20

chr20	1915303	1915303	-	GT	SIRPA	frameshift insertion	SIRPA:NM_001040023:exon2:c.284_285insGT:p.D95fs,SIRPA:NM_001040022:exon3:c.284_285insGT:p.D95fs,SIRPA:NM_080792:exon3:c.284_285insGT:p.D95fs	103	0.20
chr20	1915305	1915306	CT	-	SIRPA	frameshift deletion	SIRPA:NM_001040023:exon2:c.286_287del:p.L96fs,SIRPA:NM_001040022:exon3:c.286_287del:p.L96fs,SIRPA:NM_080792:exon3:c.286_287del:p.L96fs	104	0.20
chr3	49360961	49360961	T	C	RHOA	Non-synonymous SNV	RHOA:NM_001313943:exon5:c.A514G:p.S172G	26	0.19
chr7	152177062	152177062	-	T	KMT2C	frameshift insertion	KMT2C:NM_170606:exon38:c.8390dupA:p.K2797fs	101	0.19
chr20	1915319	1915319	C	A	SIRPA	Non-synonymous SNV	SIRPA:NM_001040023:exon2:c.C300A:p.N100K,SIRPA:NM_001040022:exon3:c.C300A:p.N100K,SIRPA:NM_080792:exon3:c.C300A:p.N100K	104	0.18
chr1	16946292	16946292	A	G	CROCC	Non-synonymous SNV	CROCC:NM_014675:exon16:c.A2170G:p.M724V	36	0.17
chr20	1915338	1915338	C	A	SIRPA	Non-synonymous SNV	SIRPA:NM_001040023:exon2:c.C319A:p.R107S,SIRPA:NM_001040022:exon3:c.C319A:p.R107S,SIRPA:NM_080792:exon3:c.C319A:p.R107S	109	0.17
chr3	49360951	49360951	T	C	RHOA	Non-synonymous SNV	RHOA:NM_001313943:exon5:c.A524G:p.Y175C	25	0.16

chr20	1915406	1915408	CGA	-	SIRPA	Non-frameshift deletion	SIRPA:NM_001040023:exon2:c.387_389del:p.129_130del,SIRPA:NM_001040022:exon3:c.387_389del:p.129_130del,SIRPA:NM_080792:exon3:c.387_389del:p.129_130del	81	0.15
chr20	1915344	1915344	G	A	SIRPA	Non-synonymous SNV	SIRPA:NM_001040023:exon2:c.G325A:p.G109S,SIRPA:NM_001040022:exon3:c.G325A:p.G109S,SIRPA:NM_080792:exon3:c.G325A:p.G109S	104	0.14
chr1	16945580	16945580	G	A	CROCC	Non-synonymous SNV	CROCC:NM_014675:exon15:c.G2110A:p.A704T	30	0.13
chrX	77681846	77681846	C	A	ATRX	Non-synonymous SNV	ATRX:NM_138270:exon8:c.G3296T:p.R1099I,ATRX:NM_000489:exon9:c.G3410T:p.R1137I	93	0.12

Table S2.21: All Cancer Associated SNVs Called for Case 16. Variants were filtered based on a list of 383 pre-defined cancer genes. The variants were maintained if they were exonic, had a damaging function, previously reported in COSMIC, and were considered pathogenic in the ClinVar Database. Variants were then excluded if they had a variant allele frequency below 0.05, depth below 10, a prevalence of greater than 10% in the 1000 Genomes Project, or if no annotations were present. SNVs are shaded dark orange if the variant allele frequency (VAF) =1, and lighter shades of orange for VAFs >=0.5 and <1 and VAFs <0.5 and >0.4. The 'AA Change' column lists the ref.seq. number, the base pair alteration, and the amino acid alteration.

	Chromosome	Start	End	Reference	Alteration	Gene	Exonic Function	AA Change	Read Depth	VAF
Case 65	chr17	7676154	7676154	G	C	<i>TP53</i>	Non-synonymous SNV	TP53:NM_001126118:exon3:c.C98G:p.P33R,TP53:NM_000546:exon4:c.C215G:p.P72R,TP53:NM_001126112:exon4:c.C215G:p.P72R,TP53:NM_001126113:exon4:c.C215G:p.P72R,TP53:NM_001126114:exon4:c.C215G:p.P72R,TP53:NM_001276695:exon4:c.C98G:p.P33R,TP53:NM_001276696:exon4:c.C98G:p.P33R,TP53:NM_001276760:exon4:c.C98G:p.P33R,TP53:NM_001276761:exon4:c.C98G:p.P33R	111	1.00
	chr4	105275794	105275794	A	G	<i>TET2</i>	Non-synonymous SNV	TET2:NM_001127208:exon11:c.A5284G:p.I1762V	35	1.00
	chr9	132595027	132595027	T	C	<i>DDX31</i>	Non-synonymous SNV	DDX31:NM_022779:exon20:c.A2395G:p.I799V	32	1.00
	chr12	6667904	6667906	TGC	-	<i>ZNF384</i>	Non-frameshift deletion	ZNF384:NM_001039920:exon9:c.1194_1196del:p.398_399del,ZNF384:NM_133476:exon10:c.1359_1361del:p.453_454del,ZNF384:NM_001135734:exon11:c.1542_1544del:p.514_515del	20	0.90
	chr7	138917686	138917686	G	C	<i>KIAA1549</i>	Non-synonymous SNV	KIAA1549:NM_001164665:exon2:c.C1940G:p.S647C,KIAA1549:NM_020910:exon2:c.C1940G:p.S647C	14	0.64
	chr13	28049450	28049450	C	T	<i>FLT3</i>	Non-synonymous SNV	FLT3:NM_004119:exon8:c.G970A:p.D324N	19	0.63

chr17	43071077	43071077	T	C	<i>BRCA1</i>	Non-synonymous SNV	BRCA1:NM_007297:exon14:c.A4696G:p.S1566G, BRCA1:NM_007298:exon14:c.A1525G:p.S509G, BRCA1:NM_007294:exon15:c.A4837G:p.S1613G, BRCA1:NM_007299:exon15:c.A1525G:p.S509G, BRCA1:NM_007300:exon16:c.A4900G:p.S1634G	16	0.63
chr6	41936060	41936060	C	A	<i>CCND3</i>	Non-synonymous SNV	CCND3:NM_001136125:exon4:c.G543T:p.E181D, CCND3:NM_001136126:exon4:c.G171T:p.E57D, CCND3:NM_001136017:exon5:c.G516T:p.E172D, CCND3:NM_001287427:exon5:c.G609T:p.E203D, CCND3:NM_001287434:exon5:c.G171T:p.E57D, CCND3:NM_001760:exon5:c.G759T:p.E253D	233	0.61
chr9	21970986	21970986	C	G	<i>CDKN2A</i>	Non-synonymous SNV	CDKN2A:NM_000077:exon2:c.G373C:p.D125H, CDKN2A:NM_001195132:exon2:c.G373C:p.D125H	117	0.57
chr18	44952641	44952641	G	A	<i>SETBP1</i>	Non-synonymous SNV	SETBP1:NM_015559:exon4:c.G3301A:p.V1101I	16	0.56
chr5	35874473	35874473	C	T	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon6:c.C731T:p.T244I	176	0.56
chr2	25799408	25799408	T	C	<i>ASXL2</i>	Non-synonymous SNV	ASXL2:NM_018263:exon4:c.A380G:p.K127R	196	0.55
chr16	72787445	72787445	G	A	<i>ZFHX3</i>	Non-synonymous SNV	ZFHX3:NM_001164766:exon9:c.C8089T:p.H2697Y, ZFHX3:NM_006885:exon10:c.C10831T:p.H3611Y	21	0.52

chr17	43092919	43092919	G	A	<i>BRCA1</i>	Non-synonymous SNV	BRCA1:NM_007297:exon9:c.C2471T:p.P824L,BRCA1:NM_007294:exon10:c.C2612T:p.P871L,BRCA1:NM_007300:exon10:c.C2612T:p.P871L	21	0.52
chr4	55113391	55113391	C	T	<i>KDR</i>	Non-synonymous SNV	KDR:NM_002253:exon7:c.G889A:p.V297I	270	0.52
chr12	49038030	49038030	G	T	<i>KMT2D</i>	Non-synonymous SNV	KMT2D:NM_003482:exon34:c.C9326A:p.P3109H	112	0.52
chr3	47083895	47083895	G	A	<i>SETD2</i>	Non-synonymous SNV	SETD2:NM_014159:exon12:c.C5885T:p.P1962L	104	0.51
chr9	5126343	5126343	G	A	<i>JAK2</i>	Non-synonymous SNV	JAK2:NM_004972:exon24:c.G3188A:p.R1063H	285	0.51
chr17	43092418	43092418	T	C	<i>BRCA1</i>	Non-synonymous SNV	BRCA1:NM_007297:exon9:c.A2972G:p.E991G,BRCA1:NM_007294:exon10:c.A3113G:p.E1038G,BRCA1:NM_007300:exon10:c.A3113G:p.E1038G	14	0.50
chr8	11548067	11548067	G	A	<i>BLK</i>	Non-synonymous SNV	BLK:NM_001715:exon4:c.G211A:p.A71T	10	0.50
chr1	245915549	245915549	C	T	<i>SMYD3</i>	Non-synonymous SNV	SMYD3:NM_001167740:exon8:c.G794A:p.R265H,SMYD3:NM_022743:exon8:c.G617A:p.R206H	496	0.48
chr9	132911517	132911517	A	G	<i>TSC1</i>	Non-synonymous SNV	TSC1:NM_001162427:exon9:c.T812C:p.M271T,TSC1:NM_000368:exon10:c.T965C:p.M322T,TSC1:NM_001162426:exon10:c.T965C:p.M322T	44	0.48

chr1	179143010	179143010	C	G	<i>ABL2</i>	Non-synonymous SNV	ABL2:NM_001136000:exon1:c.G35C:p.S12T,ABL2:NM_001168239:exon1:c.G35C:p.S12T,ABL2:NM_005158:exon1:c.G35C:p.S12T	94	0.47
chr5	35871088	35871088	G	A	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon4:c.G412A:p.V138I	271	0.45
chr4	55106807	55106807	T	A	<i>KDR</i>	Non-synonymous SNV	KDR:NM_002253:exon11:c.A1416T:p.Q472H	195	0.45
chr11	3702551	3702551	G	C	<i>NUP98</i>	Non-synonymous SNV	NUP98:NM_016320:exon23:c.C3424G:p.Q1142E,NUP98:NM_139132:exon23:c.C3424G:p.Q1142E	173	0.43
chr5	35860966	35860966	T	C	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon2:c.T197C:p.I66T	299	0.43
chr1	169613079	169613079	C	T	<i>SELP</i>	Non-synonymous SNV	SELP:NM_003005:exon5:c.G625A:p.V209M	23	0.39
chr9	131163898	131163898	A	G	<i>NUP214</i>	Non-synonymous SNV	NUP214:NM_005085:exon20:c.A2752G:p.N918D	46	0.39
chr6_GL000253_v2_alt	3528887	3528895	AGCAGCAGC	-	<i>NOTCH4</i>	Non-frameshift deletion	NOTCH4:NM_004557:exon11:c.36_44del:p.12_15del	42	0.38
chr7	152144741	152144741	T	C	<i>KMT2C</i>	Non-synonymous SNV	KMT2C:NM_170606:exon55:c.A14315G:p.N4772S	113	0.38
chr13	32332592	32332592	A	C	<i>BRCA2</i>	Non-synonymous SNV	BRCA2:NM_000059:exon10:c.A1114C:p.N372H	26	0.31
chr17	43091983	43091983	T	C	<i>BRCA1</i>	Non-synonymous SNV	BRCA1:NM_007297:exon9:c.A3407G:p.K1136R,BRCA1:NM_007294:exon10:c.A3548G:p.K1183R,BRCA1:NM_007300:exon10:c.A3548G:p.K1183R	18	0.28

chrX	77682471	77682471	C	G	<i>ATRX</i>	Non-synonymous SNV	ATRX:NM_138270:exon8:c.G2671C;p.E891Q,ATRX:NM_000489:exon9:c.G2785C;p.E929Q	133	0.26
chr1	6197724	6197724	-	T	<i>RPL22</i>	frameshift insertion	RPL22:NM_000983:exon2:c.44dupA;p.K15fs	1416	0.22
chr11	118505199	118505199	-	A	<i>KMT2A</i>	frameshift insertion	KMT2A:NM_001197104:exon27:c.9308dupA;p.Q3103fs,KMT2A:NM_005933:exon27:c.9299dupA;p.Q3100fs	48	0.21
chr7	152177062	152177062	-	T	<i>KMT2C</i>	frameshift insertion	KMT2C:NM_170606:exon38:c.8390dupA;p.K2797fs	85	0.20
chr3	49360951	49360951	T	C	<i>RHOA</i>	Non-synonymous SNV	RHOA:NM_001313943:exon5:c.A524G;p.Y175C	26	0.19
chr1	16938425	16938425	C	T	<i>CROCC</i>	Non-synonymous SNV	CROCC:NM_014675:exon11:c.C1316T;p.A439V	21	0.19
chrX	77681846	77681846	C	A	<i>ATRX</i>	Non-synonymous SNV	ATRX:NM_138270:exon8:c.G3296T;p.R1099I,ATRX:NM_000489:exon9:c.G3410T;p.R1137I	104	0.16
chr15	41762197	41762197	-	A	<i>MGA</i>	frameshift insertion	MGA:NM_001080541:exon21:c.6953dupA;p.Q2318fs,MGA:NM_001164273:exon22:c.7580dupA;p.Q2527fs	77	0.16
chr10	63214489	63214489	-	T	<i>JMJD1C</i>	frameshift insertion	JMJD1C:NM_001282948:exon7:c.1131dupA;p.D378fs,JMJD1C:NM_032776:exon8:c.1677dupA;p.D560fs	116	0.16
chr19	10151462	10151462	-	T	<i>DNMT1</i>	frameshift insertion	DNMT1:NM_001379:exon23:c.2152dupA;p.M718fs, DNMT1:NM_001130823:exon24:c.2200dupA;p.M734fs	148	0.15

	chr4	24548878	24548878	-	T	<i>DHX15</i>	frameshift insertion	DHX15:NM_001358:exon 6:c.1224dupA:p.Q409fs	470	0.14
	chrX	77684181	77684181	-	T	<i>ATRX</i>	frameshift insertion	ATRX:NM_138270:exon8:c.960dupA:p.L321fs,ATRX:NM_000489:exon9:c.1074dupA:p.L359fs	181	0.12
	chr11	119274893	119274893	-	T	<i>CBL</i>	frameshift insertion	CBL:NM_005188:exon5:c.810dupT:p.A270fs	157	0.11

Table S2.22: All Cancer Associated SNVs Called for Case 65. Variants were filtered based on a list of 383 pre-defined cancer genes. The variants were maintained if they were exonic, had a damaging function, previously reported in COSMIC, and were considered pathogenic in the ClinVar Database. Variants were then excluded if they had a variant allele frequency below 0.05, depth below 10, a prevalence of greater than 10% in the 1000 Genomes Project, or if no annotations were present. SNVs are shaded dark orange if the variant allele frequency (VAF) =1, and lighter shades of orange for VAFs >=0.5 and <1 and VAFs <0.5 and >0.4. The 'AA Change' column lists the ref.seq. number, the base pair alteration, and the amino acid alteration.

