University of California

Santa Cruz

Understanding how mutant KRAS signaling tunes the non-coding transcriptome and induces stemness in lung cancer

A dissertation submitted in satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In

MOLECULAR, CELL AND DEVELOPMENTAL BIOLOGY

By

David Andrew Robert Carrillo

March 2023

The dissertation of
David Andrew Robert Carrillo is approved:

_____
Professor Melissa Jurica, chair

_____
Professor Daniel Kim

_____
Professor Bree Grillo-Hill

_____
Peter Biehl
Vice Provost and Dean of Graduate Studies

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

Understanding how mutant KRAS signaling tunes the non-coding transcriptome and induces stemness in lung cancer

David Andrew Robert Carrillo

Cancer remains a chief health concern across the developed world. With the expansion of life extending technologies, the aging population is set only to increase drastically in the coming years. As a result, novel modes of treating and identifying cancer early in its development is a goal held by researchers across the globe. This thesis focuses on one of the most common oncogenes and cancer drivers, Kirsten rat sarcoma virus oncogene, also known as KRAS. KRAS is a small GTPase responsible for relaying proliferative, differentiation and cell survival signals to the cell. Mutations in the RAS family of GTPases are present in almost a third of all cancers and are frequently overrepresented in lung and pancreatic cancers. However, RAS driven cancers have only until very recently been able to treated with single molecule inhibitors. Furthermore, not much is known about how the KRAS signaling pathway may interact with other established cellular pathways to remodel the transcriptome, an area that remains largely uninvestigated. Resolving these interconnected relationships holds much promise for the development of combinatorial therapies targeting these pathways as well as unlocking the potential for the identification of novel biomarkers. Here we aim to understand how KRAS signaling may coordinate

with the hypoxia response pathway to induce a cellular phenotype similar to that of a stem cell, i.e., a cancer stem cell, as it is widely believed that such phenotypes confer a survival advantage. We also investigate the modes in which KRAS (G12C) signaling tunes the non-coding transcriptome and how this may be exploited for the development of novel biomarkers. Our findings indicate an inconclusive relationship between hypoxia gene signatures and “stemness,” a gene signature score as determined by a machine learning algorithm. We also find that KRAS (G12C) signaling tightly regulates the expression of key transposable elements (TEs) known as Alus, and that these TE’s may serve as suitable biomarkers. Taken together, our findings challenge conventional dogma regarding the formation of cancer stem cells and reveal novel interconnected relationships between KRAS signaling and TE regulation and expression.

ACKNOWLEDGEMENTS

In the fore, let me pay homage and offer gratefulness to my sagacious mentor, Daniel Kim, whose unwavering efforts and assiduous guidance have been instrumental in my academic triumphs. This PhD voyage, fraught with trials that my mind could scarcely conceive, was one that Daniel and I, at times, viewed from different vantage points. Nevertheless, despite the divergences that lay between us, Daniel's unfailing amiability and his deeds bore testimony to his ardor for my success.

From our mentor-pupil association, one principle stands preeminent: that even when two parties espouse contrary convictions, if the mentor can impart a palpable sense of encouragement, both shall emerge from the ordeal transformed for the better, with the prospect of realizing their collective and individual objectives. Let us hope that this symbiotic bond endures for aeons to come.

I must also express my gratitude and pay homage to my external committee member, Bree Grillo-Hill. Bree and I share a common past, having both been mentored by the same erudite luminary at UCSF. In the bygone days, she was a post-doctoral researcher while I was a humble master's student. When I entreated her to serve on my committee, I sought to bask in her profound understanding of the developmental pathways and other such mysteries. Yet, there was something deeper that beckoned her presence, a

feeling that I could not articulate, but I knew, deep down, that I needed her on my committee.

Truly, the tribulations that besieged me during this PhD expedition were manifold and formidable. But Bree remained steadfast in her unwavering support, lending me her hand during my moments of direst need. Amidst the ever-shifting tides and tumultuous undercurrents that defined the relationship I shared with my erstwhile thesis advisory committee, Bree was a constant lighthouse, illuminating the path ahead. Without Bree, it is certain that this dissertation would be naught but an elusive chimera.

In verity, I must express my gratitude to Melissa Jurica, who, in the eleventh hour, acceded to take up the mantle of chairmanship of my committee. Her demeanor, so exact and unambiguous, proved a necessary respite from the former chair's confounding ways. It was by her exacting directives, wrought in unison with my other committee members, that my path to the defense of my dissertation was made lucid and swift.

In conclusion, I must also acknowledge those among the ranks of the Kim laboratory and its associates, foremost among them Roman Reggiardo and Sree Maroli, for their invaluable assistance in my endeavors.

DEDICATION

To my beloved grandmother, Louise Carrillo, You, who were a beacon of light in my life, have left an indelible imprint upon my soul. Your kindness and warmth were as a soothing balm that eased the scars of the world. Your gentle embrace was a sanctuary, a haven from the horrors that stalked the earth.

Through your eyes, I saw the world anew, and in your words, I heard the wisdom of the ages. Though now you have departed this mortal coil, your memory lives on in my heart, guiding me through the darkest of nights.
As I pen these words, I am reminded of the unbounded love and affection that you bestowed upon me, and I am forever grateful. *Sí, realmente ahora están dando doctorados a los mexicanos, y esta tesis va dedicada a ti*.

CHAPTER 1 (OVERVIEW): KRAS signaling tunes both the protein-coding and non-coding transcriptome

Cancer remains a global health issue

Cancer remains a chief health concern in the western world. Approximately 1.9M people in the US are expected to get cancer in their lifetime, and approximately 600k are expected to die[1]. While advances are being made in the treatment of many cancers, recurrence and drug evasion remain a seemingly insurmountable obstacle[2]. This is in part due to the stochastic nature of cancer mutation accumulation via unstable genomes, but also novel cancer phenotypes that arise in individual cells as a result of global transcriptional changes[2]. Indeed, for a future free of cancer to exist, a proper understanding of these phenomena is paramount. This dissertation aims to unveil new knowledge regarding the aforementioned biological phenomena. To do so, a number of new technologies were implemented such as bulk RNAseq library preparation containing unique molecular identifiers and machine learning algorithms that score gene expression signatures for “stemness”. The findings of these studies are described in the following pages.

The cancer Stem Cell Hypothesis

One theory on cancer resistance and metastasis is highlighted by the concept of a cancer stem cell (CSC). These specialized cancer cells often possess a higher degree of "stemness," i.e., a dedifferentiated state that distinguishes them from other surrounding cancer cells within the tumor[3]. In addition to being able to repopulate a tumor, these dedifferentiated cells are able to resist chemo and radiation therapies due in part to their quiescent nature, and they often express pluripotency-related genes such as OCT4 and SOX2[4]. In induced pluripotent stem cell (iPSC) reprogramming models, expression of these factors in differentiated cells results in the acquisition and maintenance of stem cell identity[5]. However, whether comparable mechanisms of epigenetic reprogramming contribute to the acquisition and maintenance of stemness in lung cancers has yet to be determined.

Mutant KRAS signaling, the non-coding transcriptome and stemness

Another avenue of interest that lies within the same vein of understanding cancer survival is unraveling the modes in which the RAS signaling pathway remodels the transcriptome in a manner beneficial to cancer survival and metastasis. RAS proteins are membrane bound GTPases that work to influence a number of proliferative signals to maintain a mode of tumorigenesis and growth when inappropriately executed (Figure 1)[6]. The primary isoforms are *KRAS4A, KRAS4B, HRAS*, and *NRAS* genes, and are

amongst the most frequently mutated in a wide array of cancers[7]. Previous studies in our lab found KRAS mutation variants G12V and G12D regulate the non-coding transcriptome and interferon genes in unique ways[8]. The work in this dissertation seeks to highlight and identify unique differences in non-coding RNA expression in lung cancer cells harboring the G12C mutation. We believe a proper understanding of how mutant KRAS remodels the non-coding transcriptome holds the potential for the development of novel therapeutic strategies. One other avenue we feel holds promise in this regard, is the identification of novel cancer biomarkers. We believe leveraging the non-coding transcriptional changes prescribed by aberrant KRAS signaling may pave the way for identifying these markers.

Cancer Biomarkers and the KRAS connection

The National Cancer Institute (NCI) defines a biomarker as "a biological molecule found in blood, other body fluids, or tissues that is a sign of a normal or abnormal process, or of a condition or disease[9]". Biomarkers are useful as they are able to delineate who has the disease vs. those who do not have said disease. The differences exploited by a biomarker can range from somatic mutations, to abnormal post-transcriptional changes. As far as physical biomarkers go, these can span everything from protein to nucleic acid, but a biomarker can also be a collection of events such as altered gene expression signatures or changes in a cell's metabolic preferences[9].

Biomarkers can also readily be sampled from blood and or any other bodily fluid or excretion. For these reasons, identifying novel biomarkers represents a promising mode for cancer treatment and prevention. Recently, exploitation of the non-coding RNA transcriptome for identification of new biomarkers has become a subject of interest[10].

To study the role of KRAS signaling mediated transcriptome remodeling, we implemented the use of human cancer cell lines and lung airway cells harboring mutant KRAS (either G12V, G12D or G12C). First, we investigated the role of mutant KRAS signaling for transcriptome remodeling in a lung airway epithelial cell line. These cells were transformed with a KRAS (G12V) construct, and subsequent transcription changes assessed via RNAseq. The primary goal of this study was to elucidate early transformation events in a lung cancer model. My role was to investigate and confirm previously established KRAS gene targets for their altered expression. A number of key dysregulated proteins and non-coding RNAs such as XIST and MALAT were investigated using an array of molecular biological techniques including microscopy.

Tools to investigate the CSC hypothesis

To investigate of the CSC hypothesis, we first questioned if the hypoxia signaling network worked in concert with the KRAS signaling pathway to induce stemness. This is because many of the downstream targets are

shared between the two networks[11]. Additionally, we aimed to utilize a 3D spheroid model to investigate these questions. The spheroid model provides both a hypoxic environment at its core, and better recapitulates the tumor microenvironment[12]. Stemness, or a dedifferentiated state was assessed using a machine learning tool developed by Malta et. al[13]. This tool relies on training data of both stem cell and non-stem cell sample. The tool is compatible with both bulk RNAseq data sets as well as single cell RNAseq data sets. Upon acquisition of the necessary tool sets, I first set out to see if hypoxic cells could be readily identified from embryoid bodies derived from CRISPR dcas9 iPSCs. These cells have been induced to knockdown KRAS and have a subsequent scramble control. Additionally, a publicly available glioma single cell RNAseq data set was also analyzed in a similar manner[14]. I then scored all cells for stemness to see which cell clusters (hypoxic vs. not hypoxic) correlated for having a higher overall stemness score.

Results indicated no correlation between hypoxic cells and increased stemness (chapter 3) in both iPSCs grown in 3D conditions (embryoid bodies) as well as in a disassociated glioma. Furthermore, the cluster identified as being most enriched for stemness was also most active with entry to the cell cycle for the glioma dataset, as determined via Seurat (chapter 3). This stands in contrast with current dogma regarding CSC, whereas they evade cancer therapy by virtue of their quiescent state[15]. These findings

subsequently influenced the direction for alternative studies, highlighted in this thesis.