	Chromosome	Start	End	Reference	Alteration	Gene	Exonic Function	AA Change	Read Depth	VAF
Case 66	chr3	47083895	47083895	G	A	SETD2	Non-synonymous SNV	SETD2:NM_014159:exon12:c.C5885T:p.P1962L	60	1.00
	chr5	35860966	35860966	T	C	IL7R	Non-synonymous SNV	IL7R:NM_002185:exon2:c.T197C:p.I66T	19	1.00
	chr5	35871088	35871088	G	A	IL7R	Non-synonymous SNV	IL7R:NM_002185:exon4:c.G412A:p.V138I	16	1.00
	chr9	132595027	132595027	T	C	DDX31	Non-synonymous SNV	DDX31:NM_022779:exon20:c.A2395G:p.I799V	18	1.00
	chr5	180630293	180630293	T	C	FLT4	Non-synonymous SNV	FLT4:NM_002020:exon4:c.A445G:p.N149D,FLT4:NM_182925:exon4:c.A445G:p.N149D	27	0.70
	chr6_GL000253_v2_alt	3528887	3528895	AGCAGCAGC	-	NOTCH4	Non-frameshift deletion	NOTCH4:NM_004557:exon1:c.36_44del:p.12_15del	13	0.69
	chr5	180612606	180612606	C	T	FLT4	Non-synonymous SNV	FLT4:NM_002020:exon26:c.G3437A:p.R1146H,FLT4:NM_182925:exon26:c.G3437A:p.R1146H	68	0.68
	chr6	43771215	43771215	G	A	VEGFA	Non-synonymous SNV	VEGFA:NM_001025366:exon1:c.G509A:p.S170N,VEGFA:NM_001025367:exon1:c.G509A:p.S170N,VEGFA:NM_001025368:exon1:c.G509A:p.S170N,VEGFA:NM_001025369:exon1:c.G509A:p.S170N,VEGFA:NM_001025370:exon1:c.G509A:p.S170N,VEGFA:NM_001033756:exon1:c.G509A:p.S170N,VEGFA:NM_001171622:exon1:c.G509A:p.S170N,VEGFA:	35	0.66

							NM_001204385:exon1:c.G509A:p.S170N,VEGFA: NM_003376:exon1:c.G509A:p.S170N		
chr4	105275843	105275843	A	G	<i>TET2</i>	Non-synonymous SNV	TET2:NM_001127208:exon11:c.A5333G:p.H1778R	13	0.54
chr5	38923171	38923171	G	A	<i>OSMR</i>	Non-synonymous SNV	OSMR:NM_003999:exon13:c.G1787A:p.R596Q	223	0.54
chr5	180603313	180603313	C	A	<i>FLT4</i>	Non-synonymous SNV	FLT4:NM_182925:exon30:c.G3971T:p.R1324L	71	0.51
chr5	35874473	35874473	C	T	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon6:c.C731T:p.T244I	14	0.50
chr7	106882147	106882147	A	G	<i>PIK3CG</i>	Non-synonymous SNV	PIK3CG:NM_001282426:exon7:c.A2569G:p.T857A,PIK3CG:NM_001282427:exon7:c.A2569G:p.T857A,PIK3CG:NM_002649:exon7:c.A2569G:p.T857A	18	0.50
chr17	16086391	16086391	C	T	<i>NCOR1</i>	Non-synonymous SNV	NCOR1:NM_001190440:exon22:c.G3116A:p.R1039Q,NCOR1:NM_006311:exon23:c.G3068A:p.R1023Q	36	0.47
chr9	21816759	21816759	G	A	<i>MTAP</i>	Non-synonymous SNV	MTAP:NM_002451:exon3:c.G166A:p.V56I	72	0.47

chr16	72794794	72794794	C	T	ZFH3	Non-synonymous SNV	ZFH3:NM_001164766:exon8:c.G5146A:p.D1716N,ZFH3:NM_006885:exon9:c.G7888A:p.D2630N	12	0.42
chr4	105236541	105236541	T	C	TET2	Non-synonymous SNV	TET2:NM_001127208:exon3:c.T2599C:p.Y867H,TE2:NM_017628:exon3:c.T2599C:p.Y867H	12	0.42
chr12	6667904	6667906	TGC	-	ZNF384	Non-frameshift deletion	ZNF384:NM_001039920:exon9:c.1194_1196del:p.398_399del,ZNF384:NM_133476:exon10:c.1359_1361del:p.453_454del,ZNF384:NM_001135734:exon11:c.1542_1544del:p.514_515del	17	0.41
chr3	128486108	128486108	C	T	GATA2	Non-synonymous SNV	GATA2:NM_001145662:exon3:c.G490A:p.A164T,GATA2:NM_032638:exon3:c.G490A:p.A164T,GATA2:NM_001145661:exon4:c.G490A:p.A164T	17	0.41
chr11	3702551	3702551	G	C	NUP98	Non-synonymous SNV	NUP98:NM_016320:exon23:c.C3424G:p.Q1142E,NUP98:NM_139132:exon23:c.C3424G:p.Q1142E	66	0.41
chr7	116771869	116771869	C	T	MET	Non-synonymous SNV	MET:NM_000245:exon14:c.C2908T:p.R970C,MET:NM_001127500:exon14:c.C2962T:p.R988C	35	0.37
chr7	148828812	148828812	C	G	EZH2	Non-synonymous SNV	EZH2:NM_152998:exon5:c.G436C:p.D146H,EZH2:NM_001203247:exon6:c.G553C:p.D185H,EZH2:NM_001203248:exon6:c.G526C:p.D176H,EZH2:NM_001203249:exon6:c.G526C:p.D176H,EZH2:NM_004456:exon6:c.G553C:p.D185H	17	0.35

chr22	22893809	22893809	A	T	<i>IGLL5</i>	Non-synonymous SNV	IGLL5:NM_001178126:exon2:c.A316T:p.T106S	112	0.35
chr2	46145188	46145188	C	T	<i>PRKCE</i>	Non-synonymous SNV	PRKCE:NM_005400:exon12:c.C1688T:p.T563M	13	0.31
chr11	108304735	108304735	G	A	<i>ATM</i>	Non-synonymous SNV	ATM:NM_000051:exon37:c.G5557A:p.D1853N	11	0.27
chr17	16146494	16146494	-	T	<i>NCOR1</i>	frameshift insertion	NCOR1:NM_001190438:exon7:c.636dupA:p.V213fs, NCOR1:NM_001190440:exon9:c.963dupA:p.V322fs, NCOR1:NM_006311:exon10:c.963dupA:p.V322fs	35	0.26
chr16	2058786	2058786	-	T	<i>TSC2</i>	frameshift insertion	TSC2:NM_000548:exon10:c.889dupT:p.V296fs, TSC2:NM_001077183:exon10:c.889dupT:p.V296fs, TSC2:NM_001114382:exon10:c.889dupT:p.V296fs	21	0.24
chr1	6197724	6197724	-	T	<i>RPL22</i>	frameshift insertion	RPL22:NM_000983:exon2:c.44dupA:p.K15fs	2957	0.22
chr12	6601979	6601981	TCC	-	<i>CHD4</i>	Non-frameshift deletion	CHD4:NM_001297553:exon3:c.396_398del:p.132_133del, CHD4:NM_001273:exon4:c.417_419del:p.139_140del	64	0.20
chr22	22893804	22893804	A	C	<i>IGLL5</i>	Non-synonymous SNV	IGLL5:NM_001178126:exon2:c.A311C:p.K104T	105	0.19
chr4	24548878	24548878	-	T	<i>DHX15</i>	frameshift insertion	DHX15:NM_001358:exon6:c.1224dupA:p.Q409fs	215	0.12

Table S2.23: All Cancer Associated SNVs Called for Case 66. Variants were filtered based on a list of 383 pre-defined cancer genes. The variants were maintained if they were exonic, had a damaging function, previously reported in COSMIC, and were considered pathogenic in the ClinVar Database. Variants were then excluded if they had a variant allele frequency below 0.05, depth below 10, a prevalence of greater than 10% in the 1000 Genomes Project, or if no annotations were present. SNVs are shaded dark orange if the variant allele frequency (VAF) =1, and lighter shades of orange for VAFs >=0.5 and <1 and VAFs <0.5 and >0.4. The 'AA Change' column lists the ref.seq. number, the base pair alteration, and the amino acid alteration.

Case 67	Chromosome	Start	End	Reference	Alteration	Gene	Exonic Function	AA Change	Read Depth	VAF
	chr12	6667904	6667906	TGC	-	ZNF384	Non-frameshift deletion	ZNF384:NM_001039920:exon9:c.1194_1196del:p.398_399del,ZNF384:NM_133476:exon10:c.1359_1361del:p.453_454del,ZNF384:NM_001135734:exon11:c.1542_1544del:p.514_515del	21	1.00
	chr2	29193500	29193500	G	C	ALK	Non-synonymous SNV	ALK:NM_004304:exon29:c.C4587G;p.D1529E	36	1.00
	chr4	849720	849720	C	T	GAK	Non-synonymous SNV	GAK:NM_001286833:exon25:c.G3595A;p.D1199N,GAK:NM_005255:exon28:c.G3889A;p.D1297N	210	1.00
	chr5	35860966	35860966	T	C	IL7R	Non-synonymous SNV	IL7R:NM_002185:exon2:c.T197C;p.I66T	179	1.00
	chr5	35871088	35871088	G	A	IL7R	Non-synonymous SNV	IL7R:NM_002185:exon4:c.G412A;p.V138I	151	1.00
	chr3	89472514	89472514	G	A	EPHA3	Non-synonymous SNV	EPHA3:NM_005233:exon16:c.G2741A;p.R914H	10	0.80
	chr9	132595027	132595027	T	C	DDX31	Non-synonymous SNV	DDX31:NM_022779:exon20:c.A2395G;p.I799V	14	0.64
chr4	105275794	105275794	A	G	TET2	Non-synonymous SNV	TET2:NM_001127208:exon11:c.A5284G;p.I1762V	15	0.60	

chr20	56370253	56370253	T	C	AURKA	Non-synonymous SNV	AURKA:NM_003600:exon9:c.A1117G:p.M373V,AURKA:NM_198435:exon9:c.A1117G:p.M373V,AURKA:NM_198437:exon9:c.A1117G:p.M373V,AURKA:NM_198434:exon10:c.A1117G:p.M373V,AURKA:NM_198436:exon10:c.A1117G:p.M373V,AURKA:NM_198433:exon11:c.A1117G:p.M373V	24	0.58
chr7	152162598	152162598	G	A	KMT2C	Non-synonymous SNV	KMT2C:NM_170606:exon43:c.C10979T:p.S3660L	31	0.58
chr1	81990901	81990901	G	A	ADGRL2	Non-synonymous SNV	ADGRL2:NM_012302:exon20:c.G3968A:p.R1323K,ADGRL2:NM_001297704:exon21:c.G3968A:p.R1323K	79	0.56
chr3	47083895	47083895	G	A	SETD2	Non-synonymous SNV	SETD2:NM_014159:exon12:c.C5885T:p.P1962L	53	0.55
chr20	53581599	53581599	T	C	ZNF217	Non-synonymous SNV	ZNF217:NM_006526:exon1:c.A1228G:p.M410V	11	0.55
chr18	12817349	12817349	G	T	PTPN2	Non-synonymous SNV	PTPN2:NM_001308287:exon5:c.C425A:p.T142K,PTPN2:NM_002828:exon6:c.C512A:p.T171K,PTPN2:NM_080422:exon6:c.C512A:p.T171K,PTPN2:NM_080423:exon6:c.C512A:p.T171K,PTPN2:NM_001207013:exon7:c.C581A:p.T194K	177	0.51

chr7	148828812	148828812	C	G	<i>EZH2</i>	Non-synonymous SNV	EZH2:NM_152998:exon5:c.G436C;p.D146H,EZH2:NM_001203247:exon6:c.G553C;p.D185H,EZH2:NM_001203248:exon6:c.G526C;p.D176H,EZH2:NM_001203249:exon6:c.G526C;p.D176H,EZH2:NM_004456:exon6:c.G553C;p.D185H	41	0.51
chr19	34466925	34466925	C	T	<i>UBA2</i>	Non-synonymous SNV	UBA2:NM_005499:exon16:c.C1652T;p.P551L	144	0.51
chr1	9717541	9717541	C	G	<i>PIK3CD</i>	Non-synonymous SNV	PIK3CD:NM_005026:exon8:c.C935G;p.S312C	13	0.46
chr17	50188134	50188134	C	T	<i>COL1A1</i>	Non-synonymous SNV	COL1A1:NM_000088:exon44:c.G3223A;p.A1075T	1591	0.45
chr5	180603322	180603322	C	T	<i>FLT4</i>	Non-synonymous SNV	FLT4:NM_182925:exon30:c.G3962A;p.R1321Q	68	0.43
chr1	16969820	16969820	G	C	<i>CROCC</i>	Non-synonymous SNV	CROCC:NM_014675:exon33:c.G5337C;p.E1779D	40	0.43
chr16	67636779	67636779	C	T	<i>CTCF</i>	Non-synonymous SNV	CTCF:NM_001191022:exon9:c.C943T;p.P315S,CTCF:NM_006565:exon11:c.C1927T;p.P643S	40	0.43
chr21	38383690	38383690	G	A	<i>ERG</i>	Non-synonymous SNV	ERG:NM_001136155:exon9:c.C877T;p.R293C,ERG:NM_001243429:exon9:c.C805T;p.R269C,ERG:NM_182918:exon10:c.C1153T;p.R385C,ERG:NM_004449:exon11:c.C1102T;p.R368C,ERG:NM_001136154:exon12:c.C1174T;p.R392C,ERG:NM_001243428:exon12:c.C1174T;p.R392	17	0.41

							C		
chr12	111447491	111447491	G	A	<i>SH2B3</i>	Non-synonymous SNV	SH2B3:NM_001291424:exon5:c.G577A:p.E193K,S H2B3:NM_005475:exon6: c.G1183A:p.E395K	44	0.41
chr17	7676154	7676154	G	C	<i>TP53</i>	Non-synonymous SNV	TP53:NM_001126118:exon3:c.C98G:p.P33R,TP53: NM_000546:exon4:c.C215G:p.P72R,TP53:NM_001126112:exon4:c.C215G:p.P72R,TP53:NM_001126113:exon4:c.C215G:p.P72R,TP53:NM_001126114:exon4:c.C215G:p.P72R,TP53:NM_001276695:exon4:c.C98G:p.P33R,TP53:NM_001276696:exon4:c.C98G:p.P33R,TP53:NM_001276760:exon4:c.C98G:p.P33R,TP53:NM_001276761:exon4:c.C98G:p.P33R	22	0.41
chrX	154400814	154400814	G	A	<i>RPL10</i>	Non-synonymous SNV	RPL10:NM_001256577:exon6:c.G442A:p.V148I,RPL10:NM_001256580:exon6:c.G497A:p.S166N,RPL10:NM_001303624:exon6:c.G605A:p.S202N,RPL10:NM_001303625:exon7:c.G605A:p.S202N,RPL10:NM_006013:exon7:c.G605A:p.S202N	2283	0.40

chr3	49360990	49360990	T	C	<i>RHOA</i>	Non-synonymous SNV	RHOA:NM_001313943:exon5:c.A485G:p.K162R	25	0.40
chr7	152180875	152180875	C	T	<i>KMT2C</i>	Non-synonymous SNV	KMT2C:NM_170606:exon36:c.G6985A:p.G2329R	30	0.40
chr20	53576098	53576098	T	C	<i>ZNF217</i>	Non-synonymous SNV	ZNF217:NM_006526:exon3:c.A2666G:p.D889G	28	0.39
chr11	118505199	118505199	-	A	<i>KMT2A</i>	frameshift insertion	KMT2A:NM_001197104:exon27:c.9308dupA:p.Q3103fs,KMT2A:NM_005933:exon27:c.9299dupA:p.Q3100fs	25	0.28
chr1	6197724	6197724	-	T	<i>RPL22</i>	frameshift insertion	RPL22:NM_000983:exon2:c.44dupA:p.K15fs	2036	0.22
chr3	49360949	49360949	T	C	<i>RHOA</i>	Non-synonymous SNV	RHOA:NM_001313943:exon5:c.A526G:p.R176G	14	0.21
chr9	37357252	37357252	-	A	<i>ZCCHC7</i>	frameshift insertion	ZCCHC7:NM_001289119:exon9:c.1617dupA:p.R539fs,ZCCHC7:NM_001289120:exon9:c.1617dupA:p.R539fs,ZCCHC7:NM_001289121:exon9:c.1617dupA:p.R539fs,ZCCHC7:NM_032226:exon9:c.1617dupA:p.R539fs	38	0.21
chr17	16146494	16146494	-	T	<i>NCOR1</i>	frameshift insertion	NCOR1:NM_001190438:exon7:c.636dupA:p.V213fs,NCOR1:NM_001190440:exon9:c.963dupA:p.V322fs,NCOR1:NM_006311:exon10:c.963dupA:p.V322fs	53	0.17
chr16	2058786	2058786	-	T	<i>TSC2</i>	frameshift insertion	TSC2:NM_000548:exon10:c.889dupT:p.V296fs,TSC2:NM_001077183:exon10:c.889dupT:p.V296fs,TSC2:NM_001114382:exon10:c	30	0.17

							.889dupT:p.V296fs		
chrX	77682737	77682737	-	T	<i>ATRX</i>	frameshift insertion	ATRX:NM_138270:exon8:c.2404dupA:p.R802fs,ATRX:NM_000489:exon9:c.2518dupA:p.R840fs	54	0.17
chr4	24548878	24548878	-	T	<i>DHX15</i>	frameshift insertion	DHX15:NM_001358:exon6:c.1224dupA:p.Q409fs	104	0.13

Table S2.24: All Cancer Associated SNVs Called for Case 67. Variants were filtered based on a list of 383 pre-defined cancer genes. The variants were maintained if they were exonic, had a damaging function, previously reported in COSMIC, and were considered pathogenic in the ClinVar Database. Variants were then excluded if they had a variant allele frequency below 0.05, depth below 10, a prevalence of greater than 10% in the 1000 Genomes Project, or if no annotations were present. SNVs are shaded dark orange if the variant allele frequency (VAF) =1, and lighter shades of orange for VAFs \geq .5 and $<$ 1 and VAFs $<$.5 and $>$.4. The 'AA Change' column lists the ref.seq. number, the base pair alteration, and the amino acid alteration.