A new class of KRAS (G12C) inhibitors to identify uniquely regulated transposable element RNA

Another avenue of interest was understanding how a new class of KRAS G12C inhibitors act to ablate RAS signaling and contribute to transcriptome remodeling. To do this, experiments in the wet lab utilized a new class of KRAS (G12C) inhibitors[16] in conjunction with a 3D spheroid model. Spheroid models are believed to better recapitulate the tumor microenvironment and also provide a hypoxic core, the likes of which was necessary to address the aforementioned questions in previous chapters. We found treatment of spheroids with this class of drugs resulted in a severely compromised spheroid morphology. Because of this, hypoxic cores were rendered negligible and subsequent experimental strategies were implemented. Retaining the previous 3D model, we exposed spheroids for approximately 3 days at an ARS-1620 concentration of 500 nM.

We found these conditions to be most ideal for the retainment of spheroid viability and sustained KRAS signaling ablation. Subsequent analysis consisted of RNAseq using approximately 100 spheroids (from each experimental condition) using a new cDNA library synthesis protocol referred to as smartseq3. This library preparation method differs from its predecessors

in that it makes use of unique molecular identifiers to control for PCR duplications, a common issue in modern RNAseq analysis. Computational analysis found a wide variety of differentially expressed protein-coding and non-coding genes.

Perhaps most interestingly, we found a specific subclass of transposable elements known as Alus to be preferentially upregulated in spheroids with drug induced KRAS signaling ablation. Specifically, we found the youngest subfamily AluY elements to be upregulated following KRAS (G12C) inhibition. We believe Alus hold promise for serving as a novel and effective biomarker. Both in part to their specificity but also their abundance. Future studies may support Alu expression as being a mode for assessing drug efficacy and tumor response in patients with cancers harboring KRAS G12C mutations. Indeed, additional work done in our lab indicates some fraction of these Alu transcripts may be excreted in microvesicles from cells exposed to the KRAS G12C inhibitor AMG-510[17]. However, further studies are needed to assess the viability of such a process using human blood samples and additional computational datasets.

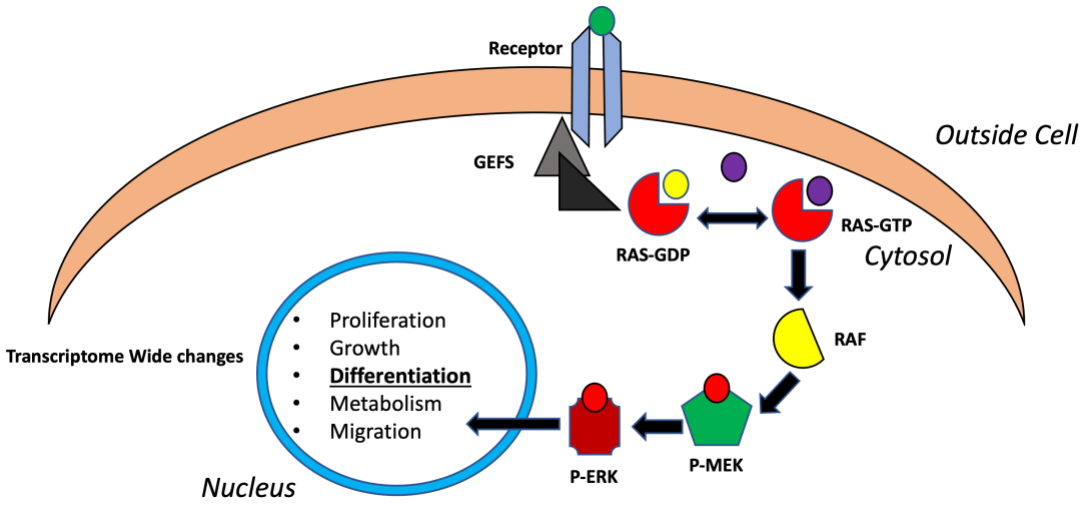

Figure 1: Schematic demonstrating KRAS signaling roles in numerous cellular pathways and behaviors. RAS signaling has a wide number of roles in cellular processes.

CHAPTER 2 (Cell Reports): Differential Expression of protein-coding and non-coding genes in lung airway epithelial cells harboring the KRAS G12V mutation

Introduction

Mutations in the small GTPase RAS are one of the most frequent drivers of cancer, particularly lung cancer[18]. Indeed, mutations in KRAS are represented in approximately one third of all lung cancers[19]. Additionally, mutations in RAS are known to be effective drivers of cancer initiation and metastasis[20]. In addition to being able to trigger entry into the cell cycle, RAS is also known to regulate the protein-coding transcriptome to control a wide array of cell behaviors including: differentiation, cell growth and metabolism[21]. However, not much is known about how RAS signaling reprograms the non-coding transcriptome in the context of cancer. This is of key interest because emerging studies suggest non-coding RNAs, particularly non-coding RNAs play important roles in the regulation of a wide variety of cellular processes[22].

This study first aimed to investigate the transcriptomes of human airway and bronchial epithelial cells transformed with a KRAS G12V construct to assess the broad effects on KRAS regulated noncoding RNAs. Ultimate findings reveal KRAS signaling increases the expression of many non-coding RNAs, and many these non-coding RNAs originate from transposable elements (TEs). We found that these TE's are differentially expressed, are

exported in microvesicles and are widely regulated by KRAB zinc- finger (KZNF) genes, the likes of which are largely downregulated in cells harboring mutant KRAS and many adenocarcinomas *in vivo*. We also observed mutant KRAS inducing an intrinsic IFN-stimulated gene (ISG) signature which is consistent in a large number of cancers[8]. The results of this paper suggest mutant KRAS is responsible for the remodeling of the non-coding transcriptome, having broad implications for intracellular and extracellular RNAs regulated by the KRAS signaling pathway.

Here, I identify and confirm the differential expression of key proteins and non-coding RNAs believed to be regulated by the KRAS signaling pathway. Of the targets, a number of these genes are believed to be important for cancer cell evasion from the immune system as well as the regulation of stem cell identity and cancer progression. Initial investigations into these targets sought to identify key proteins and transcription factors that would provide KRAS driven cancers with a cell survival advantage. To investigate these factors, I utilized fluorescent light microscopy and a lung airway cell model possessing stable expression of KRAS G12V. Targets of interest include the likes of: PDL1, DDX60, MALAT, XIST and more. I also utilized a number of other techniques such as qPCR and immunoblot (data not shown) to validate previous observations pertaining to these protein targets in primary literature[23]. My findings indicate and validate KRAS regulation of some previously established targets. In particular, MALAT, XIST

and NEAT1 demonstrate a strong regulatory effect via KRAS signaling. And all of these non-coding RNAs have significant relevance for the progression and spread of numerous cancers[24,25,26]. Protein coding targets of interest were PDL1 and DDX60. Immunofluorescence microscopy found a potential regulatory effect on PDL1s cellular location, from that of extracellular to intracellular, though further studies would need to be performed to validate this.

Results

Protein-Coding Targets Regulated by mutant KRAS signaling

To validate the effects of expressing oncogenic KRAS on global gene expression, we first sought to confirm the altered expression of known or hypothesized targets. One target of interest in particular was the programmed death ligand 1 (PDL-1). PDL-1 serves to modulate the adaptive immune response by binding to the PD1 receptor on activated T-cells to inhibit their activity[27]. In the context of cancer, this is particularly important as immune system evasion is a common occurrence in drug resistant cancers. Prior to these experiments, bulk RNAseq was implemented to identify potential targets and key dysregulated genes. PDL-1 was one notable finding, where data would indicate marginal changes in mRNA content for this gene (not published).

To assess the expression pattern of this transmembrane protein, I employed fluorescent light microscopy and immunofluorescent methods to compare the physical expression of PDL-1 between KRAS transformed cells and control cells (Figure 2.1). Interestingly, PDL-1 appeared to have an intracellular local in contrast with its predominantly transmembrane localization, potentially due to mutant KRAS signaling. Further experiments to validate this observation were attempted via the use of flow cytometry, but were inconclusive (data not shown).

In addition to PDL-1, we also found increased expression of DDX60 (Figure 2.2), which we believe pertains to the viral mimicry phenotype prescribed by aberrant KRAS signaling. DDX60 is an RNA helicase with antiviral properties. It is widely implicated in a number of cellular processes, such as translation initiation and splicing[28]. These early findings would go on to aid in the formation of a hypothesis aimed at understanding how KRAS signaling induces an interferon response in lung airway cells, as evident by the produced Cell Reports paper. Investigation into the localization and expression of DDX60 was carried out in the same manner described for PDL-1.

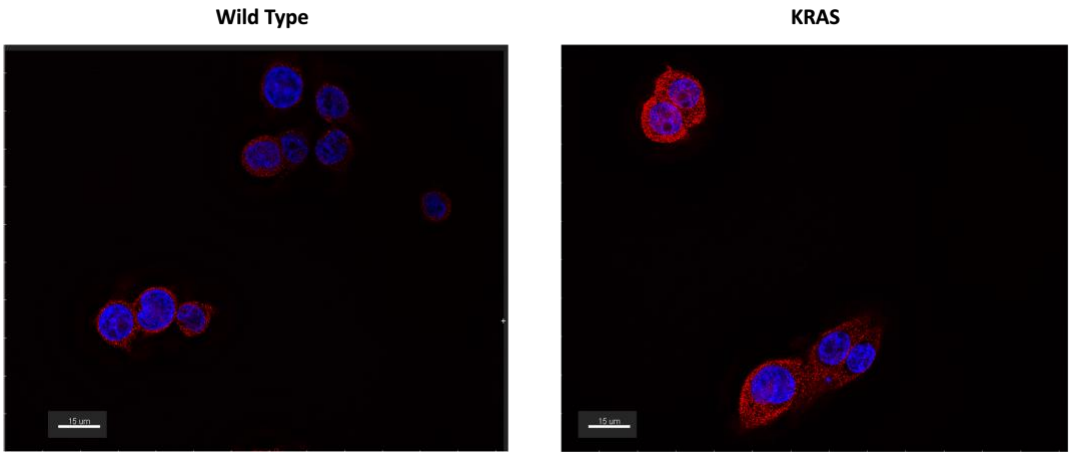

Figure 2.1: PDL1 protein expression between mutant KRAS vs. control AALE cells. Left panel, wild type AALE cells immune-stained for PDL-1, right panel, mutant KRAS containing AALE cells immune-stained for PDL-1.