Case 69	Chromosome	Start	End	Reference	Alteration	Gene	Exonic Function	AA Change	Read Depth	VAF
	chr12	6667904	6667906	TGC	-	ZNF384	Non-frameshift deletion	ZNF384:NM_001039920:exon9:c.1194_1196del:p.398_399del,ZNF384:NM_133476:exon10:c.1359_1361del:p.453_454del,ZNF384:NM_001135734:exon11:c.1542_1544del:p.514_515del	17	1.00
	chr20	1915303	1915303	-	GT	SIRPA	frameshift insertion	SIRPA:NM_001040023:exon2:c.284_285insGT:p.D95fs,SIRPA:NM_001040022:exon3:c.284_285insGT:p.D95fs,SIRPA:NM_080792:exon3:c.284_285insGT;p.D95fs	54	1.00
	chr20	1915305	1915306	CT	-	SIRPA	frameshift deletion	SIRPA:NM_001040023:exon2:c.286_287del:p.L96fs,SIRPA:NM_001040022:exon3:c.286_287del:p.L96fs,SIRPA:NM_080792:exon3:c.286_287del:p.L96fs	54	1.00
	chr20	1915319	1915319	C	A	SIRPA	Non-synonymous SNV	SIRPA:NM_001040023:exon2:c.C300A:p.N100K,SIRPA:NM_001040022:exon3:c.C300A:p.N100K,SIRPA:NM_080792:exon3:c.C300A:p.N100K	50	1.00
chr20	1915338	1915338	C	A	SIRPA	Non-synonymous SNV	SIRPA:NM_001040023:exon2:c.C319A:p.R107S,SIRPA:NM_001040022:exon3:c.C319A:p.R107S,SIRPA:NM_080792:exon3:c.C319A:p.R107S	42	1.00	

chr20	1915344	1915344	G	A	<i>SIRPA</i>	Non-synonymous SNV	SIRPA:NM_001040023:exon2:c.G325A:p.G109S,SIRPA:NM_001040022:exon3:c.G325A:p.G109S,SIRPA:NM_080792:exon3:c.G325A:p.G109S	37	1.00
chr20	1915406	1915408	CGA	-	<i>SIRPA</i>	Non-frameshift deletion	SIRPA:NM_001040023:exon2:c.387_389del:p.129_130del,SIRPA:NM_001040022:exon3:c.387_389del:p.129_130del,SIRPA:NM_080792:exon3:c.387_389del:p.129_130del	32	1.00
chrX	77682471	77682471	C	G	<i>ATRX</i>	Non-synonymous SNV	ATRX:NM_138270:exon8:c.G2671C:p.E891Q,ATRX:NM_000489:exon9:c.G2785C:p.E929Q	33	1.00
chr3	89472514	89472514	G	A	<i>EPHA3</i>	Non-synonymous SNV	EPHA3:NM_005233:exon16:c.G2741A:p.R914H	26	0.96
chr7	152162598	152162598	G	A	<i>KMT2C</i>	Non-synonymous SNV	KMT2C:NM_170606:exon43:c.C10979T:p.S3660L	17	0.65
chr15	41736622	41736622	A	G	<i>MGA</i>	Non-synonymous SNV	MGA:NM_001080541:exon13:c.A4358G:p.Y1453C,MGA:NM_001164273:exon13:c.A4358G:p.Y1453C	11	0.64
chr2	25750067	25750067	C	T	<i>ASXL2</i>	Non-synonymous SNV	ASXL2:NM_018263:exon11:c.G1489A:p.A497T	11	0.64
chr3	47121396	47121396	C	T	<i>SETD2</i>	Non-synonymous SNV	SETD2:NM_014159:exon3:c.G3240A:p.M1080I	31	0.61
chr9	136523857	136523857	C	T	<i>NOTCH1</i>	Non-synonymous SNV	NOTCH1:NM_017617:exon3:c.G263A:p.S88N	12	0.58
chr16	72788746	72788746	G	A	<i>ZFH3</i>	Non-synonymous SNV	ZFH3:NM_001164766:exon9:c.C6788T:p.S2263L,ZFH3:NM_006885:exon10:c.C9530T:p.S3177L	16	0.56

chr2	43225479	43225479	C	T	<i>ZFP36L2</i>	Non-synonymous SNV	ZFP36L2:NM_006887:exon2:c.G325A:p.G109S	241	0.53
chr3	66380453	66380453	G	C	<i>LRIG1</i>	Non-synonymous SNV	LRIG1:NM_015541:exon19:c.C3092G:p.P1031R	97	0.53
chr3	66380763	66380763	C	T	<i>LRIG1</i>	Non-synonymous SNV	LRIG1:NM_015541:exon18:c.G2869A:p.A957T	74	0.50
chr22	22895680	22895680	A	G	<i>IGLL5</i>	Non-synonymous SNV	IGLL5:NM_001256296:exon2:c.A406G:p.T136A,IGLL5:NM_001178126:exon3:c.A631G:p.T211A	763	0.50
chr1	16130268	16130268	C	T	<i>EPHA2</i>	Non-synonymous SNV	EPHA2:NM_004431:exon15:c.G2627A:p.R876H	57	0.49
chr4	849720	849720	C	T	<i>GAK</i>	Non-synonymous SNV	GAK:NM_001286833:exon25:c.G3595A:p.D1199N,GAK:NM_005255:exon28:c.G3889A:p.D1297N	116	0.47
chr9	132911517	132911517	A	G	<i>TSC1</i>	Non-synonymous SNV	TSC1:NM_001162427:exon9:c.T812C:p.M271T,TSC1:NM_000368:exon10:c.T965C:p.M322T,TSC1:NM_001162426:exon10:c.T965C:p.M322T	43	0.47
chr4	108048727	108048727	G	A	<i>LEF1</i>	Non-synonymous SNV	LEF1:NM_001130714:exon10:c.C1106T:p.T369M	69	0.46
chr11	3679646	3679646	C	T	<i>NUP98</i>	Non-synonymous SNV	NUP98:NM_139132:exon30:c.G4759A:p.E1587K,NUP98:NM_016320:exon31:c.G4981A:p.E1661K	131	0.45
chr17	16061887	16061887	G	A	<i>NCOR1</i>	Non-synonymous SNV	NCOR1:NM_001190440:exon36:c.C5443T:p.P1815S,NCOR1:NM_006311:exon37:c.C5395T:p.P1799S	49	0.45
chr19	15161407	15161407	G	A	<i>NOTCH3</i>	Non-synonymous SNV	NOTCH3:NM_000435:exon33:c.C6221T:p.P2074L	29	0.45

chr4	849932	849932	T	C	<i>GAK</i>	Non-synonymous SNV	GAK:NM_001286833:exon24:c.A3500G:p.K1167R, GAK:NM_005255:exon27:c.A3794G:p.K1265R	136	0.44
chr4	105275672	105275672	T	G	<i>TET2</i>	Non-synonymous SNV	TET2:NM_001127208:exon11:c.T5162G:p.L1721W	16	0.44
chr5	35860966	35860966	T	C	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon2:c.T197C:p.I66T	277	0.42
chr17	7676154	7676154	G	C	<i>TP53</i>	Non-synonymous SNV	TP53:NM_001126118:exon3:c.C98G:p.P33R, TP53:NM_000546:exon4:c.C215G:p.P72R, TP53:NM_001126112:exon4:c.C215G:p.P72R, TP53:NM_001126113:exon4:c.C215G:p.P72R, TP53:NM_001126114:exon4:c.C215G:p.P72R, TP53:NM_001276695:exon4:c.C98G:p.P33R, TP53:NM_001276696:exon4:c.C98G:p.P33R, TP53:NM_001276760:exon4:c.C98G:p.P33R, TP53:NM_001276761:exon4:c.C98G:p.P33R	61	0.41
chr19	41352971	41352971	C	G	<i>TGFB1</i>	Non-synonymous SNV	TGFB1:NM_000660:exon1:c.G74C:p.R25P	48	0.40
chr7	152162938	152162938	A	G	<i>KMT2C</i>	Non-synonymous SNV	KMT2C:NM_170606:exon43:c.T10639C:p.S3547P	18	0.39
chr1	16938425	16938425	C	T	<i>CROCC</i>	Non-synonymous SNV	CROCC:NM_014675:exon11:c.C1316T:p.A439V	16	0.38

chr20	32795432	32795432	G	A	<i>DNMT3B</i>	Non-synonymous SNV	DNMT3B:NM_001207056:exon8:c.G862A:p.A288T, DNMT3B:NM_001207055:exon9:c.G964A:p.A322T, DNMT3B:NM_175848:exon10:c.G1090A:p.A364T, DNMT3B:NM_175849:exon10:c.G1090A:p.A364T, DNMT3B:NM_175850:exon10:c.G1126A:p.A376T, DNMT3B:NM_006892:exon11:c.G1150A:p.A384T	46	0.37
chr22	22888120	22888120	C	T	<i>IGLL5</i>	Non-synonymous SNV	IGLL5:NM_001178126:exon1:c.C67T:p.R23C	11	0.36
chr11	118505199	118505199	-	A	<i>KMT2A</i>	frameshift insertion	KMT2A:NM_001197104:exon27:c.9308dupA:p.Q3103fs, KMT2A:NM_005933:exon27:c.9299dupA:p.Q3100fs	17	0.35
chr5	35871088	35871088	G	A	<i>IL7R</i>	Non-synonymous SNV	IL7R:NM_002185:exon4:c.G412A:p.V138I	206	0.34
chr1	26773690	26773690	-	GCA	<i>ARID1A</i>	Non-frameshift insertion	ARID1A:NM_006015:exon16:c.3977_3978insGCA:p.P1326delinsPQ, ARID1A:NM_139135:exon16:c.3977_3978insGCA:p.P1326delinsPQ	14	0.29
chr7	148828812	148828812	C	G	<i>EZH2</i>	Non-synonymous SNV	EZH2:NM_152998:exon5:c.G436C:p.D146H, EZH2:NM_001203247:exon6:c.G553C:p.D185H, EZH2:NM_001203248:exon6:c.G526C:p.D176H, EZH2:NM_001203249:exon6:c.G526C:p.D176H, EZH2:NM_004456:exon6:c.G553C:p.D185H	14	0.29

chr1	16946274	16946274	G	A	<i>CROCC</i>	Non-synonymous SNV	CROCC:NM_014675:exon16:c.G2152A:p.V718M	11	0.27
chr1	16946292	16946292	A	G	<i>CROCC</i>	Non-synonymous SNV	CROCC:NM_014675:exon16:c.A2170G:p.M724V	13	0.23
chr11	108293436	108293436	-	A	<i>ATM</i>	frameshift insertion	ATM:NM_000051:exon31:c.4736dupA:p.Q1579fs	26	0.23
chr1	6197724	6197724	-	T	<i>RPL22</i>	frameshift insertion	RPL22:NM_000983:exon2:c.44dupA:p.K15fs	2179	0.20
chr4	105234028	105234028	C	G	<i>TET2</i>	Non-synonymous SNV	TET2:NM_001127208:exon3:c.C86G:p.P29R,TET2:NM_017628:exon3:c.C86G:p.P29R	16	0.19
chr9	132595027	132595027	T	C	<i>DDX31</i>	Non-synonymous SNV	DDX31:NM_022779:exon20:c.A2395G:p.I799V	29	0.17
chr3	49360949	49360949	T	C	<i>RHOA</i>	Non-synonymous SNV	RHOA:NM_001313943:exon5:c.A526G:p.R176G	19	0.16
chr3	49360951	49360951	T	C	<i>RHOA</i>	Non-synonymous SNV	RHOA:NM_001313943:exon5:c.A524G:p.Y175C	19	0.16
chr19	10151462	10151462	-	T	<i>DNMT1</i>	frameshift insertion	DNMT1:NM_001379:exon23:c.2152dupA:p.M718fs, DNMT1:NM_001130823:exon24:c.2200dupA:p.M734fs	73	0.15
chr4	24548878	24548878	-	T	<i>DHX15</i>	frameshift insertion	DHX15:NM_001358:exon6:c.1224dupA:p.Q409fs	166	0.14
chr11	3699158	3699158	-	T	<i>NUP98</i>	frameshift insertion	NUP98:NM_016320:exon25:c.3932dupA:p.N1311fs, NUP98:NM_139132:exon25:c.3932dupA:p.N1311fs	98	0.13
chr12	6601979	6601981	TCC	-	<i>CHD4</i>	Non-frameshift deletion	CHD4:NM_001297553:exon3:c.396_398del:p.132_133del,CHD4:NM_001273:exon4:c.417_419del:p.139	51	0.12

							_140del		
chr5	35867438	35867438	-	A	<i>IL7R</i>	frameshift insertion	IL7R:NM_002185:exon3:c.355dupA:p.C118fs	238	0.11

Table S2.25: All Cancer Associated SNVs Called for Case 69. Variants were filtered based on a list of 383 pre-defined cancer genes. The variants were maintained if they were exonic, had a damaging function, previously reported in COSMIC, and were considered pathogenic in the ClinVar Database. Variants were then excluded if they had a variant allele frequency below 0.05, depth below 10, a prevalence of greater than 10% in the 1000 Genomes Project, or if no annotations were present. SNVs are shaded dark orange if the variant allele frequency (VAF) =1, and lighter shades of orange for VAFs >=0.5 and <1 and VAFs <0.5 and >0.4. The 'AA Change' column lists the ref.seq. number, the base pair alteration, and the amino acid alteration.

Category	Term	Count	P Value	Fold Enrichment	FDR
Reactome	VEGF ligand-receptor interactions (R-HSA-194313)	2	9.36E-05	> 100	4.65E-02
Reactome	VEGF binds to VEGFR leading to receptor dimerization (R-HSA-195399)	2	9.36E-05	> 100	3.72E-02
Reactome	PKMTs methylate histone lysines (R-HSA-3214841)	3	2.74E-05	55.04	1.82E-02
Reactome	Chromatin modifying enzymes (R-HSA-3247509)	5	1.96E-05	15.15	3.90E-02
Reactome	Chromatin organization (R-HSA-4839726)	5	1.96E-05	15.15	1.95E-02
Molecular Function	vascular endothelial growth factor-activated receptor activity (GO:0005021)	2	7.49E-05	> 100	4.36E-02
Molecular Function	histone-lysine N-methyltransferase activity (GO:0018024)	3	5.39E-05	43.33	5.02E-02
Molecular Function	helicase activity (GO:0004386)	4	5.86E-05	19.26	4.54E-02

Table S2.26: Gene Ontology Terms for Commonly Shared SNVs. Panther Gene Ontology Term Enrichment analysis was performed on the set of 30 genes which harbored the most commonly shared SNVs amongst all 18 samples evaluated by RNAseq. The 'category' column references the type of test performed. The 'count' column displays the number of genes associated with the term that were in the set of 30 evaluated. Statistics are calculated based on the count and the expected count.

#	Official Ab Name	Ab Name Reported on	Gene Name	Company	Catalog #	Species	Validation Status*	RPPA Dilution
		Dataset						
1	14-3-3 beta	14-3-3-beta	YWHAB	Santa Cruz	sc-628	Rabbit	Valid	1:75
2	14-3-3 zeta	14-3-3-zeta	YWHAZ	Santa Cruz	sc-1019	Rabbit	Valid	1:5000
3	4E-BP1	4E-BP1	EIF4EBP1	CST	9452	Rabbit	Valid	1:100
4	4E-BP1 (phospho S65)	4E-BP1_pS65	EIF4EBP1	CST	9456	Rabbit	Valid	1:250
5	53BP1	53BP1	TP53BP1	CST	4937	Rabbit	Valid	1:300
6	Acetyl CoA Carboxylase (phospho S79)	ACC_pS79	ACACA, ACACB	CST	3661	Rabbit	Valid	1:500
7	Acetyl CoA Carboxylase 1	ACC1	ACACA	Abcam	ab45174	Rabbit	Use with Caution	1:20000
8	ADAR1	ADAR1	ADAR	Abcam	ab88574	Mouse	Valid	1:200
9	Akt	Akt	AKT1,2,3	CST	4691	Rabbit	Valid	1:10000
10	Akt (phospho S473)	Akt_pS473	AKT1,2,3	CST	9271	Rabbit	Valid	1:150
11	Akt (phospho T308)	Akt_pT308	AKT1,2,3	CST	2965	Rabbit	Valid	1:500
12	AMPK alpha	AMPKa	PRKAA1	CST	2532	Rabbit	Use with Caution	1:200
13	AMPK alpha (phospho T172)	AMPKa_pT172	PRKAA1	CST	2535	Rabbit	Use with Caution	1:100
14	AMPK alpha 2 (Phospho S345)	AMPK-a2_pS345	PRKAA2	Abcam	ab129081	Rabbit	Valid	1:500
15	Androgen Receptor	AR	AR	CST	5153	Rabbit	Valid	1:250
16	Annexin I	Annexin-I	ANXA1	BD Biosciences	610066	Mouse	Valid	1:5000
17	Annexin VII	Annexin-VII	ANXA7	BD Biosciences	610668	Mouse	Valid	1:30
18	A-Raf	A-Raf	ARAF	CST	4432	Rabbit	Valid	1:150
19	ARID1A	ARID1A	ARID1A	Sigma-Aldrich	HPA005456	Rabbit	Use with Caution	1:1000
20	Atg3	Atg3	ATG3	CST	3415	Rabbit	Valid	1:750
21	Atg7	Atg7	ATG7	CST	8558	Rabbit	Valid	1:1000
22	ATM	ATM	ATM	CST	2873	Rabbit	Valid	1:250
23	ATM(phospho S1981)	ATM_pS1981	ATM	CST	5883	Rabbit	Valid	1:25
24	ATR (Phospho S428)	ATR_pS428	ATR	Abcam	ab178407	Rabbit	Use with Caution	1:1000
25	Aurora B/AIM1	Aurora-B	AIM1	CST	3094	Rabbit	Valid	1:50
26	Axl	Axl	AXL	CST	8661	Rabbit	Valid	1:1000
27	B7-H4	B7-H4	VTCN1	CST	14572	Rabbit	Use with Caution	1:50
28	Bad (phospho S112)	Bad_pS112	BAD	CST	9291	Rabbit	Valid	1:50
29	Bak	Bak	BAK1	Abcam	ab32371	Rabbit	Use with Caution	1:30

30	BAP1	BAP1	BAP1	Santa Cruz	sc-28383	Mouse	Valid	1:125
31	Bax	Bax	BAX	CST	2772	Rabbit	Valid	1:100
32	Bcl2	Bcl2	BCL2	Dako	M0887	Mouse	Valid	1:50
33	Bcl-xL	Bcl-xL	BCL2L1	CST	2762	Rabbit	Valid	1:100
34	Beclin	Beclin	BECN1	Santa Cruz	sc-10086	Goat	Use with Caution	1:250
35	beta Actin	b-Actin	ACTB	CST	4970	Rabbit	Use with Caution	1:75
36	beta Catenin	b-Catenin	CTNNB1	CST	9562	Rabbit	Valid	1:1500
37	beta Catenin (phospho T41/S45)	b-Catenin_pT41_S45	CTNNB1	CST	9565	Rabbit	Valid	1:30
38	Bid	Bid	BID	CST	2002	Rabbit	Use with Caution	1:1000
39	Bim	Bim	BCL2L11	Abcam	ab32158	Rabbit	Valid	1:400
40	BiP/GRP78	BiP-GRP78	HSPA5	BD Biosciences	610978	Mouse	Use with Caution	1:750
41	B-Raf	B-Raf	BRAF	CST	14814	Rabbit	Use with Caution	1:500
42	B-Raf (phospho S445)	B-Raf_pS445	BRAF	CST	2696	Rabbit	Valid	1:1000
43	BRD4	BRD4	BRD4	CST	13440	Rabbit	Valid	1:200
44	c-Abl	c-Abl	ABL	CST	2862	Rabbit	Valid	1:50
45	c-IAP2	c-IAP2	BIRC3	CST	3130	Rabbit	Use with Caution	1:750
46	Caspase-3 active	Caspase-3	CASP3	Abcam	ab32042	Rabbit	Use with Caution	1:250
47	Caspase-7 (cleaved D198)	Caspase-7-cleaved	CASP7	CST	9491	Rabbit	Use with Caution	1:75
48	Caspase-8	Caspase-8	CASP8	CST	9746	Mouse	**Used for QC**	1:150
49	Caveolin-1	Caveolin-1	CAV1	CST	3238	Rabbit	Valid	1:5000
50	CD4	CD4	CD4	Abcam	ab133616	Rabbit	Valid	1:500
51	CD20	CD20	MS4A1	Abcam	ab78237	Rabbit	Use with Caution	1:75
52	CD29	CD29	ITGB1	BD Biosciences	610467	Mouse	Valid	1:30
53	CD31	CD31	PECAM1	Dako	M0823	Mouse	Valid	1:30
54	CD44	CD44	CD44	CST	3570	Mouse	Use with Caution	1:50
55	CD45	CD45	CD45	DAKO	M070129-2	Mouse	Valid	1:1000
56	CD49b	CD49b	ITGA2	BD Biosciences	611016	Mouse	Valid	1:50
57	CD134/OX40	CD134	CD134	Abcam	ab76000	Rabbit	Valid	1:100
58	cdc2 (Phospho Y15)	cdc2_pY15	CDK1	CST	4539	Rabbit	Use with Caution	1:250

59	cdc25C	cdc25C	CDC25C	CST	4688	Rabbit	Valid	1:500
60	CDK1	CDK1	CDK1	Abcam	ab32384	Rabbit	Use with Caution	1:1000
61	CDKN2A/p16INK4a	p16INK4a	CDKN2A	Abcam	ab81278	Rabbit	Valid	1:500
62	Chk1	Chk1	CHEK1	CST	2360	Mouse	Use with Caution	1:250
63	Chk1 (phospho S296)	Chk1_pS296	CHEK1	Abcam	ab79758	Rabbit	Valid	1:125
64	Chk2	Chk2	CHEK2	CST	3440	Mouse	Valid	1:50
65	Chk2 (phospho T68)	Chk2_pT68	CHEK2	CST	2197	Rabbit	Use with Caution	1:125
66	c-Jun (phospho S73)	c-Jun_pS73	JUN	CST	9164	Rabbit	Valid	1:30
67	c-Kit	c-Kit	KIT	Abcam	ab32363	Rabbit	Valid	1:30
68	Claudin 7	Claudin-7	CLDN7	Novus Biologicals	NB100-91714	Rabbit	Valid	1:300
69	c-Met (phospho Y1234/Y1235)	c-Met_pY1234_Y1235	MET	CST	3129	Rabbit	Valid	1:100
70	c-Myc	c-Myc	MYC	Santa Cruz	sc-764	Rabbit	Use with Caution	1:125
71	COG3	COG3	COG3	ProteinTech	11130-1-AP	Rabbit	Valid	1:750
72	COL6A1	Collagen-VI	COL6A1	Santa Cruz	sc-20649	Rabbit	Valid	1:5000
73	Connexin 43	Connexin-43	CNST43	CST	3512	Rabbit	Use with Caution	1:150
74	Cox2	Cox2	PTGS2	CST	4842	Rabbit	Use with Caution	1:50
75	Cox-IV	Cox-IV	PTGS3	CST	4850	Rabbit	Valid	1:1000
76	C-Raf (phospho S338)	C-Raf_pS338	RAF1	CST	9427	Rabbit	Valid	1:100
77	C-Raf/Raf-1	C-Raf	RAF1	Millipore	04-739	Rabbit	Use with Caution	1:200
78	CREB	CREB	CREB1	CST	9197	Rabbit	Use with Caution	1:1000
79	Cyclin B1	Cyclin-B1	CCNB1	Epitomics	1495-1	Rabbit	Valid	1:1500
80	Cyclin D1	Cyclin-D1	CCND1	Santa Cruz	sc-718	Rabbit	Valid	1:200
81	Cyclin D3	Cyclin-D3	CCND3	CST	2936	Mouse	Valid	1:1000
82	Cyclin E1	Cyclin-E1	CCNE1	Santa Cruz	sc-247	Mouse	Valid	1:30
83	Cyclophilin F	Cyclophilin-F	PPIF	Abcam	ab110324	Mouse	Valid	1:50000
84	Detyrosinated alpha-Tubulin	D-a-Tubulin	TUBA1A	Abcam	ab48389	Rabbit	Valid	1:500
85	Dimethyl-Histone H3 (Lys4)	DM-Histone-H3	HIST1H3A	CST	9725	Rabbit	Valid	1:100
86	Dimethyl-K9 Histone H3	DM-K9-Histone-H3	H3K9ME2	Abcam	ab32521	Rabbit	Use with Caution	1:250
87	DUSP4/MKP2	DUSP4	DUSP4	CST	5149	Rabbit	Valid	1:250
88	E2F-1	E2F1	E2F1	Santa Cruz	sc-251	Mouse	Valid	1:30

89	E-Cadherin	E-Cadherin	CDH1	CST	3195	Rabbit	Valid	1:300
90	eEF2	eEF2	EEF2	CST	2332	Rabbit	Use with Caution	1:50
91	eEF2K	eEF2K	EEF2K	CST	3692	Rabbit	Valid	1:50
92	EGFR	EGFR	EGFR	CST	2232	Rabbit	Valid	1:100
93	EGFR (phospho Y1173)	EGFR_pY1173	EGFR	Abcam	ab32578	Rabbit	Valid	1:50
94	eIF4E	eIF4E	EIF4E	CST	9742	Rabbit	Valid	1:75
95	eIF4E (Phospho S209)	eIF4E_pS209	EIF4E	Abcam	ab76256	Rabbit	Valid	1:500
96	eIF4G	eIF4G	EIF4G1	CST	2498	Rabbit	Use with Caution	1:1000
97	Elk1 (phospho S383)	Elk1_pS383	ELK1	CST	9181	Rabbit	Use with Caution	1:50
98	ENY2	ENY2	ENY2	GeneTex	GTX629542	Mouse	Use with Caution	1:1000
99	Epithelial Membrane Antigen	EMA	EMA	Dako	M061329-2	Mouse	Use with Caution	1:1000
100	ErbB2/HER2	HER2	ERBB2	Lab Vision	MS-325-P1	Mouse	Valid	1:3000
101	ErbB2/HER2 (phosphoY1248)	HER2_pY1248	ERBB2	R&D Systems	AF1768	Rabbit	Use with Caution (likely sees pEGFR)	1:1500
102	ErbB3/HER3	HER3	ERBB3	Santa Cruz	sc-285	Rabbit	Valid	1:300
103	ErbB3/HER3 (phosphoY1289)	HER3_pY1289	ERBB3	CST	4791	Rabbit	Use with Caution	1:50
104	ERCC1	ERCC1	ERCC1	Santa Cruz	sc-17809	Mouse	Valid	1:30
105	ERCC5	ERCC5	ERCC5	ProteinTech	11331-1-AP	Rabbit	Use with Caution	1:250
106	ERRF1/MIG6	MIG6	ERRF1	Sigma-Aldrich	WH0054206M1	Mouse	Valid	1:50
107	Estrogen Receptor	ER	ESR1	Lab Vision	RM-9101	Rabbit	Valid	1:40
108	Estrogen Receptor alpha (Phospho S118)	ER-a_pS118	ESR1	Abcam	ab32396	Rabbit	Valid	1:1000
109	Ets-1	Ets-1	ETS1	Bethyl	A303-501A	Rabbit	Valid	1:100
110	FAK	FAK	PTK2	Abcam	ab40794	Rabbit	Use with Caution	1:1000
111	FAK (phospho Y397)	FAK_pY397	PTK2	CST	3283	Rabbit	Valid	1:30
112	Fatty Acid Synthase	FASN	FASN	CST	3180	Rabbit	Valid	1:1000
113	Fibronectin	Fibronectin	FN1	Epitomics	1574-1	Rabbit	Valid	1:10000
114	FoxM1	FoxM1	FOXO3	CST	5436	Rabbit	Valid	1:30
115	FoxO3a	FoxO3a	FOXO3	CST	2497	Rabbit	Use with Caution	1:25
116	FoxO3a (phospho S318/S321)	FoxO3a_pS318_S321	FOXO3	CST	9465	Rabbit	Use with Caution	1:30
117	G6PD	G6PD	G6PD	CST	8866	Rabbit	Valid	1:1000

118	Gab2	Gab2	GAB2	CST	3239	Rabbit	Valid	1:300
119	GAPDH	GAPDH	GAPDH	Life Technologies	AM4300	Mouse	Use with Caution	1:50000
120	GATA3	GATA3	GATA3	BD Biosciences	558686	Mouse	Valid	1:300
121	GATA6	GATA6	GATA6	CST	4253	Rabbit	Use with Caution	1:500
122	GCLM	GCLM	GCLM	Abcam	ab124827	Rabbit	Use with Caution	1:500
123	GCN5L2	GCN5L2	KAT2A	CST	3305	Rabbit	Valid	1:30
124	Glutamate Dehydrogenase1/2	Glutamate-D1-2	GLUD	CST	12793	Rabbit	Use with Caution	1:500
125	Glutaminase	Glutaminase	GLS	Abcam	ab156876	Rabbit	Use with Caution	1:250
126	Glycogen Synthase	Gys	GYS1	CST	3886	Rabbit	Valid	1:1500
127	Glycogen Synthase (phospho S641)	Gys_pS641	GYS1	CST	3891	Rabbit	Valid	1:250
128	Granzyme B	Granzyme-B	GZMB	CST	4275	Rabbit	Valid	1:500
129	GSK-3alpha/beta	GSK-3a-b	GSK3A, GSK3B	Santa Cruz	sc-7291	Mouse	Valid	1:750
130	GSK-3alpha/beta (phospho S21/S9)	GSK-3a-b_pS21_S9	GSK3A, GSK3B	CST	9331	Rabbit	Valid	1:200
131	H2AX (phospho S140)	H2AX_pS140	H2AX	Pierce Biotechnology	MA1-2022	Mouse	Use with Caution	1:400
132	Heregulin	Heregulin	NRG1	CST	2573	Rabbit	Valid	1:30
133	HES1	HES1	HES1	CST	11988	Rabbit	Valid	1:1000
134	Hexokinase II	Hexokinase-II	HK2	CST	2106	Rabbit	Valid	1:100
135	Hif-1 alpha	Hif-1-alpha	HIF1A	BD Biosciences	610958	Mouse	Use with Caution	1:50
136	Histone H3	Histone-H3	H3F3A, H3F3B	Abcam	ab1791	Rabbit	Valid	1:3000
137	HLA-DR/DP/DQ/DX	HLA-DR-DP-DQ-DX	HLA	Santa Cruz	sc-53302	Mouse	Valid	1:1000
138	HSP27	HSP27	HSP27	CST	2402	Mouse	Use with Caution	1:100
139	HSP27 (phospho S82)	HSP27_pS82	HSBP1	CST	2401	Rabbit	Valid	1:75
140	HSP70	HSP70	HSP70	CST	4872	Rabbit	Use with Caution	1:150
141	IGF1R (phospho Y1135/Y1136)	IGF1R_pY1135_Y1136	IGF1R	CST	3024	Rabbit	Valid	1:30
142	IGFBP2	IGFBP2	IGFBP2	CST	3922	Rabbit	Valid	1:50
143	IGFRb	IGFRb	INSR	CST	3027	Rabbit	Use with Caution	1:250
144	INPP4b	INPP4b	INPP4B	CST	4039	Rabbit	Valid	1:25
145	Insulin Receptor beta	IR-b	INSRB	CST	3025	Rabbit	Use with Caution	1:750
146	IRF-1	IRF-1	IRF1	Santa Cruz	sc-497	Rabbit	Use with Caution	1:200