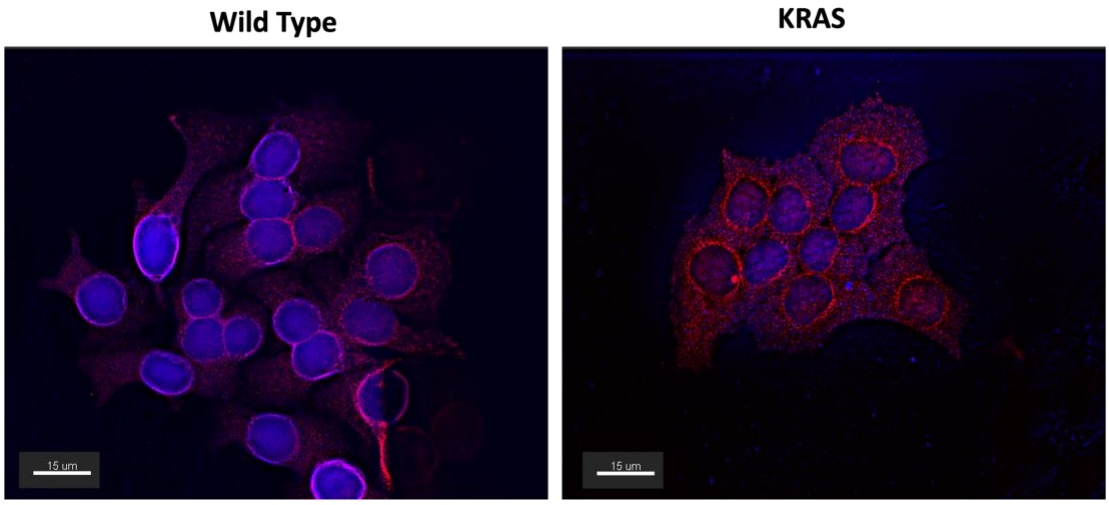

Figure 2.2: DDX60 protein expression between mutant KRAS vs. control AALE cells. Left panel, wild type AALE cells stained for DDX60, right panel, mutant KRAS containing AALE cells stained for DDX60. Matched exposures were carried out between experimental conditions to control for staining variability.

Non-Coding RNAs Regulated by mutant KRAS

Our next primary focus was to validate the expression and localization of some long non-coding RNAs known or hypothesized to be regulated by KRAS signaling, via primary literature and preliminary results respectively. The targets observed were XIST, MALAT, NEAT1 and a novel long non-coding RNA known as lnc00707, nick named SPECTER. We observed defined patterns of gene expression and localization for the aforementioned non-coding RNAs. Specifically, we found XIST to be downregulated in cells possessing mutant KRAS and MALAT, NEAT1 and SPECTER to be upregulated. XIST, NEAT1 and MALAT demonstrated nuclear localization as per previous literature[24–26] and SPECTER appeared to have both nuclear and cytosolic localization (Figure 2.3) in both experimental conditions. Taken together, these results reaffirm previously recorded data from the literature and shed light on the expression and localization of novel long non-coding RNAs that appear to be regulated via oncogenic KRAS signaling.

.

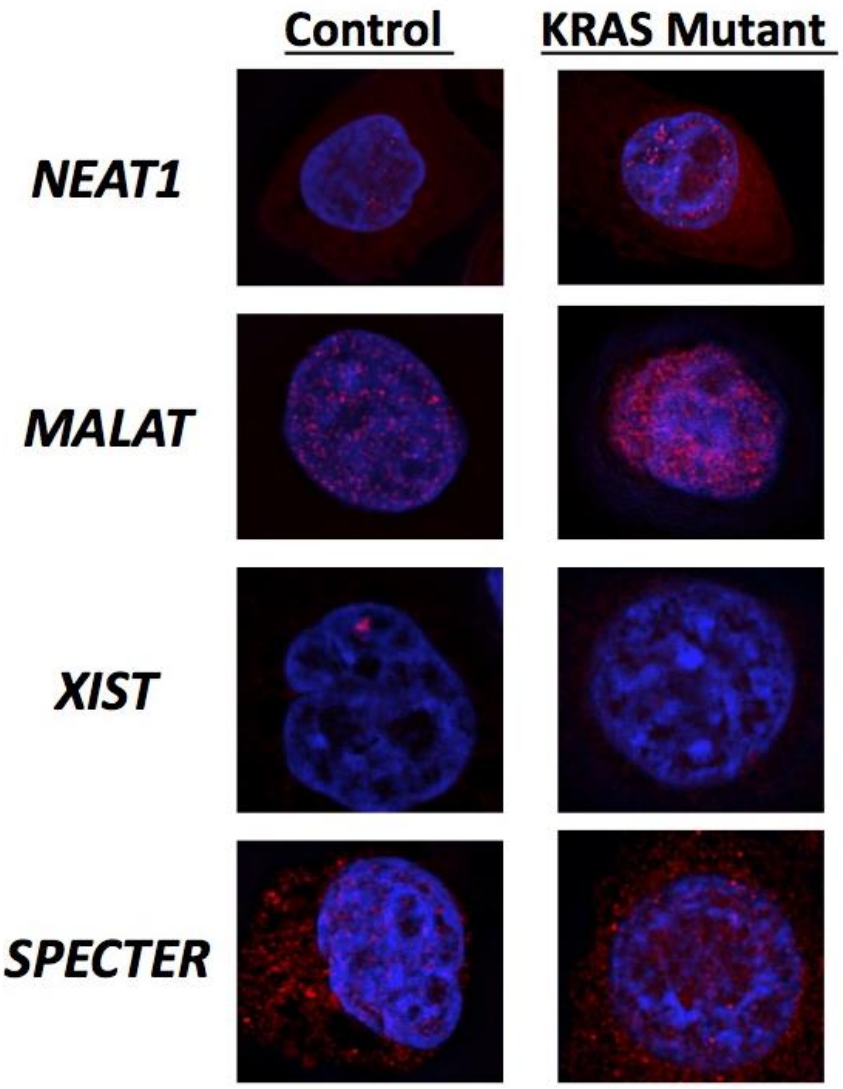

Figure 2.3: Differential expression of the non-coding RNAs XIST, MALAT, NEAT1 and SPECTER via RNA FISH in AALE Cells. Top to bottom, left column is RNA FISH for NEAT1, MALAT, XIST and SPECTER. Right column represents AALE cells harboring mutant KRAS stained in same order via RNA FISH. Matched exposures were used between conditions to control for staining variability.

Materials and Methods

Cell Lines

Immortalized lung epithelial cells (AALE) were obtained from the laboratory of Eric Collison (University of California, San Francisco). The AALE stable cell lines pBABE-mCherry Puro (control) (Lu et al., 2017) and pBABE-FLAG-KRAS(G12D) Zeo (mutant KRAS) were generated using retroviral transduction, followed by selection in puromycin or zeocin. Cells were cultured at 37C and 5% CO2 in SABM Basal Medium (Lonza SABM basal medium, CC-3119) with supplements and growth factors (Lonza SAGM SingleQuots Kit Suppl. & Growth Factors, CC- 4124).

Single Molecule RNA FISH

Complementary fluorescent oligos for MALAT-1, NEAT-1, XIST and SPECTER were obtained from Biosearch Technologies. The protocol for single molecule RNA FISH is derived from the Biosearch Technologies Stellaris recommended guidelines. Briefly, cells were grown in 6 well plates on acid washed 18mm round coverslips. Media was aspirated and cells were rinsed twice with PBS. Cells were then fixed in 1ml of 3.7% formaldehyde and incubated for 10 minutes at room temperature. Formaldehyde was removed and cells were rinsed twice with PBS. Cells were then permeabilized for 1 hour at 4 degrees Celsius in 70% ethanol. Next, ethanol was aspirated and 1ml of wash buffer A was applied to the cells for 5 minutes. After the wash,

100ul of hybridization buffer containing 1ul of probe was applied to the cells by placing the cell side of the coverslip face down into 100ul of hybridization buffer containing appropriate probe on paraffin. Coverslips were then placed in a humidified chamber (150mm cell culture dish with moist paper towel) and sealed with paraffin and incubated at 37 Celsius for 4 hours. After the incubation cells were transferred to a new well plate and washed with wash buffer A for 30 minutes at 37 Celsius. Wash buffer A was aspirated and 1ml of DAPI stain was applied and incubated again at 37 Celsius for 30 minutes. After, DAPI solution was aspirated and wash buffer B was applied for 5 minutes. Slides were mounted using vectashield medium cell side down. Nail polish was used to seal coverslips to the microscope slide and allowed to dry for 10 minutes in the dark.

Immunofluorescence

Briefly, cells were grown on acid washed coverslips (or sectioned spheroids on microscope slides) were rinsed with PBS. Fixation was then performed using 4% PFA for 10 minutes. PFA was then aspirated and 3 washes with PBST followed, each for 5 minutes. Slips and or slides were then blocked using 3% BSA in PBST for 1 hour at room temperature. An additional 3 washes followed. Primary antibody (see table) was incubated in blocking buffer (5% BSA in PBST) overnight with slips or slides at 4C. The following day, blocking buffer and primary antibody were washed away with PBST each

for 5 minutes. Secondary antibody in blocking buffer was added to the slides and slips and allowed to incubate 1 hour at room temperature. 3 additional washes proceeded, each for 5 minutes. DAPI was added at a dilution of 1:10,000 and allowed to incubate in the dark for 5 minutes. Lastly, DAPI was removed using 3 washes of PBST, 5 minutes each. Slides and slips were mounted using vectashield anti-fade mounting medium and sealed with clear nail polish.

Microscope Imaging

All images were taken on a Zeiss Axioimager. Exposures were matched between experimental and control conditions to ensure validity of results. Multiple color channels were used in the case of multiplex experiments. Unless otherwise noted, all images were taken at 60x magnification. Multiple images were taken per sample for the production of a 3D stack and subsequent image processing. Stack numbers were matched between samples.

Image processing

Image processing was performed first by utilizing the imaging software Autoquant to execute 3D deconvolution on czi. Image stacks. Following 3D deconvolution, fully processed image stacks were exported to Imaris. In imaris, image processing consisted of the placement of scale bars and color coordination. Images were matched for their changes in adjusted color contrast and brightness to maintain experimental integrity. In the cases imaris was unavailable, similar adjustments were made in FIJI.

Discussion

My findings indicate a strong regulatory effect via KRAS G12V signaling on the long non-coding RNAs: XIST, MALAT1 and NEAT 1. A marginal effect on the long non-coding RNA dubbed “specter,” and an inconclusive effect on the cellular localization of PDL-1. Each of the above targets are known to play key roles in cancer progression and metastasis. Combined with the aberrant proliferative signals prescribed by oncogenic KRAS signaling, uncoupling of each cellular process becomes a significant issue worthy of future study.

One such study may be uncovering which cancers harboring abnormal expression of the above genes in a similar pattern to that seen in this study, also have mutations in KRAS. The question of whether mutant KRAS signaling is necessary for the pattern observed above is also of interest, and loss of function studies in the form of new KRAS inhibitors or small interfering RNA studies may be implemented to address these questions.

CHAPTER 3: Hypoxia and KRAS signaling work in concert to drive "Stemness" in a lung cancer model

Introduction

This work aimed to determine how properties of the intratumoral microenvironment can confer stem cell properties to lung cancers with oncogenic KRAS. Recent advances in the treatment of many cancers have led to improved survival rates over the past decade. However, cancer metastasis and tumor recurrence remain the chief cause of cancer-related mortality[4]. Cancer recurrence is believed to be a result of the propagation and migration of a subset of specialized cancer cells within a tumor[2]. These specialized cancer cells often possess a higher degree of "stemness," i.e., a dedifferentiated state that distinguishes them from other surrounding cancer cells within the tumor[2].