147	IRS1	IRS1	IRS1	Millipore	06-248	Rabbit	Valid	1:400
148	Jagged1	Jagged1	JAG1	Abcam	ab109536	Rabbit	Valid	1:750
149	Jak2	Jak2	JAK2	CST	3230	Rabbit	Valid	1:750
150	JNK/SAPK (phospho T183/Y185)	JNK_pT183_Y185	MAPK8	CST	4668	Rabbit	Valid	1:30
151	JNK2	JNK2	MAPK9	CST	4672	Rabbit	Use with Caution	1:30
152	LC3A/B	LC3A-B	LC3AB	CST	4108	Rabbit	Use with Caution	1:500
153	Lck	Lck	LCK	CST	2752	Rabbit	Valid	1:100
154	LDHA	LDHA	LDHA	CST	3582	Rabbit	Use with Caution	1:250
155	LRP6 (phospho S1490)	LRP6_pS1490	LRP6	CST	2568	Rabbit	Valid	1:500
156	MAPK (phospho T202/Y204)	MAPK_pT202_Y204	MAPK1, MAPK3	CST	4377	Rabbit	Valid	1:30
157	Mcl 1	Mcl-1	MCL1	CST	5453	Rabbit	Valid	1:100
158	MDM2 (phospho S166)	MDM2_pS166	MDM2	CST	3521	Rabbit	Valid	1:50
159	MEK1	MEK1	MAP2K1	Abcam	ab32576	Rabbit	Valid	1:1500
160	MEK1 (phospho S217/S221)	MEK1_pS217_S221	MAP2K1 MAP2K2	CST	9154	Rabbit	Valid	1:50
161	MERIT40 (Phospho S29)	MERIT40_pS29	BABAM1	CST	12110	Rabbit	Valid	1:1000
162	Merlin/NF2	Merlin	NF2	Novus Biologicals	22710002	Rabbit	Use with Caution	1:250
163	MIF	MIF	MIF	Santa Cruz	sc-20121	Rabbit	Use with Caution	1:300
164	MMP2	MMP2	MMP2	CST	4022	Rabbit	Valid	1:75
165	Mnk1	Mnk1	MKNK1	CST	2195	Rabbit	Valid	1:1000
166	Monocarboxylic Acid Transporter 4	MCT4	SLC16A4	Millipore	AB3314P	Rabbit	Valid	1:500
167	MSH6	MSH6	MSH6	Novus Biologicals	22030002	Rabbit	Use with Caution	1:1000
168	MSI2	MSI2	MSI2	Abcam	ab76148	Rabbit	Use with Caution	1:4000
169	mTOR	mTOR	MTOR	CST	2983	Rabbit	Valid	1:1000
170	mTOR (phospho S2448)	mTOR_pS2448	MTOR	CST	2971	Rabbit	Use with Caution	1:50
171	Myosin heavy chain 11	Myosin-11	MYH11	Novus Biologicals	21370002	Rabbit	Valid	1:5000
172	Myosin IIa (phospho S1943)	Myosin-IIa_pS1943	MYH9	CST	5026	Rabbit	Valid	1:1000
173	Myt1	Myt1	MYT1	CST	4282	Rabbit	Use with Caution	1:2000
174	NAPSIN A	NAPSIN-A	NAPSA	Abcam	ab129189	Rabbit	Use with Caution	1:150
175	N-Cadherin	N-Cadherin	CDH2	CST	4061	Rabbit	Valid	1:30

176	NDRG1 (phospho T346)	NDRG1_pT346	NDRG1	CST	3217	Rabbit	Valid	1:100
177	NDUFB4	NDUFB4	NDUFB4	Abcam	ab110243	Mouse	Valid	1:30
178	NF-kappaB p65 (phospho S536)	NF-kB-p65_pS536	RELA	CST	3033	Rabbit	Use with Caution	1:30
179	Notch1	Notch1	NOTCH1	CST	3268	Rabbit	Valid	1:30
180	Notch3	Notch3	NOTCH3	Santa Cruz	sc-5593	Rabbit	Use with Caution	1:300
181	N-Ras	N-Ras	NRAS	Santa Cruz	sc-31	Mouse	Valid	1:50
182	Oct-4	Oct-4	OCT4	CST	2750	Rabbit	Use with Caution	1:200
183	p21	p21	CDKN1A	Santa Cruz	sc-397	Rabbit	Valid	1:150
184	p27 KIP 1	p27-Kip-1	CDKN1B	Abcam	ab32034	Rabbit	Valid	1:50
185	p27/KIP 1 (phospho T198)	p27_pT198	CDKN1B	Abcam	ab64949	Rabbit	Valid	1:30
186	p38 MAPK	p38	MAPK14	CST	9212	Rabbit	Valid	1:1500
187	p38 MAPK (phosphorT180/Y182)	p38_pT180_Y182	MAPK14	CST	9211	Rabbit	Valid	1:50
188	p44/42 MAPK	p44-42-MAPK	MAPK3	CST	4695	Rabbit	Valid	1:2000
189	p53	p53	TP53	CST	9282	Rabbit	Use with Caution	1:2500
190	p70 S6 Kinase (phospho T389)	p70-S6K_pT389	RPS6KB1	CST	9205	Rabbit	Valid	1:50
191	p70/S6K1	p70-S6K1	RPS6KB1	Abcam	ab32529	Rabbit	Valid	1:300
192	p90RSK (phospho T573)	p90RSK_pT573	RPS6K	CST	9346	Rabbit	Use with Caution	1:30
193	PAI-1	PAI-1	SERPINE1	BD Biosciences	612024	Mouse	Valid	1:100
194	PAICS	PAICS	PAICS	Sigma-Aldrich	HPA035895	Rabbit	Use with Caution	1:250
195	PAK1	PAK1	PAK1	CST	2602	Rabbit	Valid	1:1000
196	PAK4	PAK4	PAK4	CST	3242	Rabbit	Valid	1:750
197	PAR	PAR	PAR	Trevigen	4336-BPC-100	Rabbit	Use with Caution	1:15000
198	PARK7/DJ1	DJ1	PARK7	Abcam	ab76008	Rabbit	Valid	1:5000
199	PARP	PARP	PARP1	CST	9532	Rabbit	Valid	1:1000
200	Paxillin	Paxillin	PXN	CST	2542	Rabbit	Use with Caution	1:250
201	P-Cadherin	P-Cadherin	CDH3	CST	2130	Rabbit	Use with Caution	1:50
202	PCNA	PCNA	PCNA	CST	2586	Mouse	Use with Caution	1:1000
203	PD-1	PD-1	PD1	CST	43248	Rabbit	Valid	1:500
204	Pdcd4	Pdcd4	PDCD4	Rockland	600-401-965	Rabbit	Use with Caution	1:750
205	PDGFR beta	PDGFR-b	PDGFRB	CST	3169	Rabbit	Valid	1:100

206	PDHK1	PDHK1	PDHK1	CST	3820	Rabbit	Use with Caution	1:500
207	PDK1	PDK1	PDPK1	CST	3062	Rabbit	Valid	1:50
208	PDK1 (phospho S241)	PDK1_pS241	PDPK1	CST	3061	Rabbit	Valid	1:50
209	PD-L1	PD-L1	CD274	CST	13684	Rabbit	Use with Caution	1:150
210	PEA-15	PEA-15	PEA15	CST	2780	Rabbit	Valid	1:100
211	PED/PEA-15 (phospho S116)	PEA-15_pS116	PEA15	Invitrogen	44-836G	Rabbit	Valid	1:1000
212	PI3K p110 alpha	PI3K-p110-a	PIK3CA	CST	4255	Rabbit	Use with Caution	1:75
213	PI3K p110 beta	PI3K-p110-b	PIK3CB	Santa Cruz	sc-376412	Mouse	Use with Caution	1:60
214	PI3K p85	PI3K-p85	PIK3R1	Millipore	06-195	Rabbit	Valid	1:15000
215	PKA RI alpha	PKA-a	PRKAR1A	CST	5675	Rabbit	Valid	1:200
216	PKCalpha	PKCa	PRKCA	CST	2056	Rabbit	Valid	1:200
217	PKC beta II (phospho S660)	PKC-b-II_pS660	PRKCA, PRKCB PRKCD, PRKCE PRKCH, PRKCQ	CST	9371	Rabbit	Valid	1:200
218	PKC delta (phospho S664)	PKC-delta_pS664	PRKCD	Millipore	07-875	Rabbit	Valid	1:100
219	PKM2	PKM2	PKM2	CST	4053	Rabbit	Use with Caution	1:300
220	PLC gamma2 (phospho Y759)	PLC-gamma2_pY759	PLCG2	CST	3874	Rabbit	Use with Caution	1:30
221	PLK1	PLK1	PLK1	CST	4513	Rabbit	Use with Caution	1:125
222	PMS2	PMS2	PMS2	Novus Biologicals	22510002	Rabbit	Valid	1:1500
223	PRAS40	PRAS40	AKT1S1	Invitrogen	AHO1031	Mouse	Use with Caution	1:250
224	PRAS40 (phospho T246)	PRAS40_pT246	AKT1S1	Life Technologies	441100G	Rabbit	Valid	1:500
225	PREX1	PREX1	PREX1	Abcam	ab102739	Rabbit	Valid	1:150
226	Progesterone Receptor	PR	PGR	Abcam	ab32085	Rabbit	Valid	1:50
227	PTEN	PTEN	PTEN	CST	9552	Rabbit	Valid	1:500
228	Rab11	Rab11	RAB11A,B	CST	3539	Rabbit	Use with Caution	1:30
229	Rab25	Rab25	RAB25	CST	4314	Rabbit	Valid	1:30
230	Rad50	Rad50	RAD50	CST	3427	Rabbit	Valid	1:250
231	Rad51	Rad51	RAD51	CST	8875	Rabbit	Valid	1:30
232	Raptor	Raptor	RPTOR	CST	2280	Rabbit	Valid	1:300
233	Rb	Rb	RB1	CST	9309	Mouse	**Used for QC**	1:100
234	Rb (phospho S807/S811)	Rb_pS807_S811	RB1	CST	9308	Rabbit	Valid	1:500

235	RBM15	RBM15	RBM15	Novus Biologicals	21390002	Rabbit	Valid	1:5000
236	Rheb	Rheb	RHEB	R&D Systems	MAB3426	Mouse	Use with Caution	1:75
237	Rictor	Rictor	RICTOR	CST	2114	Rabbit	Use with Caution	1:100
238	Rictor (phospho T1135)	Rictor_pT1135	RICTOR	CST	3806	Rabbit	Valid	1:200
239	RIP	RIP	RIP	CST	4926	Rabbit	Use with Caution	1:250
240	RPA32	RPA32	RPA32	CST	2208	Rat	Use with Caution	1:500
241	RPA32 (PhosphoS4/S8)	RPA32_pS4_S8	RPA32	Bethyl	A300-245A	Rabbit	Use with Caution	1:250
242	RSK	RSK	RPS6KA1 RPS6KA2 RPS6KA3	CST	9347	Rabbit	Use with Caution	1:150
243	S6 (phospho S235/S236)	S6_pS235_S236	RPS6	CST	2211	Rabbit	Valid	1:2500
244	S6 (phospho S240/S244)	S6_pS240_S244	RPS6	CST	2215	Rabbit	Valid	1:1000
245	S6 Ribosomal Protein	S6	RPS6	CST	2317	Mouse	Valid	1:1000
246	SCD	SCD	SCD	Santa Cruz	sc-58420	Mouse	Valid	1:30
247	SDHA	SDHA	SDHA	CST	11998	Rabbit	Valid	1:250
248	SF2/ASF	SF2	SRSF1	Invitrogen	32-4500	Mouse	Valid	1:150
249	Shc (phospho Y317)	Shc_pY317	SHC1	CST	2431	Rabbit	Valid	1:30
250	SHP-2 (phospho Y542)	SHP-2_pY542	PTPN11	CST	3751	Rabbit	Use with Caution	1:75
251	SLC1A5	SLC1A5	SLC1A5	Sigma-Aldrich	HPA035240	Rabbit	Use with Caution	1:15000
252	Slfn1	Slfn1	SLFN11	Santa Cruz	sc-136891	Goat	Use with Caution	1:250
253	Smac/Diablo	Smac	DIABLO	CST	2954	Mouse	**Used for QC**	1:150
254	Smad1	Smad1	SMAD1	Abcam	ab33902	Rabbit	Valid	1:750
255	Smad3	Smad3	SMAD3	Abcam	ab40854	Rabbit	Valid	1:150
256	Smad4	Smad4	SMAD4	Santa Cruz	sc-7966	Mouse	Valid	1:30
257	Snail	Snail	SNAI1	CST	3895	Mouse	**Used for QC**	1:50
258	SOD1	SOD1	SOD1	CST	4266	Mouse	Valid	1:500
259	SOD2	SOD2	SOD2	CST	13194	Rabbit	Valid	1:200
260	Sox2	Sox2	SOX2	CST	2748	Rabbit	Valid	1:200
261	Src	Src	SRC	Millipore	05-184	Mouse	Valid	1:200
262	Src (phospho Y527)	Src_pY527	SRC, YES1, FYN FGR	CST	2105	Rabbit	Valid	1:30
263	Src Family (phospho Y416)	Src_pY416	SRC, LYN, FYN LCK, YES1, HCK	CST	2101	Rabbit	Valid	1:500