In addition to being able to repopulate a tumor, these dedifferentiated cells are able to resist chemo and radiation therapies due in part to their quiescent nature, and they often express pluripotency-related genes such as OCT4 and SOX2[4]. In induced pluripotent stem cell (iPSC) reprogramming models, expression of these factors in differentiated cells results in the acquisition and maintenance of stem cell identity[5]. However, whether comparable mechanisms of epigenetic reprogramming contribute to the acquisition and maintenance of stemness in lung cancers has yet to be determined.

Recent evidence suggests that cancer cells can express the pluripotency-related factors SOX2 and OCT4 when exposed to conditions arising within the intratumoral microenvironment, such as hypoxia[29]. The hypoxia regulatory network depends on the expression and stabilization of hypoxia-inducible factors (HIFs). In the absence of oxygen, prolyl hydroxylase domain containing protein 2 (PHD2) is unable to hydroxylate two proline residues on HIFs, preventing them from entering a von Hippel–Lindau tumor suppressor (pVHL)-mediated degradation pathway[30].

Likewise, lactic acid can interfere with the PHD2 reaction in a similar manner. Following lactate import into the cell via monocarboxylate transporter 1, lactate is converted to pyruvate via lactate dehydrogenase. Accumulation of lactate derived pyruvate results in pyruvate acting as a competitive inhibitor with 2-oxoglutarate for the PHD2 reaction[30,31].

Stabilization of HIF via lactic acidosis and hypoxia results in its accumulation and dimerization with the HIFb subunit, allowing them to translocate to the nucleus, where they are believed to have more than 200 transcriptional targets[32]. However, how readily this hypoxia response can elicit the expression of stem cell pathways may depend on the mutational background of lung cancers.

One of the more frequently mutated genes associated with tumorigenesis is the protooncogene KRAS. In lung cancer, the presence of mutant KRAS is often associated with poor clinical outcome[20] and may also be overrepresented in people who smoke compared to never-smokers[33]. KRAS signaling also shares many of the same downstream gene targets with the hypoxia response pathway[34]. Shared gene targets include pathways that control proliferation, metabolism and differentiation. Furthermore, current studies have demonstrated that oncogenic KRAS can facilitate increased expression of HIFs and vice versa, suggesting a positive feedback loop when both are present[35].

For these reasons, I am interested in understanding how oncogenic KRAS signaling may coordinate synergistically with the hypoxia response pathway to reprogram lung cancer cells to obtain a more stem cell-like state. To address these questions, I used a 3D in vitro cell culture systems to derive TS from A549 lung cancer cells with oncogenic KRAS. These TS recapitulate the intratumoral conditions of hypoxia and acidosis observed in solid tumors. I then used single cell RNA sequencing and functional assays to characterize and identify subpopulations that possess hypoxia-driven stemness. I believe this study has significant implications for the potential development of novel therapies associated with lung cancer progression and metastasis, as very little is known as to how oncogenic drivers interact with the intratumoral microenvironment to elicit the ability to resist conventional therapies.

Results

Oncogenic KRAS can induce the expression of non-coding RNAs involved in stem cell maintenance and cancer metastasis

To determine if oncogenic KRAS is capable of inducing the expression of genes associated with stemness and cancer resistance, we did the following: first, we transfected non-cancerous lung airway epithelial cells (AALE) with a mutant KRAS plasmid and compared these cells to control AALE cells. We found that the long non-coding RNA MALAT-1 was upregulated 3-fold in mutant KRAS cells. MALAT-1 has been shown to be involved in the maintenance of stem cell properties and is often associated with cancer metastasis and cancer stem cell formation when highly expressed13. To validate these results, I performed single-molecule RNA fluorescence in situ hybridization (FISH) for MALAT-1 (Figure 3.1).

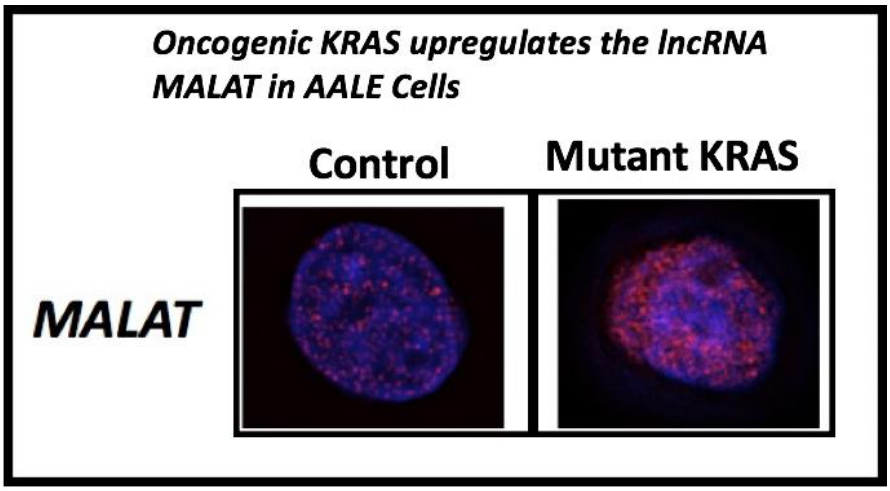

Figure 3.1: MALAT mRNA localization and expression in AALE Cells. Left image is control AALE cells and right is AALE cells harboring mutant KRAS, both stained via single molecule RNA FISH.

A machine learning pipeline can be applied to single cell RNAseq data to characterize stemness

The machine learning pipeline used relies on a linear regression model to derive a stemness index (mRNAsi) based off gene expression data derived from RNAseq. Because of this, a large number of samples are required to get a reliable result. Single cell RNAseq data is ideally suited for the use of this algorithm as each cell is a data point for the described linear regression. To prepare for the use of this model in aim 1, I derived the stemness index for each single cell from a
publicly available (NCBI GEO)[14] glioma single cell RNAseq data set (figure 2b), using the stem cell signature produced from training the model on RNAseq samples derived from stem cell RNAseq datasets (figure 2a).

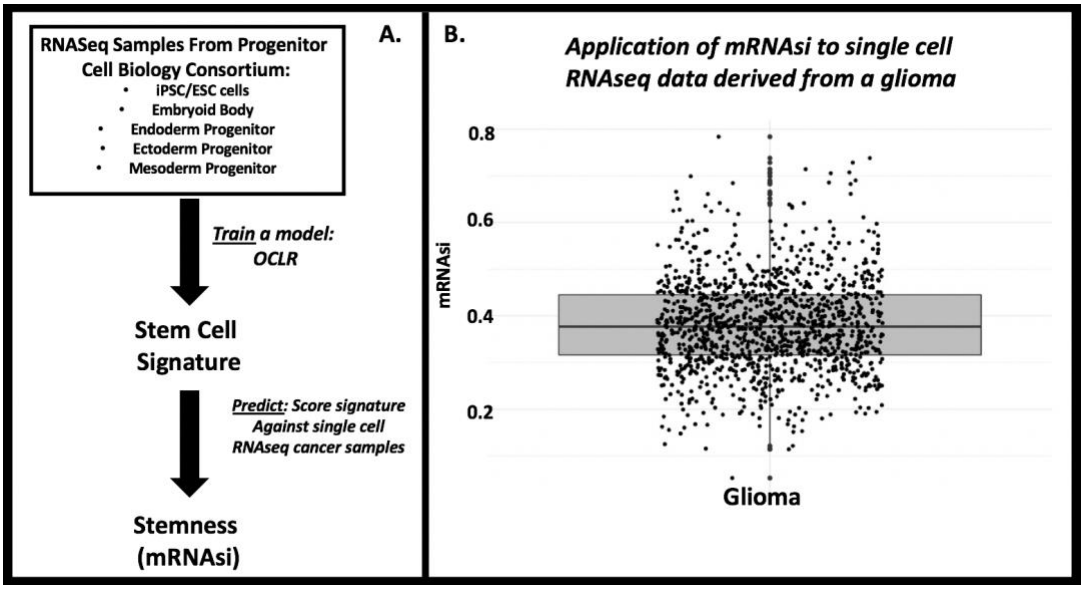

Figure 3.2: Machine learning model and application. A) flowchart describing the machine learning tool and procedure to derive mRNAsi using one class linear regression (OCLR). B) Box plot of mRNAsi for a single cell RNAseq experiment involving a disassociated glioma tumor.

Knocking down endogenous KRAS in human iPSCs results in distinct subpopulations and differential gene expression

As mentioned in earlier chapters, KRAS is the only embryonic lethal RAS isoform. As such, we hypothesized that KRAS signaling may be crucial for lineage commitment in iPSC differentiation models. To determine if knocking down endogenous KRAS results in a diminished pool of pluripotent stem cells or specific lineage commitments, we took advantage of a CRISPRi Gen1c induced pluripotent stem cell line[14]. We then transfected a KRAS guide RNA and performed single cell RNAseq for assessment of lineage commitments after 5 days of growth under KRAS knockdown conditions (>90% KRAS
knockdown). These experiments were compared to control cells transfected with a scramble control guide RNA. I then performed clustering and differential expression analysis in R. The results of this experiment are shown in figure 3. Results reveal distinct subpopulations and altered gene expression in cells with KRAS knockdown compared to control. Genes that play roles in differentiation and metabolism (S100A11 and PHGDH) are some of the more notable upregulated targets in KRAS knockdown iPSCs (figure 3b).

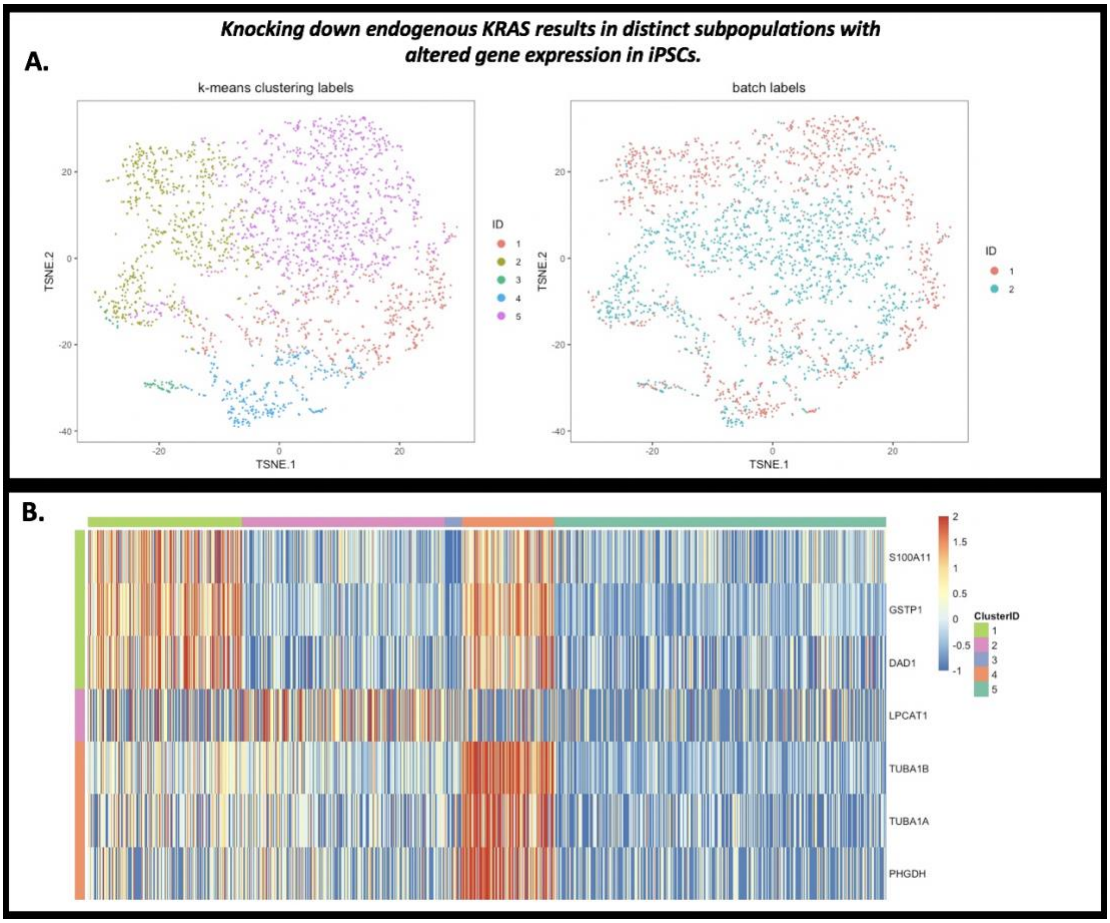

Figure 3.3: Knocking down endogenous KRAS results in distinct subpopulations with altered gene expression. A) TSNE plots of aggregated KRAS knockdown and scramble control samples. Left panel demonstrates 5 distinct subpopulations. Right panel highlights the cells from each sample; ID 1 is KRAS knockdown and ID 2 is scramble control. B) Heatmap demonstrating the most differentially expressed genes for each of the 5 clusters: S100A11, GSTP1, DAD1, LPCAT1, TUBA1B, TUBA1A, PHGDH.

Embryoid bodies derived from iPSCs demonstrate a hypoxic signature when compared to iPSCs grown in 2D conditions with or without endogenous KRAS

To investigate if a hypoxic signature can be ascertained in the 3D conditions prescribed by embryoid bodies, I utilized single cell RNAseq from disassociated EBs derived from iPSCs with KRAS knocked down (compared to control). The hypoxic signature was derived from the GSEA hypoxia hallmark dataset. This gene set consist of more than 200 genes known to be regulated by hypoxia. Due to low sequencing depth, we did not discriminate on gene set composition. I then carried out this same analysis on iPSCs grown in a monolayer to assess if any hypoxic signature is diminished or not present. Embryoid bodies should theoretically possess a hypoxic and necrotic center as with spheroids when grown in 3D conditions, therefore it was prudent that any detected hypoxic signature also be contrasted to that of cells that should possess no hypoxic signature (cells grown in 2D).

Single cell gene expression analysis indicates that a hypoxic signature is detectable only in iPSC embryoid bodies (figure 3.3b), and not iPSCs grown in monolayer (figure 3.3a), supporting the notion of 3D growth conditions being conducive to hypoxic physiology.

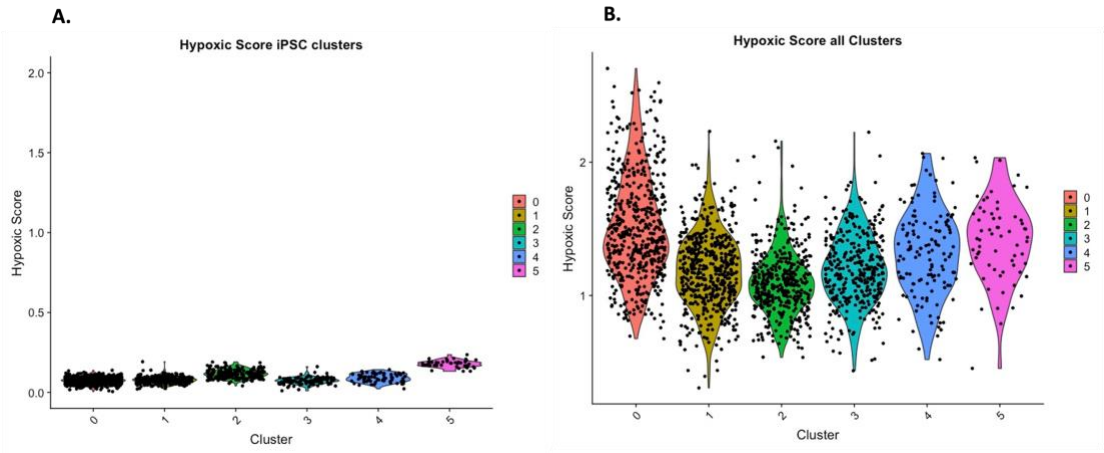

Figure 3.4: Single cell RNAseq iPSC clusters in 2D conditions via Seurat. Violin plots for clustered 2D iPSCs. Y axis is overall hypoxic score and X axis is cluster number.

Embryoid bodies demonstrate no statistical significance between KRAS knockdown and control cells for clusters scoring highest for hypoxia

To assess if there was a statistical significance between single cell iPSCs isolated from embryoid bodies (control vs. KRAS knockdown), I subset cells scoring highest for hypoxia and compared their individual stemness via boxplot (SEM) and corresponding regressions via scatter plot in R. We found there to be no statistical significance between KRAS knockdown cells and control cells in the cluster scoring highest for hypoxia (figure 3.5). There were dramatically less KRAS knockdown cells in hypoxia positive clusters, but it is hypothesized that the reason for this is the overlap in KRAS and hypoxia regulated genes, specifically the MAPK targets within the RAS pathway (figure 3.5).

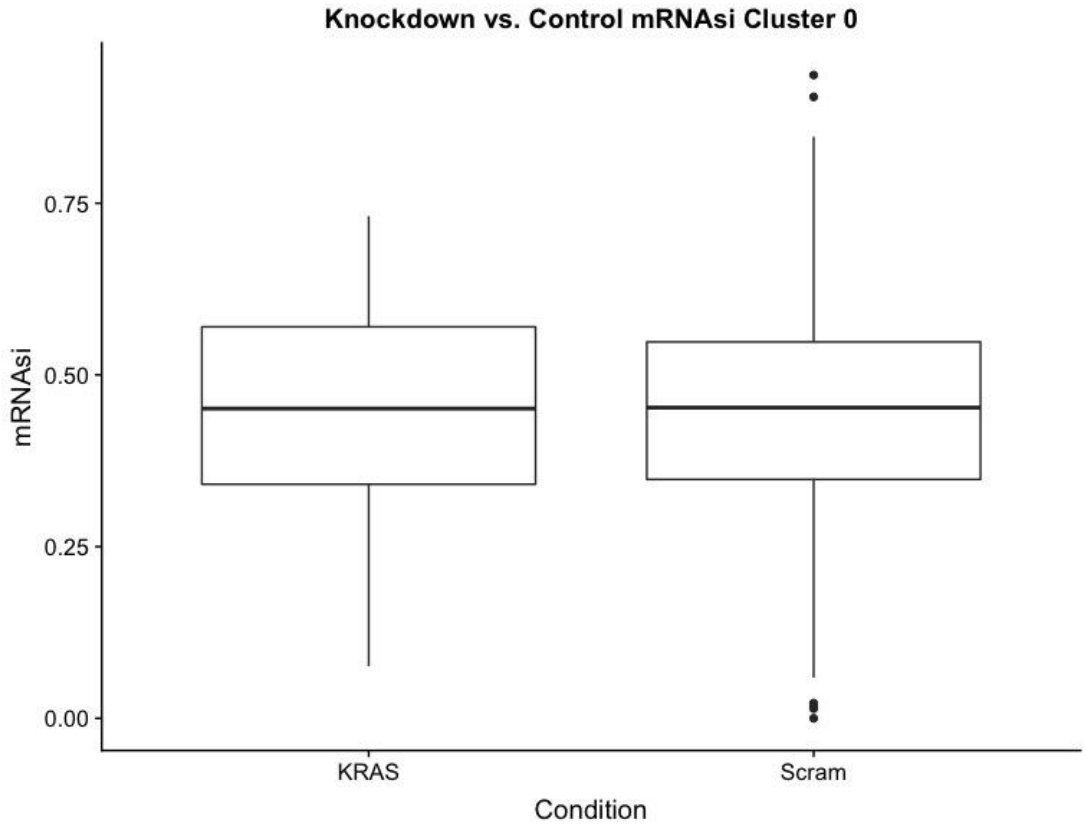

Figure 3.5: Box plot comparing stemness index (mRNAsi) between EB iPSCs in KRAS knockdown vs. control cells. Y axis is mRNAsi via box plot, X axis is experimental condition.

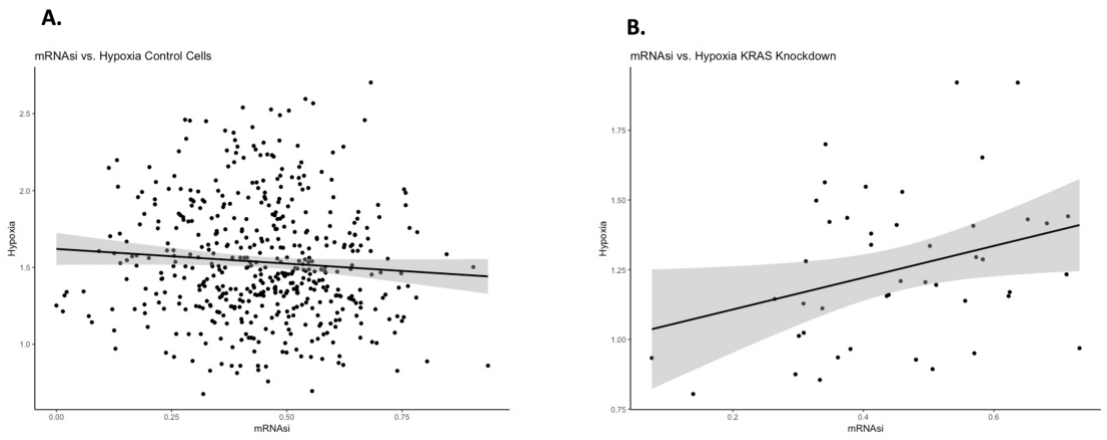

Figure 3.6: Scatter plot with linear regression, stemness index vs hypoxic score between experimental conditions. A) Scatter plot with linear regression of cluster 0 control cells, X axis is stemness index and Y axis is hypoxic score. B) Scatter plot with linear regression of cluster 0 KRAS knockdown cells, X axis is stemness index and Y axis is hypoxic score.

A single cell RNAseq glioma dataset finds no relationship between hypoxia and stemness using a machine learning tool

Before creation of a lung cancer CRISPR cas9 cell line, we found it prudent to first demonstrate that there exists a relationship between cells scoring high for hypoxia and as such, concurrent stemness. To do this, I made use of a publicly available single cell RNAseq data set, the likes of which was derived from a disassociated human glioma tumor. I then implemented the use of a single cell RNAseq analysis tool known as Seurat. Upon clustering of these individual cells (figure 3.4) I then sought to see which cluster scored highest for a hypoxia gene set signature using a hypoxia hallmark gene set from the GSEA database.