264	Stat3	Stat3	STAT3	CST	4904	Rabbit	Use with Caution	1:3000
265	Stat3 (phospho Y705)	Stat3_pY705	STAT3	CST	9145	Rabbit	Valid	1:100
266	Stat5a	Stat5a	STAT5A	Abcam	ab32043	Rabbit	Valid	1:250
267	Stathmin 1	Stathmin-1	STMN1	Abcam	ab52630	Rabbit	Valid	1:75
268	Sting	STING	STING	CST	13647	Rabbit	Valid	1:250
269	Syk	Syk	SYK	Santa Cruz	sc-1240	Mouse	Valid	1:3000
270	Tau	Tau	TAU	Millipore	05-348	Mouse	Use with Caution	1:100
271	TAZ	TAZ	WWTR1	CST	4883	Rabbit	Valid	1:300
272	TFAM	TFAM	TFAM	CST	7495	Rabbit	Valid	1:300
273	TIGAR	TIGAR	C12ORF5	Abcam	ab137573	Rabbit	Valid	1:100
274	Transferrin Receptor	TFRC	TFRC	Novus Biologicals	22500002	Rabbit	Valid	1:15000
275	Trans-glutaminase II	Trans-glutaminase	TGM2	Lab Vision	MS-224-P1	Mouse	Valid	1:150
276	TRIM25	TRIM25	TRIM25	Abcam	ab167154	Rabbit	Use with Caution	1:2000
277	TSC1/Hamartin	TSC1	TSC1	CST	4906	Rabbit	Use with Caution	1:200
278	TSC2/Tuberin (phosphoT1462)	Tuberin_pT1462	TSC2	CST	3617	Rabbit	Valid	1:30
279	TTF1**	TTF1	NKX2-1	Abcam	ab76013	Rabbit	Valid	1:250
280	Tuberin	Tuberin	TSC2	Abcam	ab32554	Rabbit	Valid	1:2500
281	TUFM	TUFM	TUFM	Abcam	ab173300	Rabbit	Valid	1:300
282	Twist	TWIST	TWIST2	Santa Cruz	sc-81417	Mouse	Use with Caution	1:30
283	Tyro3	Tyro3	TYRO3	CST	5585	Rabbit	Valid	1:30
284	UBAC1	UBAC1	UBAC1	Sigma-Aldrich	HPA005651	Rabbit	Valid	1:250
285	Ubiquityl Histone H2B	Ubq-Histone-H2B	HIST1H2BB	CST	5546	Rabbit	Use with Caution	1:500
286	UGT1A	UGT1A	UGT1A1	Santa Cruz	sc-271268	Mouse	Valid	1:75
287	ULK1 (phospho S757)	ULK1_pS757	ULK1	CST	6888	Rabbit	Use with Caution	1:1000
288	VASP	VASP	VASP	CST	3112	Rabbit	Valid	1:250
289	VDAC1/Porin	Porin	VDAC1	Abcam	ab14734	Mouse	Valid	1:300
290	VEGF Receptor 2	VEGFR-2	KDR	CST	2479	Rabbit	Valid	1:12000
291	VHL/EPPK1**	VHL-EPPK1	EPPK1	BD Biosciences	556347	Mouse	Use with Caution (Targets EPPK1)	1:500
292	Vimentin	Vimentin	VIM	Dako	M0725	Mouse	Use with Caution	1:250

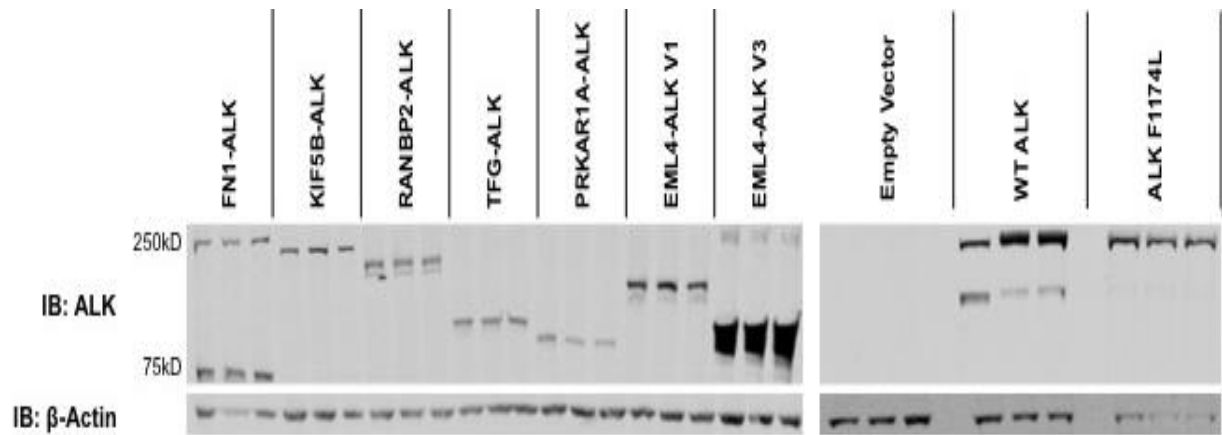
293	Wee1	Wee1	WEE1	CST	4936	Rabbit	Use with Caution	1:2000
294	Wee1 (Phospho S642)	Wee1_pS642	WEE1	CST	4910	Rabbit	Use with Caution	1:75
295	WIPI1	WIPI1	WIPI1	CST	12124	Rabbit	Use with Caution	1:1000
296	WIPI2	WIPI2	WIPI2	CST	8567	Rabbit	Use with Caution	1:1000
297	XBP1	XBP1	XBP1	Santa Cruz	sc-32136	Goat	Use with Caution	1:200
298	XPA	XPA	XPA	Santa Cruz	sc-56813	Mouse	Valid	1:75
299	XPF	XPF	XPF	Abcam	ab73720	Rabbit	Use with Caution	1:100
300	XRCC1	XRCC1	XRCC1	CST	2735	Rabbit	Use with Caution	1:30
301	YAP	YAP	YAP1	Santa Cruz	sc-15407	Rabbit	Use with Caution	1:750
302	YAP (phospho S127)	YAP_pS127	YAP1	CST	4911	Rabbit	Valid	1:250
303	YB1 (phospho S102)	YB1_pS102	YBX1	CST	2900	Rabbit	Valid	1:50
304	ZAP-70	ZAP-70	ZAP70	CST	2705	Rabbit	Use with Caution	1:1000

Supplementary Table S3.1: MD Anderson RPPA Standard Antibody List. This table lists the antibodies used on the RPPA to detect total proteins and phospho-proteins. This table was adapted from the MD Anderson Cancer Center RPPA Core website: (<https://www.mdanderson.org/research/research-resources/core-facilities/functional-proteomics-rppa-core.html>)

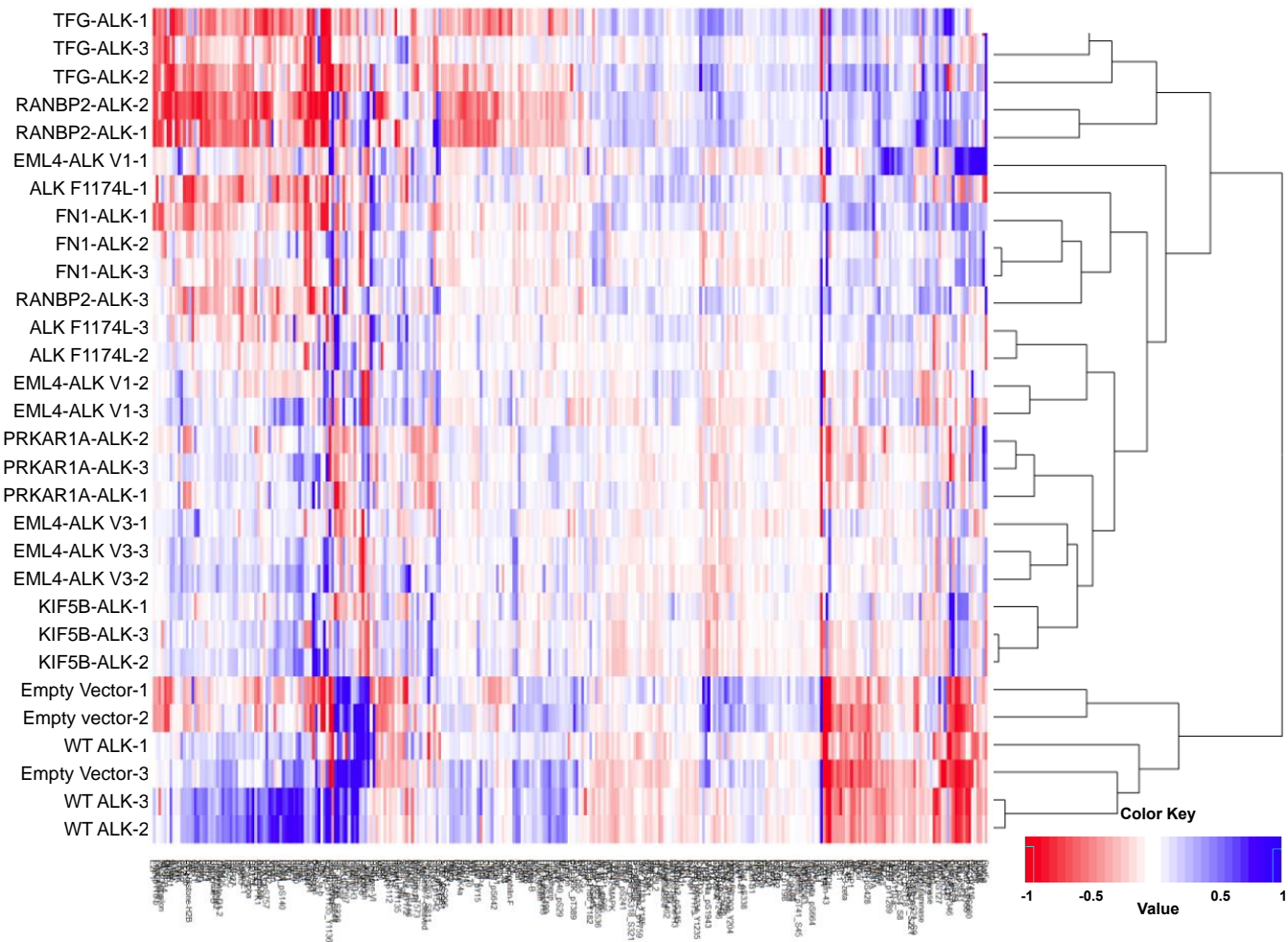
*"Validation Status" column: *Valid* = RPPA and Western blot correlation > 0.7; *Use with Caution* = RPPA and Western blot correlation < 0.7; *Under Evaluation* = Antibody has given mixed results and/or evaluated by another lab; We are in the process of (re)validating; *Used for QC* = These antibodies are used for tissue sample quality control (QC); WILL NOT report to Datasets containing tissue samples, but WILL report to Datasets containing cell line samples.

** VHL Antibody has been identified to recognize EPPK1 by MS, RNAi and correlation with RNA expression.

Ab = antibody



Supplementary Figure S3.1. Expression of ALK fusion variants. Lysates from the ALK variant 3T3 cell line clones selected for experiments were assessed for ALK expression by immunoblot. Lysates were run in triplicate.



Supplementary Figure S3.2. Active ALK variants change the protein expression profile of NIH3T3 cells. This heatmap is a depiction of the expression level for all 304 interrogated proteins and/or phosphorylation sites for all cell lines and all replicates. This data was generated using a Reverse Phase Protein Array (RPPA). The experiment was run using biological replicates run in triplicate. Note that blue is representative of increased expression (from the mean), and red is decreased expression (from the mean). A hierarchical tree is depicted on the right to further distinguish delineation of the cell lines. The protein names and phosphorylation sites are represented at the bottom. This heatmap exhibits unsupervised clustering of all cell lines and replicates.

RANBP2-ALK			TFG-ALK		
Protein	Fold Change (Log)	Adjusted p Value	Protein	Fold Change (Log)	Adjusted p Value
GAPDH	4.87	8.45E-05	GAPDH	7.79	3.24E-06
Stat5a	4.64	8.99E-06	Stat5a	3.37	1.45E-04
Connexin-43	4.53	1.54E-06	ACC pS79	2.78	6.11E-04
Stat3 pY705	4.13	3.71E-03	UGT1A	2.68	8.31E-05
NDRG1 pT346	3.38	4.03E-04	Stat3	2.63	3.60E-04
ACC pS79	3.00	3.02E-04	NDRG1 pT346	2.61	3.29E-03
53BP1	2.32	4.89E-10	Mcl-1	2.51	2.95E-09
Notch1	2.31	1.70E-12	Notch1	2.37	9.26E-13
UGT1A	2.25	6.13E-04	PKM2	2.28	8.08E-04
Stat3	2.21	2.17E-03	Connexin-43	2.12	3.19E-03
MDM2 pS166	2.15	9.90E-04	mTOR	2.12	6.99E-08
Gys pS641	2.12	1.02E-04	c-Jun pS73	1.88	4.31E-03
PAK1	2.07	1.55E-08	Transglutaminase	1.86	3.21E-08
Gys	2.04	1.55E-05	Akt	1.85	4.68E-07
mTOR	1.89	8.25E-07	Gab2	1.81	1.58E-07
4E-BP1	1.85	7.82E-05	4E-BP1	1.81	1.16E-04
Shc pY317	1.84	1.13E-05	MDM2 pS166	1.80	7.69E-03
PKM2	1.83	8.99E-03	53BP1	1.79	2.14E-07
c-Jun pS73	1.82	6.16E-03	Bim	1.75	3.53E-09
PKC-b-II pS660	1.78	5.73E-03	TSC1	1.70	3.61E-05
Transglutaminase	1.75	1.59E-07	14-3-3-zeta	1.67	9.79E-07
PRAS40 pT246	1.69	4.23E-06	Rictor	1.63	8.68E-09
Akt	1.69	4.53E-06	p44/42 MAPK	1.60	9.25E-07
GSK-3a-b pS21/S9	1.68	2.09E-04	Gys pS641	1.59	7.11E-03
Rictor	1.68	3.26E-09	Cyclin-D3	1.58	3.05E-06
SHP-2 pY542	1.64	2.57E-03	FAK	1.57	9.73E-08
Mcl-1	1.62	4.01E-05	PEA-15	1.52	6.95E-05

eIF4E	1.60	7.98E-07	PRAS40 pT246	1.52	7.30E-05
TSC1	1.59	1.68E-04	Gys	1.52	3.50E-03
ARID1A	1.57	1.05E-04	PAR	-8.78	1.69E-03
PKA-a	1.56	5.02E-04	IGF1R pY1135/Y1136	-5.55	3.19E-04
Bim	1.53	2.75E-07	FASN	-5.37	1.52E-04
Caveolin-1	1.52	2.80E-02	Src pY527	-4.25	3.07E-04
FAK	1.51	3.72E-07	Chk1	-3.87	9.38E-04
Chk1	-5.71	6.99E-05	GSK-3a-b	-3.18	5.29E-04
FASN	-4.41	5.43E-04	S6 pS235/S236	-2.88	2.55E-04
HES1	-3.35	3.04E-04	Vimentin	-2.85	2.05E-06
b-Catenin	-3.25	2.14E-04	Rad50	-2.58	1.91E-04
Annexin-I	-2.97	7.38E-07	ADAR1	-2.56	1.45E-04
ADAR1	-2.82	4.73E-05	CD171	-2.46	3.73E-07
CD171	-2.77	5.34E-08	Annexin-I	-2.42	1.27E-05
Rad50	-2.67	1.31E-04	ULK1 pS757	-2.38	1.48E-06
Porin	-2.55	1.55E-04	SOD1	-2.28	1.38E-05
Vimentin	-2.54	9.76E-06	B-Raf	-2.26	1.52E-07
c-IAP2	-2.48	2.59E-05	XPA	-2.25	3.89E-04
Smad4	-2.44	2.38E-05	BAP1	-2.24	1.16E-04
HER3	-2.42	5.28E-08	E2F1	-2.21	4.71E-05
YAP	-2.35	3.57E-05	Histone-H3	-2.19	5.86E-04
G6PD	-2.33	3.79E-08	Porin	-2.18	9.53E-04
TRIM25	-2.32	8.24E-09	Cyclin-E1	-2.17	4.49E-06
SCD	-2.27	1.64E-05	c-IAP2	-2.15	1.83E-04
ULK1 pS757	-2.25	3.62E-06	Smad4	-2.15	1.41E-04
TFRC	-2.21	3.03E-04	Syk	-2.11	6.62E-05
GATA3	-2.18	6.01E-05	HER3	-2.10	7.18E-07
GSK-3a-b	-2.15	1.28E-02	YAP	-2.06	2.31E-04
SF2	-2.14	6.39E-04	NDUFB4	-2.03	3.32E-04
BAP1	-2.14	2.12E-04	N-Ras	-2.02	6.28E-04
N-Ras	-2.13	3.14E-04	CD31	-2.02	4.64E-03
Smac	-2.13	1.81E-07	RPA32	-2.02	9.08E-04