Following identification of the high scoring hypoxia cluster, I then utilized the machine learning algorithm across all clusters. Interestingly I found that the cluster scoring high for hypoxia was different from the cluster scoring highest for stemness (figure 3.5). Additionally, I also found that cluster 11 also appeared to have an enriched gene expression pattern for entry to the cell cycle (Figure 3.6), conflicting with the notion that cancer cells with high stemness often demonstrate a quiescent nature. Taken together, the above results indicate a negative relationship between hypoxia gene expression and subsequent stemness in individual cells.

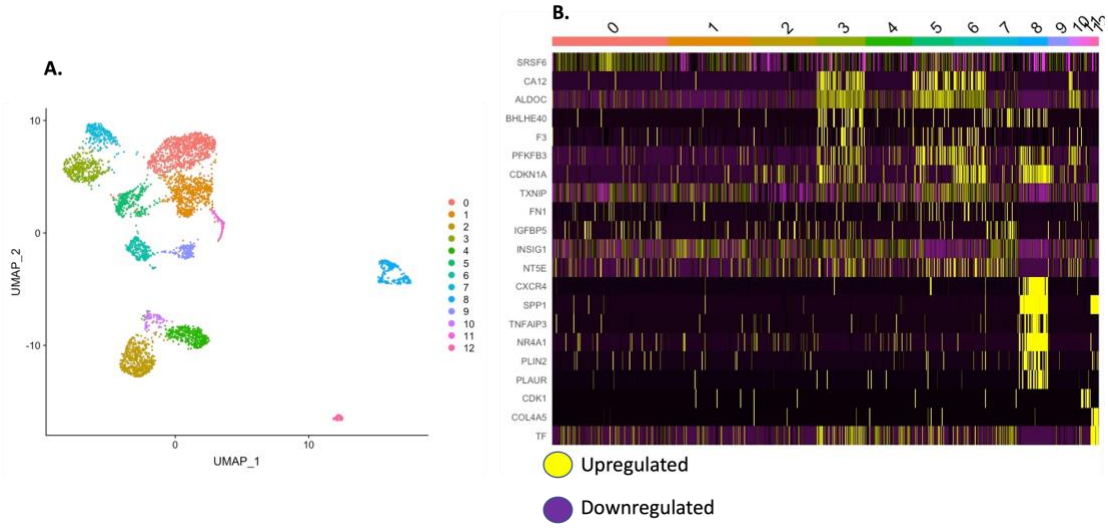

Figure 3.7: Clustering of glioma cells in Seurat and corresponding heatmap for detected hypoxic genes. A) UMAP clustering of glioma cells. B) Heatmap of detected hypoxic genes, Y axis (detected hypoxic genes) X axis is corresponding cluster. Yellow indicates upregulated, purple downregulated.

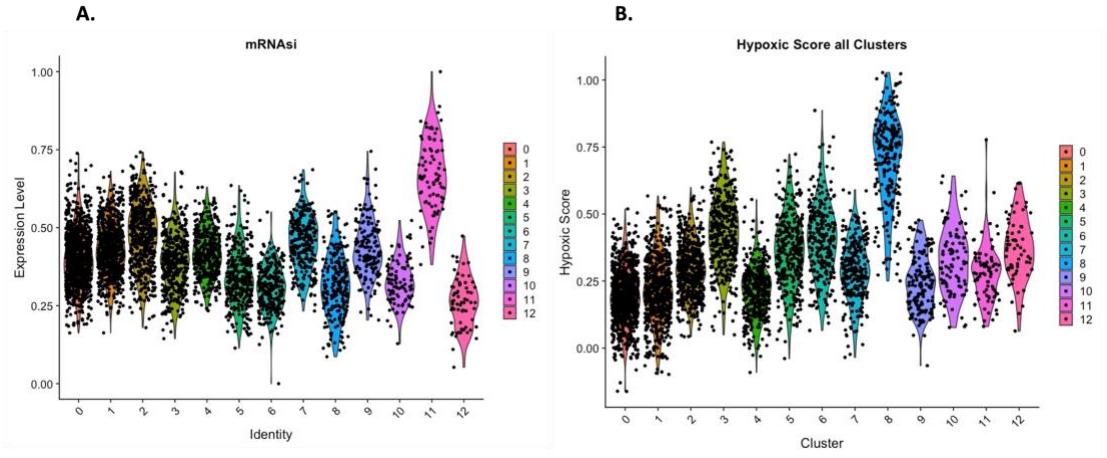

Figure 3.8: Stemness index vs hypoxic signature violin plot with corresponding clusters. A) Stemness index as determined by machine learning algorithm. Y axis is stemness index and X axis is corresponding cluster. B) Hypoxic score across clusters. Y axis hypoxic score and X axis is corresponding cluster.

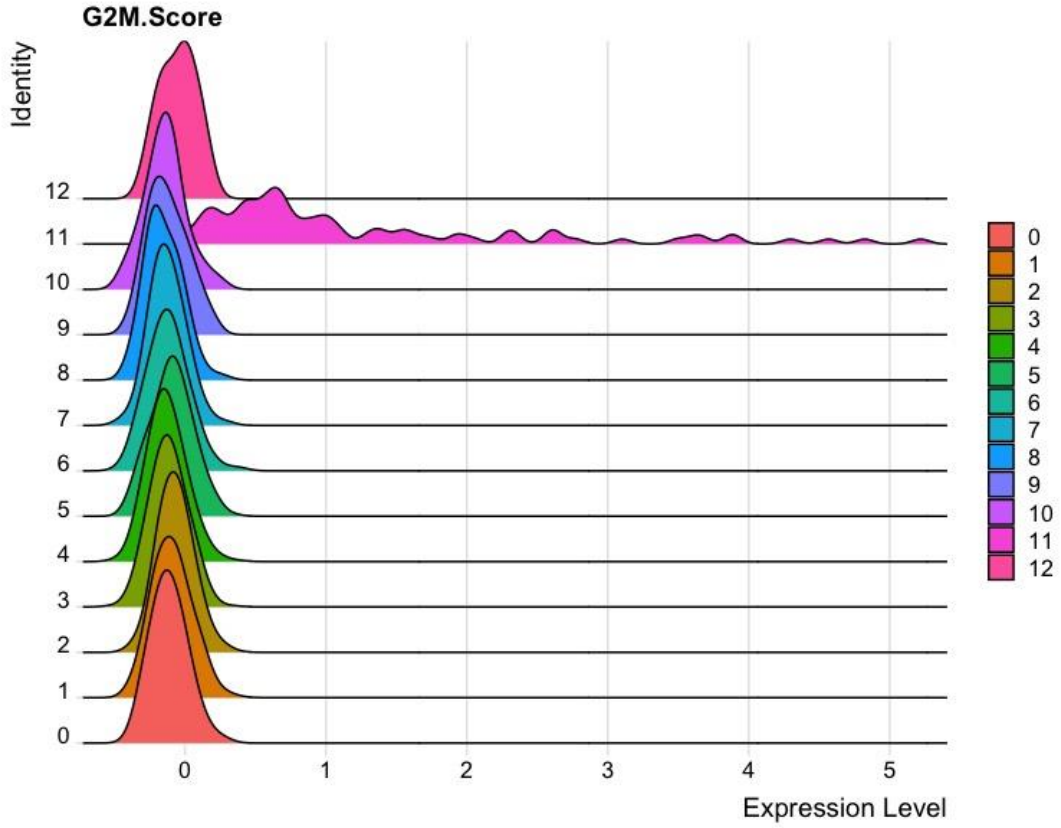

Figure 3.9: Cell cycle entry per cluster as determined by Seurat. Cell cycle entry determined by Seurats own built-in function. X axis is overall expression of a cell cycle entry gene list, X axis is corresponding cluster.

Methods and Materials

Glioma Single Cell Dataset

The disassociated glioma dataset was obtained from the GEO website as referenced[14].

Single Cell Embryoid Body Single Cell Dataset

The embryoid body dataset was produced in lab by a fellow graduate student Sree Maroli. Embryoid bodies (EBs) were derived from iPSCs, grown in low adhesion conditions and allowed to spontaneously differentiate. EBs were then disassociated and subsequent single cell protocols were followed according the 10X manufacturer.

Machine Learning Tool

The machine learning tool used to score for stemness or mRNAsi, was obtained from the following paper[13]. The tool was implemented in R according to the authors own code, using the R packages mentioned in the source paper[13].

Single Cell RNAseq Analysis

Single cell RNAseq for embryoid bodies was performed using 10X genomics manufacture kits and protocols. Alignments and counts were produced using 10x genomics own computational tool CellRanger in UCSC's computational

cluster Hummingbird. Data was then exported to R for subsequent analysis using the R package Seurat. Plots and other tables were produced either with Seurats own functions or ggplot2 in R.

Discussion

Our findings indicate a disconnect in the canonically accepted notion that hypoxia tunes a stemness identity in cancer, at least in the case of a single disassociated glioma. Additionally, we found no correlation in stemness between cells scored high for a hypoxia signature in embryoid bodies derived from iPSCs. It may be that the ability for hypoxia to provide a stem cell like signature is context dependent between cancers and may require the presence of one or more driver mutations, like KRAS. Though the iPSC data shown above does not support a KRAS connection, the cancer connection may still be of interest.

What's more, it may be of significant clinical significance to determine the conditions necessary for hypoxic conditions to provide some modicum of stem cell identity. That is, determining what conditions are necessary both endogenously and exogenously to provide stem cell gene expression signatues in the presence of hypoxia. This work, of course is beyond the scope of the current PhD project but may be of interest for future investigators. The limitations of the study outlined in this chapter are also significant, as single cell RNAseq is not as sensitive to relatively low

expressed transcripts. As such, there may be several hypoxia related genes that were unaccounted for that may have contributed to a more concrete understanding of hypoxia's relationship to both KRAS signaling and stemness as determined by the machine learning tool.

Another limitation comes from the fact that stem cell identity is more than mRNA content. Indeed, the original Yamanaka[5] paper in which iPSCs were first derived relied on epigenetic signatures to better characterize stem cell identify. This same approach was also applied to the machine learning algorithm, where bisulfite sequencing was used as a training set and prediction models were implemented on both mRNA and epigenetic signatures to score for stemness. Taken together, there exist multiple alternative approaches that may be of use for characterizing stemness and discerning the relationship between KRAS signaling and hypoxia signaling.

Chapter 4: Transposable element RNA dysregulation in mutant KRAS(G12C) 3D lung cancer spheroids

Introduction

Transposable element (TE) RNAs are recurrently dysregulated in the context of cancer[36]. In mutant KRAS lung cancer cells, KRAB zinc-finger (KZNF) genes are broadly downregulated, leading to the aberrant upregulation of TE RNAs derived from LINE, SINE, and LTR elements[8,37]. In addition to TE RNAs, long noncoding RNAs (lncRNAs) are also coordinately regulated with RAS signaling genes[38], and their expression patterns are similarly altered in many cancers[10,39–41]. For TE RNAs, their upregulation in cancer induces a state of viral mimicry, leading to the intrinsic activation of innate immunity genes such as interferon-stimulated genes (ISGs)[8,36,42]. In particular, the Alu family of SINEs are a predominant source of immunogenic TE RNAs[43], which are induced by epigenetic changes caused by DNA methyltransferase inhibitors (DNMTi) or mutant KRAS-mediated KZNF inhibition[8,36,42].

To investigate how a common mutation in KRAS affects the TE RNA landscape in lung cancer cells, we characterized the transcriptomes of 3D lung cancer spheroids that harbor KRAS(G12C) mutations in the presence or absence of mutant KRAS(G12C) inhibitor[44]. We show that KRAS(G12C) signaling is required for the expression of LINE- and LTR-derived TE RNAs,

while KRAS(G12C) inhibition specifically upregulates both SINE-derived TE RNAs and a subset of interferon (IFN)-related genes. Our findings reveal the complex interplay between mutant KRAS signaling and TE dysregulation in lung cancer cells, where a defined set of young AluY elements are upregulated in a mutation-dependent manner.

Results

KRAS(G12C) inhibition alters the coding and noncoding transcriptome

To determine how oncogenic KRAS(G12C) signaling regulates the coding and noncoding transcriptome, we performed RNA sequencing (RNA-seq) on 3D mutant KRAS(G12C) lung cancer spheroids. For RNA-seq, we used a full-length protocol with 5' unique molecular identifiers (UMIs) to enable precise RNA counting while mitigating PCR amplification biases[45] (Figure 4.1). We compared the transcriptomes of H358 lung cancer spheroids treated with the KRAS(G12C) inhibitor ARS-1620 (ARS) to control spheroids (DMSO-treated) (Figure S1A) and saw that ARS treatment substantially reduced the levels of phosphorylated ERK (p-ERK) (Figure S1B), indicating that ARS treatment was inhibiting downstream KRAS(G12C) signaling. The suppression of KRAS(G12C) signaling by ARS treatment was also evidenced by reductions in spheroid size and cell viability (Figure S1C and S1D).

At the RNA level, we assessed the relative abundances of different biotypes and TE superfamilies in ARS- or DMSO-treated 3D lung cancer spheroids, which revealed dynamic changes in TE RNA composition upon KRAS(G12C) inhibition (Figure 4.1). While we detected only 32% of GENCODE-annotated protein-coding genes and 15% of lncRNA genes, 92-96% of TE superfamilies (92% LTR, 95% SINE, 96% LINE) were represented in our UMI-tagged RNA-seq data, revealing broad dysregulation of TE RNAs in the KRAS(G12C)-driven transcriptome. Up to a quarter of all detected RNA molecules in ARS-treated lung cancer spheroids were derived from TEs, indicating that TE RNAs represent a sizable portion of the transcriptional output of mutant KRAS(G12C) lung cancer cells.

We next determined the significantly differentially expressed genes between ARS- and DMSO-treated 3D lung cancer spheroids to identify biological processes that were regulated by oncogenic KRAS(G12C) signaling. Lung cancer spheroids with intact KRAS(G12C) signaling were significantly enriched for genes involved in G2M checkpoint, E2F targets, MYC targets, and mitotic spindle (Figure 4.1). Upon KRAS(G12C) inhibition, however, lung cancer spheroids expressed significantly higher levels of genes involved in oxidation phosphorylation, complement, and the IFN alpha and gamma responses (Figure 4.1), providing further supporting evidence for the involvement of KRAS signaling in the regulation of IFN-related genes[8,37].

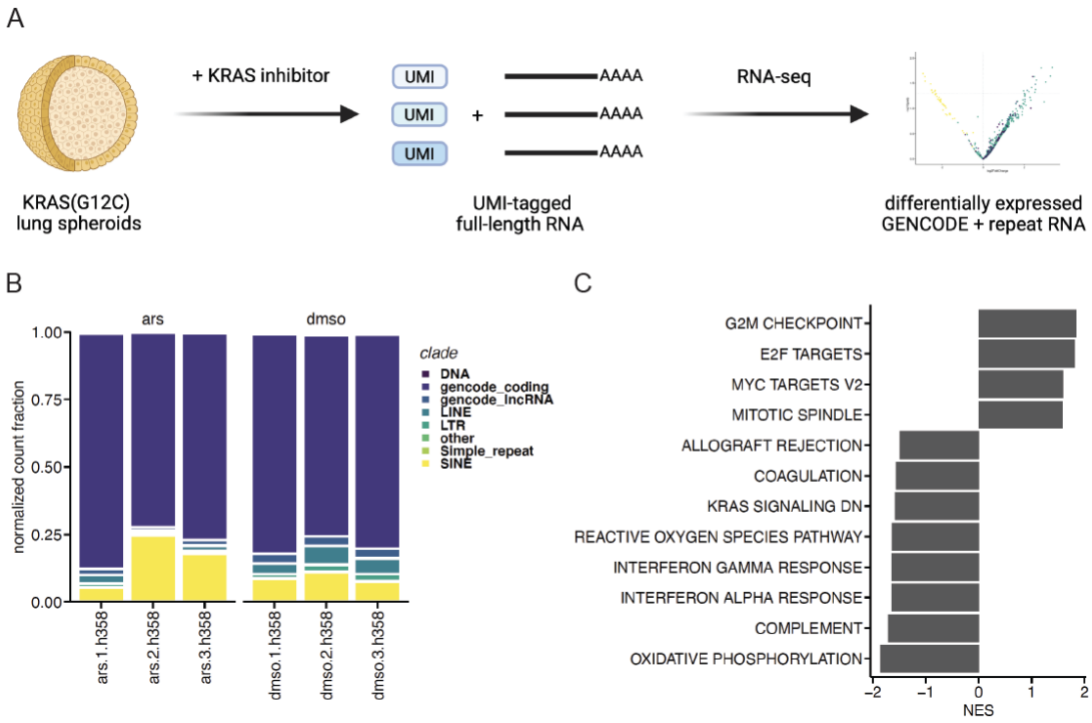

Figure 4.1. KRAS(G12C) inhibition alters the coding and noncoding transcriptome

A. Experimental schematic. **B**. Distribution of counts assigned to GENCODE coding, lncRNA, and TE/repeat superfamilies in ARS-treated (ars) or DMSO-treated (dmso) lung cancer spheroid RNA-seq libraries, where each column represents a biological replicate. **C**. Significant Gene Set Enrichment Analysis results observed in DMSO-treated (right, positive NES) or ARS-treated (left, negative NES) lung cancer spheroids using differentially expressed genes ranked by normalized enrichment score (NES).

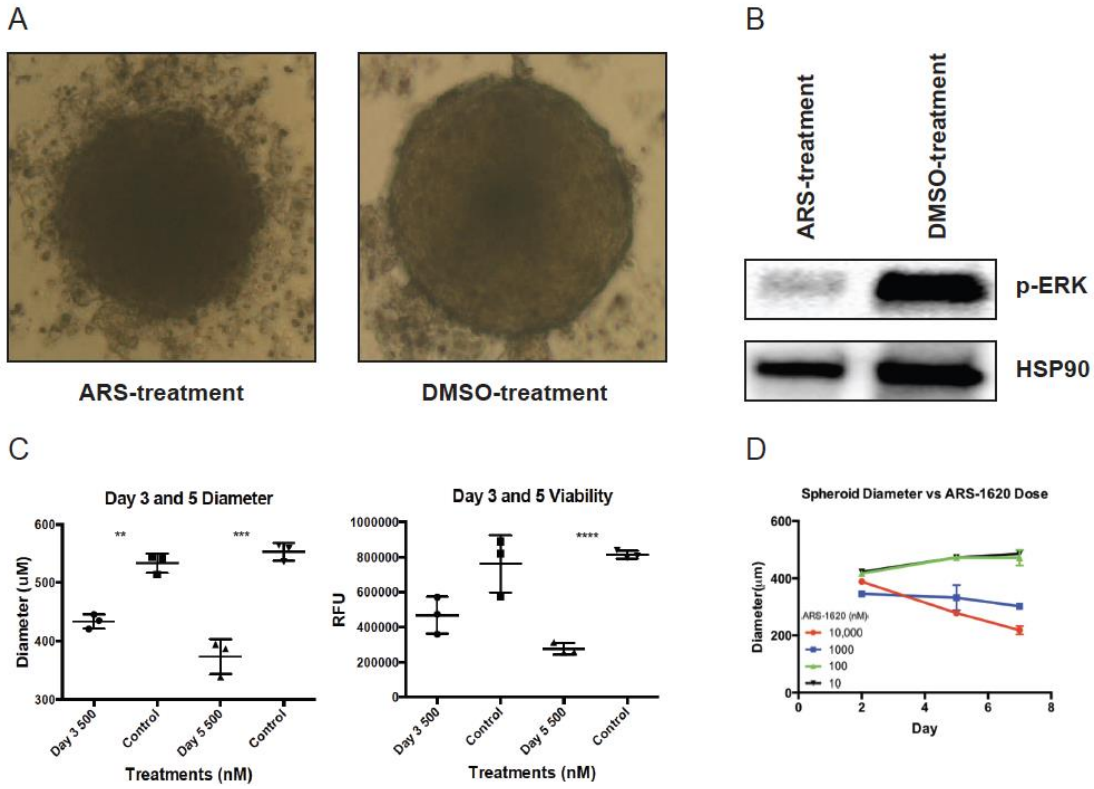

Figure S1: Viability and Spheroid Diameter

A. H358 3D lung cancer spheroids treated with ARS or DMSO. **B**. Western blot for p-ERK and HSP90 using H358 3D lung cancer spheroids treated with ARS or DMSO. **C.** Diameter measurements (in micrometers) (left plot) and cell viability (Cell Titer-Glo® luminescent cell viability in relative fluorescence units) (right plot) of H358 3D lung cancer spheroids treated with ARS or DMSO after 3 or 5 days of treatment (500 nM ARS-1620 or DMSO). **D.** Diameter measurements (in micrometers) of H358 3D lung cancer spheroids treated with different concentrations of ARS-1620 (nM) for 7 days.

KRAS(G12C) inhibition coordinately induces ISGs and young AluY elements

To further elucidate the intrinsic upregulation of ISGs upon KRAS(G12C) inhibition, we identified which genes and TEs were significantly differentially expressed in each gene set and TE superfamily, respectively. Across both IFN alpha and IFN gamma response genes, the most strongly induced gene upon KRAS(G12C) inhibition in lung cancer spheroids was RTP4 (Figure 4.2), a receptor transporter protein that negatively regulates TBK1 signaling[46]. Additionally, the MHC class I complex gene beta-2-microglobulin (B2M), which is recurrently inactivated in lung cancer[47], was also significantly upregulated in both IFN-related gene sets in ARS-treated lung cancer spheroids (Figure 4.2).