PRAS40	-2.11	5.08E-04	S6 pS240/S244	-2.01	9.94E-03
Syk	-2.10	7.11E-05	SCD	-2.00	1.16E-04
RPA32	-2.02	8.66E-04	G6PD	-1.98	9.64E-07
Bcl2	-2.02	4.85E-06	Beclin	-1.98	3.26E-04
Cyclin-E1	-2.01	1.69E-05	TRIM25	-1.96	2.96E-07
cdc25C	-2.01	8.76E-07	VHL-EPPK1	-1.93	6.61E-04
VHL-EPPK1	-2.01	3.84E-04	HES1	-1.90	3.09E-02
Src	-1.99	2.77E-05	TFRC	-1.90	2.15E-03
SOD1	-1.99	1.25E-04	SLC1A5	-1.90	1.14E-05
Notch3	-1.98	4.15E-06	Src	-1.84	1.15E-04
S6 pS235/S236	-1.97	9.95E-03	CD44	-1.83	6.22E-04
RIP	-1.97	3.41E-08	Smac	-1.81	5.41E-06
Sox2	-1.95	6.16E-06	Pdcd4	-1.81	3.57E-03
PARP	-1.90	1.79E-06	Bcl2	-1.79	5.35E-05
Cdc2 pY15	-1.88	1.44E-04	Notch3	-1.78	3.53E-05
ATM	-1.87	2.56E-04	PARP	-1.75	1.08E-05
E2F1	-1.82	9.06E-04	HER2	-1.74	8.96E-04
Hif1 α	-1.77	1.29E-05	A-Raf	-1.72	3.71E-07
Granzyme-B	-1.75	1.25E-06	PRAS40	-1.71	7.36E-03
Glutamate-D1-2	-1.75	9.20E-06	Hif1 α	-1.71	2.85E-05
HSP27	-1.74	2.02E-06	RIP	-1.68	1.86E-06
MCT4	-1.72	8.19E-07	Chk2	-1.67	1.21E-04
CD31	-1.72	2.28E-02	SHP-2 pY542	-1.67	1.84E-03
PAI-1	-1.71	4.00E-06	Caveolin-1	-1.66	9.48E-03
Chk2	-1.70	8.72E-05	PAI-1	-1.66	8.78E-06
BRD4	-1.70	1.01E-05	CD49b	-1.64	1.08E-02
B7-H4	-1.69	6.41E-05	HSP27	-1.62	1.17E-05
HSP27 pS82	-1.69	3.53E-05	E-Cadherin	-1.60	5.87E-06
Histone-H3	-1.67	1.49E-02	cdc25C	-1.60	1.36E-04
PAK4	-1.67	1.83E-06	Sox2	-1.60	3.69E-04
XPA	-1.65	1.60E-02	BRD4	-1.59	5.13E-05
MIG6	-1.64	1.32E-02	MCT4	-1.59	8.17E-06

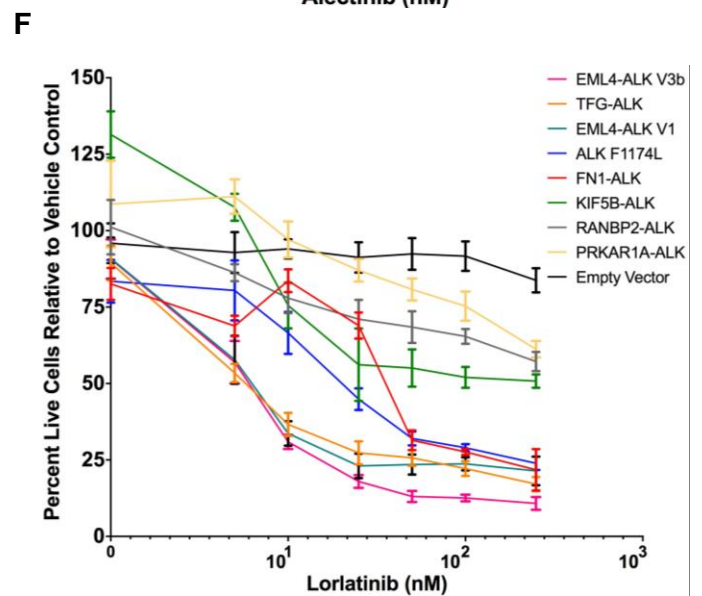
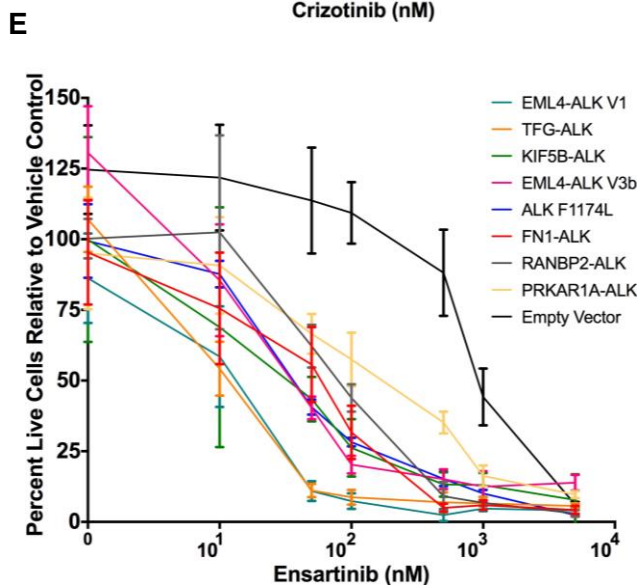
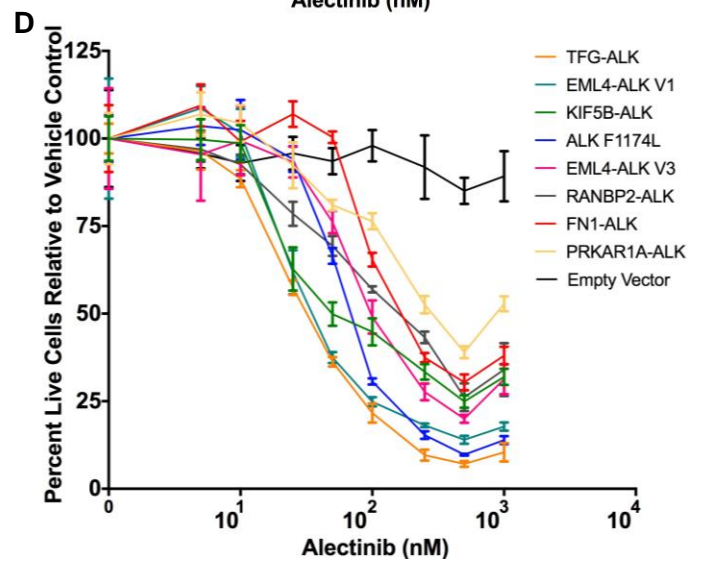
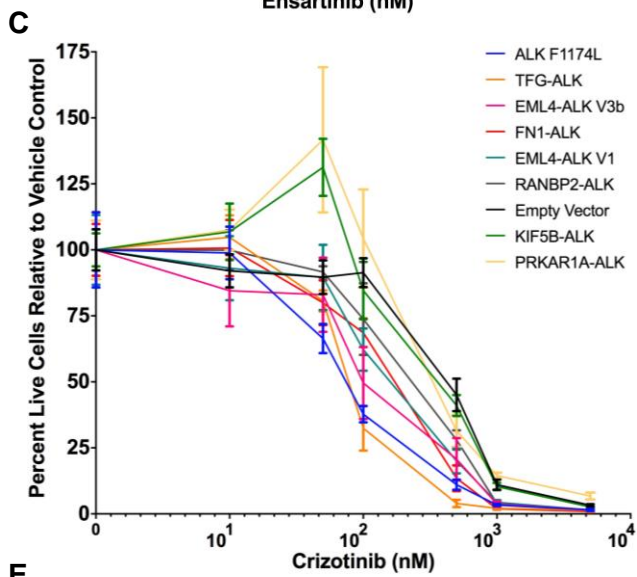
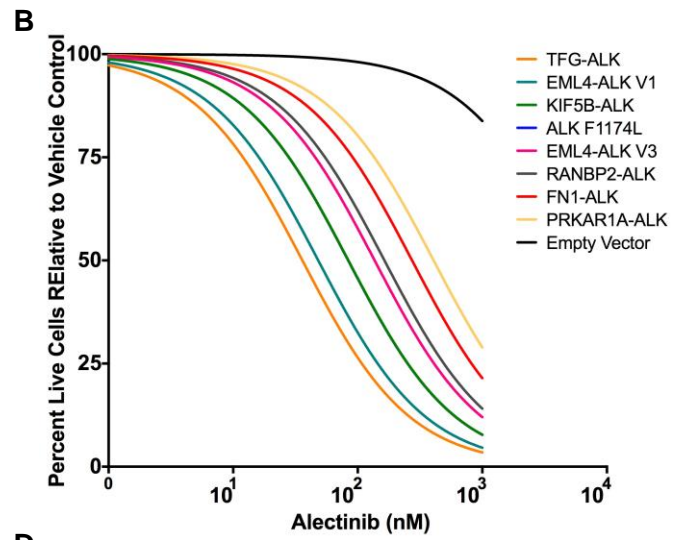
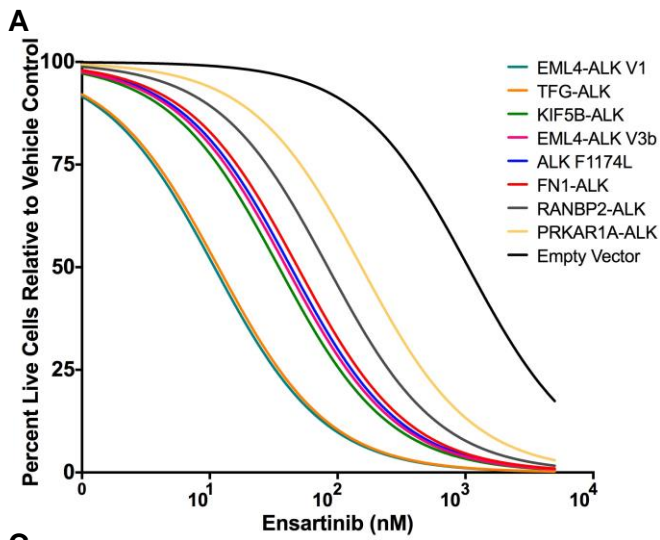
HER2	-1.63	2.66E-03
LRP6 pS1490	-1.62	1.23E-05
Beclin	-1.61	6.49E-03
MERIT40 pS29	-1.61	6.88E-06
E-Cadherin	-1.60	5.99E-06
p16INK4a	-1.56	6.56E-05
SLC1A5	-1.56	6.66E-04
A-Raf	-1.55	6.31E-06
CD44	-1.55	7.88E-03
MSI2	-1.55	6.28E-04
PMS2	-1.54	7.32E-06
CD49b	-1.54	2.32E-02
SDHA	-1.54	7.70E-03
Tau	-1.53	3.43E-04
Cox-IV	-1.51	2.67E-05

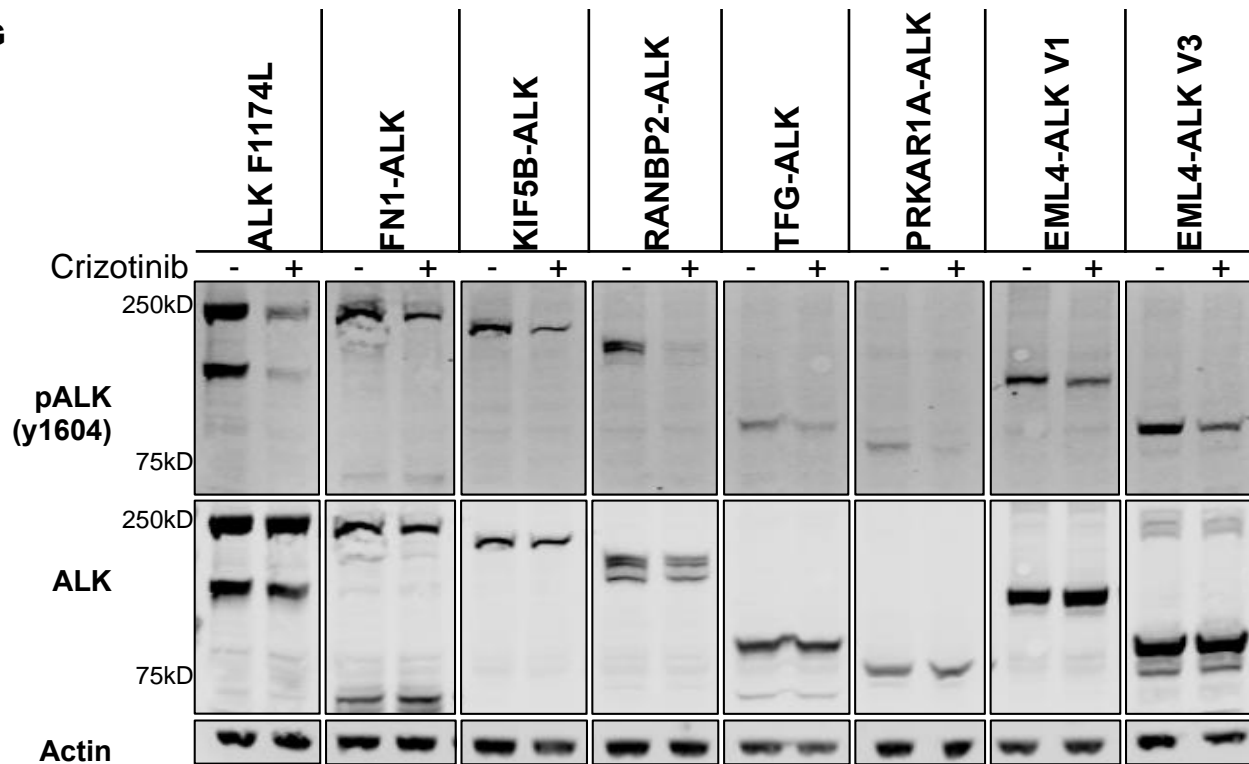
Granzyme-B	-1.58	2.06E-05
PDGFR-b	-1.58	2.22E-02
SF2	-1.55	3.01E-02
B7-H4	-1.55	4.56E-04
HSP27 pS82	-1.52	4.12E-04
BiP-GRP78	-1.50	1.04E-05

EML4-ALK V3			KIF5B-ALK			PRKAR1A-ALK		
Protein	Fold Change (Log)	Adj. p Value	Protein	Fold Change (Log)	Adj. p Value	Protein	Fold Change (Log)	Adj. p Value
GAPDH	6.73	8.73E-06	Connexin-43	4.52	1.58E-06	ACC pS79	2.98	3.20E-04
NDRG1 pT346	3.04	9.60E-04	c-Jun pS73	3.62	2.18E-06	PKC-b-II pS660	2.31	2.26E-04
ACC pS79	2.73	7.14E-04	NDRG1 pT346	3.48	3.17E-04	GAPDH	2.25	2.04E-02
Connexin-43	2.73	2.37E-04	GAPDH	3.46	1.00E-03	Stat5a	2.21	6.41E-03
Stat3 pY705	2.63	3.67E-02	PKM2	2.86	6.44E-05	Notch1	2.14	9.23E-12
PKM2	2.51	2.72E-04	ACC pS79	2.81	5.57E-04	Stat3	1.96	7.40E-03
Gys pS641	2.06	1.53E-04	PRAS40	2.40	9.72E-05	UGT1A	1.95	3.37E-03
UGT1A	2.04	1.92E-03	Gys pS641	2.37	2.00E-05	Connexin-43	1.93	8.19E-03
Gys	1.84	9.97E-05	MDM2 pS166	2.11	1.24E-03	PKM2	1.88	6.63E-03

Notch1	1.82	6.88E-10	Gys	2.04	1.61E-05	c-Jun pS73	1.66	1.76E-02
PKC-b-II pS660	1.81	4.83E-03	PKC-b-II pS660	2.02	1.21E-03	TSC1	1.66	6.12E-05
Stat5a	1.79	3.71E-02	Cox2	2.00	2.90E-05	MDM2 pS166	1.59	3.03E-02
Shc pY317	1.71	5.56E-05	Mcl-1	1.77	4.45E-06	PRAS40	1.56	2.35E-02
MDM2 pS166	1.68	1.63E-02	Gab2	1.75	3.89E-07	Akt	1.53	6.91E-05
TSC1	1.60	1.35E-04	PAK1	1.65	3.89E-06	PAR S6	-5.25	1.20E-02
4E-BP1	1.60	1.18E-03	Rictor	1.59	1.94E-08	pS235/S236	-4.53	3.54E-06
Rictor	1.58	2.42E-08	Notch1	1.54	1.48E-07	IGF1R pY1135/Y1136	-3.07	1.01E-02
PDGFR-b S6	-3.85	4.20E-07	PAR	-3.62	4.45E-02	S6 pS240/S244	-2.94	2.74E-04
pS235/S236	-3.06	1.41E-04	Vimentin	-2.70	4.18E-06	Vimentin	-2.60	7.30E-06
Vimentin	-2.69	4.53E-06	Caveolin-1	-2.43	6.95E-05	Histone-H3	-2.51	1.18E-04
Histone-H3	-2.32	2.96E-04	S6 pS235/S236	-2.26	2.75E-03	HES1	-2.19	1.03E-02
Src pY527	-2.31	2.05E-02	Histone-H3	-2.17	6.63E-04	VHL-EPPK1	-1.85	1.20E-03
Pdcd4	-2.10	5.16E-04	PDGFR-b	-2.16	4.51E-04	SOD1	-1.75	9.27E-04
S6 pS240/S244	-1.95	1.29E-02	Fibronectin	-2.07	5.31E-06	Rad50	-1.68	2.12E-02
Fibronectin	-1.83	5.61E-05	Pdcd4	-1.98	1.15E-03	CD171	-1.67	4.22E-04
Rad50	-1.61	3.40E-02	S6 pS240/S244	-1.72	3.93E-02	TFRC	-1.62	1.48E-02
CD171	-1.59	1.09E-03	CD171	-1.66	4.63E-04	Syk	-1.62	3.97E-03
SLC1A5	-1.56	6.58E-04	RPA32	-1.62	1.46E-02	ULK1 pS757	-1.61	1.47E-03
Syk	-1.53	9.51E-03	A-Raf	-1.60	2.67E-06	SLC1A5	-1.56	6.69E-04
c-IAP2	-1.53	1.98E-02	SCD	-1.55	6.62E-03	SF2	-1.55	3.14E-02
HER2	-1.52	7.62E-03	Syk	-1.53	9.34E-03	Beclin	-1.54	1.21E-02
SOD1	-1.52	8.92E-03	HER3	-1.53	6.60E-04	XPA	-1.54	3.61E-02
TFRC	-1.51	3.35E-02	VHL-EPPK1	-1.50	2.19E-02	BAP1	-1.53	2.11E-02

Supplementary Table S3.2: Proteins exhibiting most significantly different expression from the WT ALK cell line for each ALK fusion cell line. This table lists the proteins and phospho-proteins that exhibited significantly different expression for each ALK fusion cell line vs the WT ALK cell line along with the fold change and adjusted p-values. Proteins in blue are upregulated, and proteins in red are downregulated.



G

Supplementary Figure S3.3. ALK fusion variants exhibit differential response to ALK TKIs. NIH 3T3 cells expressing the indicated ALK variants were treated with increasing concentrations of ensartinib (A,E), alectinib (B,D), crizotinib (C), or lorlatinib (F) for 72 hours. Propidium iodide (PI) and Hoechst were added and allowed to incubate for 20 minutes before imaging on the ImageExpress. Dose response curves were generated by determining total live cells (# of Hoechst positive cells minus # of PI positive cells) at each dose of ALK TKI. A non-linear regression of for ensartinib and alectinib is represented in A and B. Experiments were performed with 6 replicates per drug concentration and repeated three times. One representative experiment is shown here. (G) NIH 3T3 cells expressing the indicated ALK variants were treated with 100 nM of crizotinib for 2 hours. Cell lysates were probed for pALK, ALK, and actin by infrared western blot.

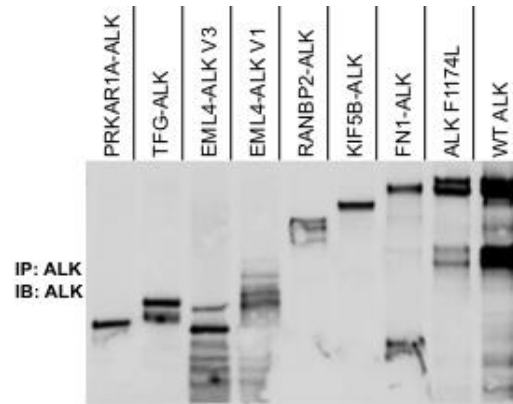
ALK TKI		Status [#]
1st generation	Crizotinib*	FDA-approved, first-line
2nd generation	Alectinib*	FDA approved, first-line FDA-approved, post crizotinib
	Ensartinib*	Investigational
	Ceritinib	FDA approved, first-line FDA-approved, post crizotinib
	Brigatinib	FDA-approved, post crizotinib
	Entrectinib	Investigational
3rd generation	Lorlatinib*	FDA breakthrough therapy designation

Supplementary Table S3.3: ALK TKIs in clinical use. This table depicts the ALK TKI in clinical use and the current approval status. [#]Information obtained from <https://www.fda.gov/>. *Denotes ALK TKIs used in these studies.

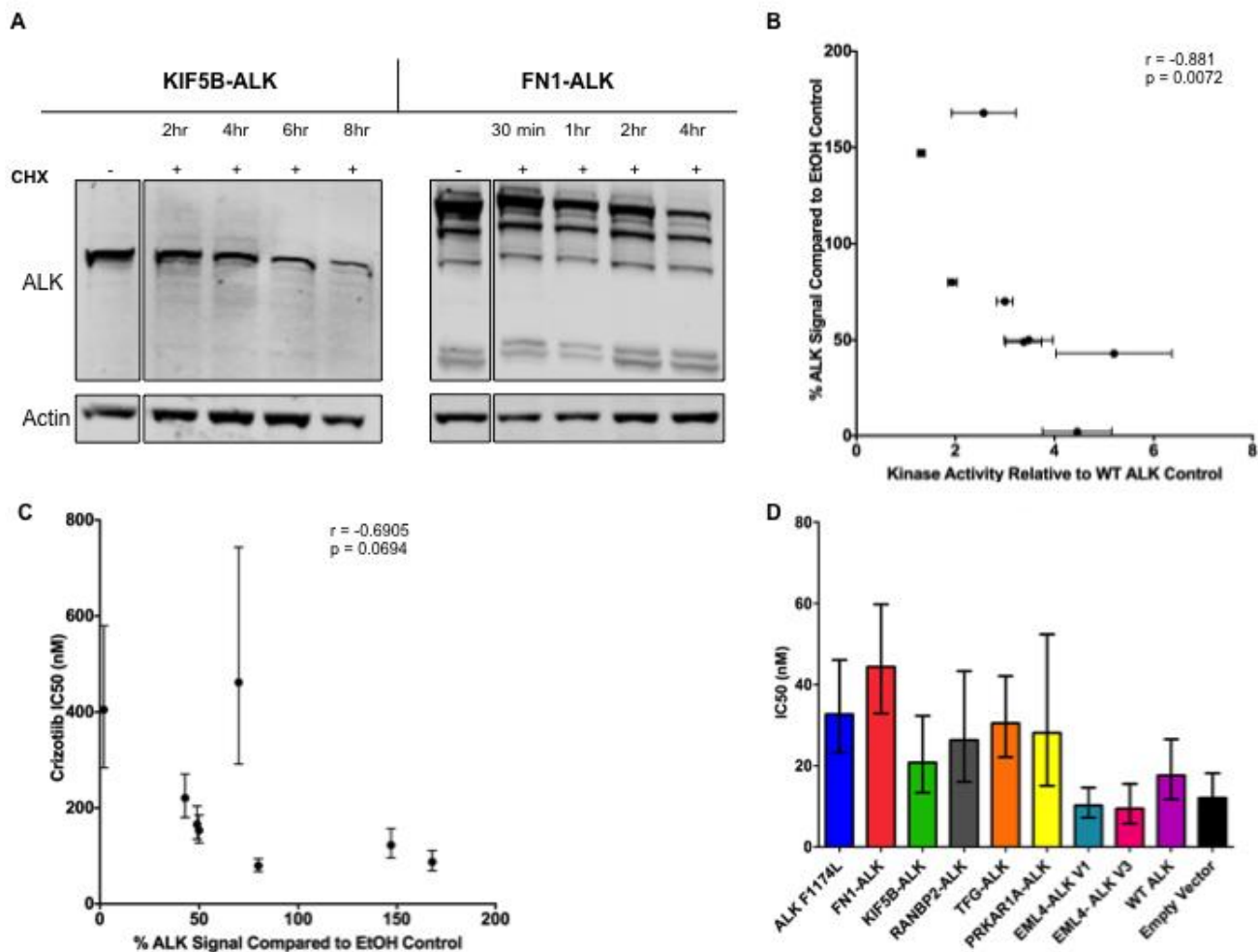
		ALK Variant							
		ALK F1174L	FN1-ALK	KIF5B-ALK	RANBP2-ALK	TFG-ALK	PRKAR1A-ALK	EML4-ALK V1	EML4-ALK V3
Apoptosis pathway	Fold Change	12.91	NS	NS	5.68	7.60	NS	4.90	4.34
	P Value	6.28E-05			2.71E-03	7.12E-04		5.26E-03	9.00E-03
Breast reactive pathway	Fold Change	-14.98	-8.67	-256.73	NS	-80.72	NS	-497.93	-57.75
	P Value	9.82E-03	3.39E-02	1.01E-05		1.62E-04		2.22E-06	3.69E-04
Cell Cycle pathway	Fold Change	6.50	8.26	NS	-3.53	NS	-5.63	NS	NS
	P Value	2.02E-03	7.01E-04		2.68E-02		3.78E-03		
Core reactive pathway	Fold Change	NS	NS	-27.22	63.08	NS	NS	-16.27	NS
	P Value			2.23E-03	2.79E-04			7.76E-03	
DNA damage response pathway	Fold Change	NS	-6.07	NS	NS	NS	NS	NS	NS
	P Value		2.22E-02						
EMT pathway	Fold Change	NS	NS	-50.51	33.50	NS	NS	NS	NS
	P Value			2.52E-03	5.77E-03				
Hormone signaling (Breast) pathway	Fold Change	-11.34	-7.27	NS	-16.73	-3.69	NS	-7.30	NS
	P Value	3.51E-04	2.20E-03		7.12E-05	3.17E-02		2.16E-03	
PI3K/AKT pathway	Fold Change	8.64	NS	10.46	NS	NS	NS	NS	NS
	P Value	2.93E-02		1.88E-02					
RAS/MAPK pathway	Fold Change	79.05	18.75	564.08	26.95	NS	NS	NS	NS
	P Value	4.47E-03	4.41E-02	1.57E-04	2.55E-02				
RTK pathway	Fold Change	NS	NS	NS	24.21	NS	NS	33.94	NS
	P Value				1.10E-03			4.28E-04	
TSC/mTOR pathway	Fold Change	NS	NS	NS	NS	NS	-57.01	NS	-15.78
	P Value						3.43E-03		3.49E-02
Kinase Activity	Relative activity	1.9	3.5	4.5	5.2	2.6	3.0	3.4	1.3
	Standard Deviation	0.1	0.5	0.7	1.2	0.7	0.2	0.4	0.1
Crizotinib Sensitivity	IC50	79.5	153.3	405.1	220.7	87.4	461.8	165.8	122.5
	Lower 95% CI	67.3	127.0	283.9	180.1	68.9	291.4	134.9	96.0
	Upper 95% CI	94.0	185.2	580.0	270.5	110.9	743.2	204.0	156.6
Alectinib Sensitivity	IC50	83.9	273.1	84.1	163.9	36.0	406.7	48.2	136.7
	Lower 95% CI	70.3	219.7	65.8	138.1	32.3	327.7	39.0	112.7
	Upper 95% CI	100.1	341.1	107.9	194.6	40.0	507.0	59.7	166.0
Ensartinib Sensitivity	IC50	43.1	48.8	24.3	91.4	11.7	123.2	11.2	39.9
	Lower 95% CI	36.9	37.6	15.8	71.6	9.4	91.3	9.0	27.3
	Upper 95% CI	50.2	62.9	36.9	117.0	14.6	167.1	14.0	58.1

Lorlatinib Sensitivity	IC50	23.8	35.2	89.1	171.6	7.8	328.5	7.7	6.1
	Lower 95% CI	19.7	28.4	56.4	128.7	6.3	262.7	6.1	5.4
	Upper 95% CI	28.8	43.6	145.0	232.4	9.6	418.8	9.7	6.9
Focus Formation	Foci	101.5	352.0	4.0	8.0	70.0	16.0	46.5	13.0
	Standard Deviation	24.7	4.9	1.7	2.0	24.3	5.7	9.2	12.7
Colony Formation	Mean Colony number	11.3	228.7	23.0	55.0	30.7	13.3	27.3	13.3
	Lower 95% CI	-5.2	105.2	-11.5	-94.9	-45.5	-14.9	-36.1	3.0
	Upper 95% CI	27.9	352.1	57.5	204.9	106.8	41.6	90.8	23.7
Protein Stability/Turnover	% ALK total protein at 16 hours	80	50	2	43	168	70	49	147
ALK autophosphorylation	pALK/ALK ratio	0.93	0.56	0.79	0.59	0.15	0.26	0.2	0.17

Supplementary Table S3.4: Compilation of all Phenotypic Data Measurements for ALK Fusion Variants and ALK F1174L. This table contains all data measurements presented in this manuscript. For the RPPA pathway analysis, each pathway listed was assessed for a statistical difference between the indicated ALK variant expressing cell line vs. the WT ALK expressing cell line. Numerical data is shown for pathways in which a statistical significant difference from the WT ALK cell line was found. NS = not significant. Protein stability/turnover is represented as the percent of ALK detected in cellular lysates by immunoblot 16 hours post cyclohexamide treatment. ALK autophosphorylation is represented as the ratio of pALK(Y1604) to total ALK detected in cellular lysates under normal culture conditions.



Supplementary Figure S3.4. Immunoprecipitated ALK fusion variants for kinase assays. ALK variants were immunoprecipitated from 3T3 cellular lysates using a polyclonal ALK antibody (Cell Signaling) and Sepharose A beads. An aliquot of the immunoprecipitate, equal to the amount used in each *in vitro* kinase assay, was assessed by immunoblot for total ALK.



Supplementary Figure S3.5. ALK fusion variants exhibit differences in protein turnover and sensitivity to ganetespib.

(A) KIF5B-ALK or FN1-ALK expressing 3T3 cells were treated with 50 μ g/mL of cycloheximide over a short time course. Lysates were run on SDS-PAGE and probed with an antibody against total ALK. Collection time points are depicted above the respective lane for each variant. **(B)** Depicted is a spearman correlation analysis between kinase activity and protein stability at 16 hours post cycloheximide treatment. All fusion variants and the ALK F1174L positive control are included. Error bars represent kinase activity standard deviations. **(C)** Depicted is a spearman correlation analysis between protein stability at 16 hours post cycloheximide treatment and the respective IC₅₀ values for crizotinib (Figure 2C). All fusion variants and the ALK F1174L positive control are included. Error bars represent IC₅₀ 95% confidence intervals. **(D)** NIH3T3 cell lines stably expressing ALK fusion variants were treated with increasing concentrations of ganetespib for 72 hours. Propidium iodide (PI) and Hoechst were added and allowed to incubate for 20 minutes before imaging on the ImageExpress. Dose response curves were generated by determining total live cells (# of Hoechst positive cells minus # of PI positive cells) at each concentration of ganetespib. IC₅₀s with 95% confidence intervals were generated with the data from the dose response curves using GraphPad Prism and are depicted here. Experiments were performed with 6 replicates per drug concentration and repeated three times. One representative experiment is shown here.