Given the direct role for oncogenic KRAS signaling in TE RNA regulation[8], we investigated which subfamilies of TE RNAs were dependent on mutant KRAS(G12C). In lung cancer spheroids with intact KRAS(G12C) signaling, we found that the LINE subfamilies L1M6B and L1PA12 were highly expressed and dependent on KRAS(G12C), as evidenced by their downregulation in lung cancer spheroids treated with ARS (Figure 4.2). Moreover, while only a single DNA subfamily MER44D was dependent on KRAS(G12C) signaling, over a dozen LTR subfamilies were regulated by mutant KRAS(G12C), including MLT1A0-int, LTR51, MER50B, and LTR1B0 (Figure 4.2). In contrast, SINE subfamilies were all significantly upregulated

upon KRAS(G12C) inhibition and coordinately induced with specific ISG genes (Figure 4.2) in lung cancer spheroids. These SINE TE RNAs were all derived from the AluY subfamily (Figure 4.2), indicating that KRAS(G12C) inhibition dysregulates a specific subset of young AluY elements.

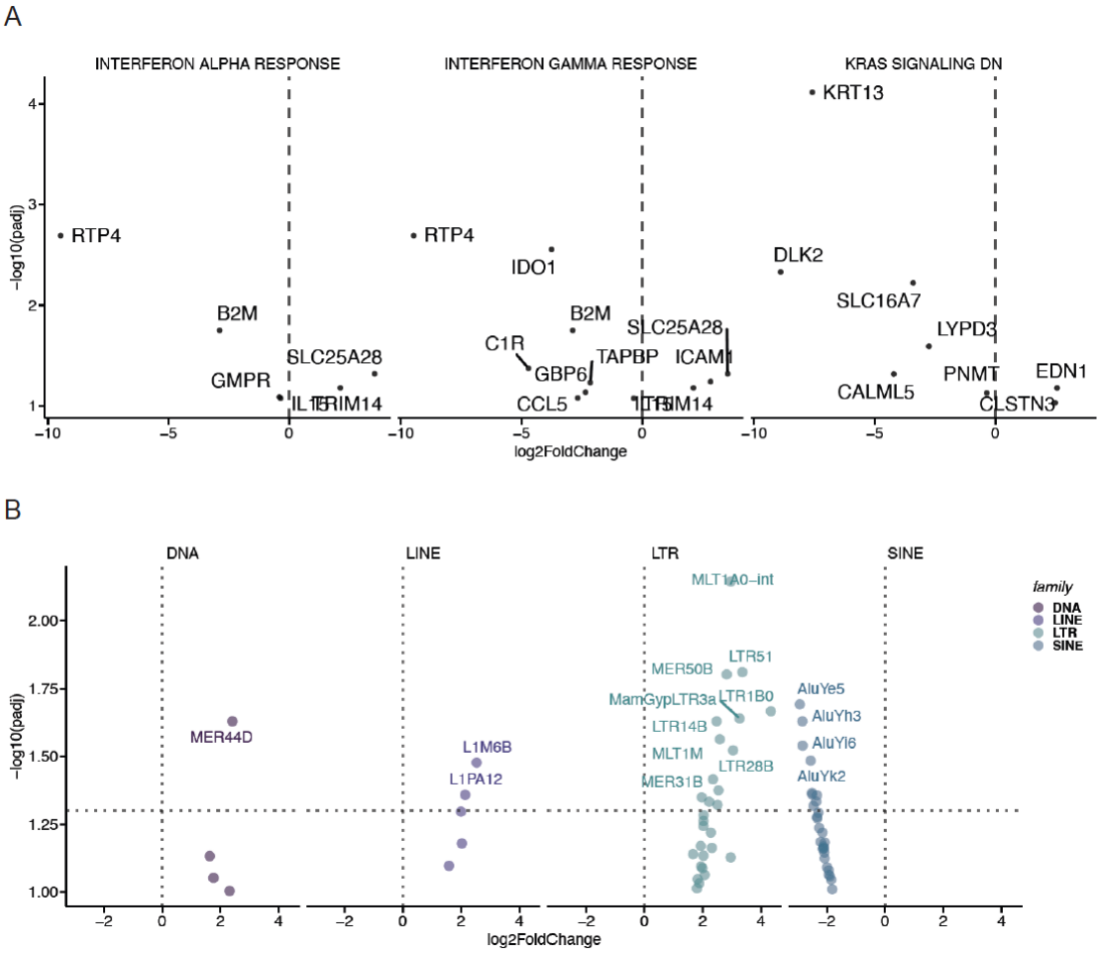

Figure 4.2. KRAS(G12C) inhibition coordinately induces ISGs and young AluY elements

A. Volcano plots depicting significant differential expression observed in key gene sets between DMSO-treated (right, positive fold-change) or ARS-treated (left, negative fold-change) lung cancer spheroids. **B**. Volcano plots depicting significant differential expression observed in TE superfamilies between DMSO-treated (right, positive fold-change) or ARS-treated (left, negative fold-change) lung cancer spheroids.

KRAS(G12C) inhibition downregulates long noncoding RNAs

To further elucidate the effects of KRAS(G12C) inhibition on the transcriptome, we examined all significantly differentially expressed lncRNAs in lung cancer spheroids treated with ARS or DMSO control. We found that a large number of lncRNAs were dependent on KRAS(G12C) signaling for their expression, as many of these lncRNAs were significantly downregulated upon KRAS(G12C) inhibition (Figure 4.3). Three of these downregulated lncRNAs, AC114546.3, NCMAP-DT, and AC073575.2, have no known functions but overlap in an antisense orientation to the coding genes ZNF770, RCAN3, and ERP29, respectively. As a class of noncoding RNA, we observed broad downregulation of lncRNAs upon ARS treatment in lung cancer spheroids (Figure 4.3), indicating that many lncRNAs are dependent on KRAS(G12C) signaling for their expression.

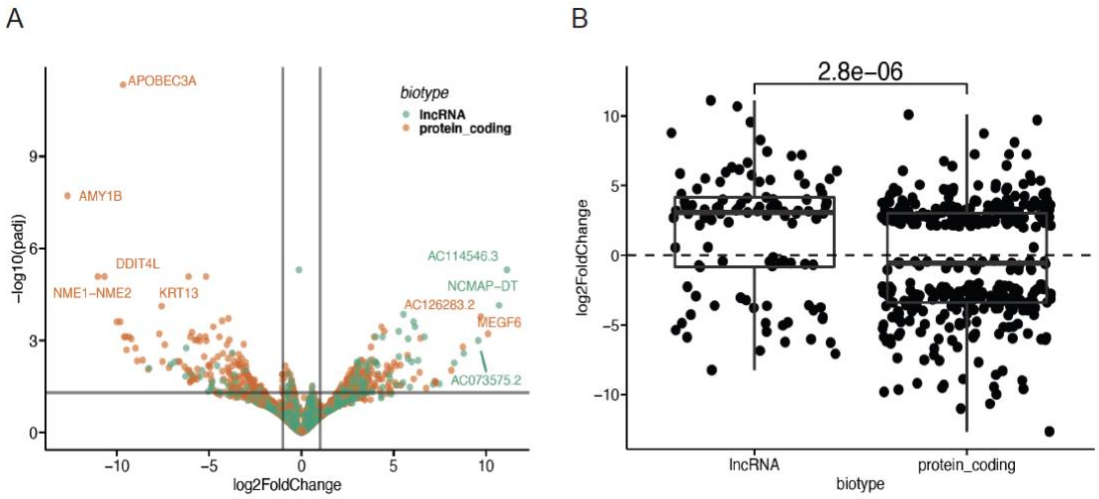

Figure 4.3. KRAS(G12C) inhibition downregulates long noncoding RNAs

A. Volcano plot of significant differential expression of GENCODE protein-coding RNAs and lncRNAs between DMSO-treated (right, positive fold-change) or ARS-treated (left, negative fold-change) lung cancer spheroids. **B.** Box plot of significant differential expression of GENCODE protein-coding RNAs and lncRNAs between DMSO-treated (top, positive fold-change) or ARS-treated (bottom, negative fold-change) lung cancer spheroids (Wilcoxon).

Methods and Materials

Cell lines

H358 lung cancer cell lines containing the KRAS(G12C) mutation were cultured in RPMI 1640 medium (Invitrogen) supplemented with 10% fetal bovine serum (Sigma) at 37°C, 5% CO2 in a humidified incubator. All cell lines tested negative for mycoplasma. Cell lines were purchased from American Type Culture Collection (ATCC).

Cell viability assays

For spheroid viability assays, 10,000 cells/well were seeded in low adhesion round bottom 96-well plates and incubated at 37°C, 5% CO2 for 24 hours. Then serially diluted ARS-1620 or DMSO were added to the cells, and plates were incubated in standard culture conditions for 72 hours, with fresh ARS and DMSO media being replaced daily. Cell viability was measured using a Cell Titer-Glo® Luminescent Cell Viability Assay kit (Promega) according to the manufacturer protocol. The luminescence signal of ARS-treated samples was normalized to DMSO control. Luminescence was measured on a SpectraMax iD3 molecular device.

RNA isolation

Total bulk RNA was isolated from approximately 100 H358 spheroids (per condition) using Quick-RNA Mini-Prep kit (Zymogen) according to the manufacturer protocol. RNA was quantified via NanoDrop-8000 Spectrophotometer.

RNA-seq library preparation

An adapted Smart-seq3 protocol[45] was used to generate RNA-seq libraries from total RNA to count and assess full-length RNA molecules. Briefly, 10ng of total RNA was reverse transcribed using a barcoded oligoDT primer (125nM) followed by template switching with a barcoded template switch oligo (125nM). These oligo sequences served as primers for PCR amplification. The Nextera HT kit (Illumina) was used to convert cDNA libraries into sequencing libraries with the addition of a UMI-specific primer to amplify the cDNA ends containing molecular barcodes as described in the Smart-seq3 protocol. cDNA and library quality were assessed using an Agilent bioanalyzer DNA high sensitivity chip and quantified using the high sensitivity DNA assay on the Qubit 3.0.

Western blot

Approximately 100 H358 spheroids (per condition) were isolated following ARS or DMSO treatment. Spheroids were then incubated on ice in RIPA buffer supplemented with protease inhibitor for 15 minutes. Lysates were then spun at 10,000 RCF for 10 minutes. Supernatant was then transferred to a new tube for subsequent SDS-PAGE sample prep in Laemmli buffer, boiled for 5 minutes at 95 C for a final concentration of 1mg/ml. SDS-PAGE was performed to separate protein by size followed by subsequent transfer to a PVDF membrane. Membranes were incubated with primary p-ERK (CST) and HSP90 (CST) antibodies overnight at 4 C. Secondary antibodies (Abcam) were then incubated on 3 times TBST washed membranes in blocking buffer for subsequent imaging.

RNA-seq analysis

All *fastq* files were trimmed with *Trimmomatic 2 (0.38)*[48] and resulting trimmed files were assessed with *FastQC*[49] and then processed with the following analytical pipeline:

Salmon (1.3.0): pseudoalignment of RNA-seq reads performed with *Salmon*[50] using the following arguments:

--validateMappings –gcBias --seqBias --recoverOrphans --rangeFactorizationBins 4

using an index created from the *GENCODE* version 35 transcriptome fasta file using decoy sequences to enable selective alignment. An additional, TE-aware index was created in a similar fashion but supplemented with sequences generated from the UCSC Repeat Masker track.

DESeq2 (1.32.0): Salmon output was imported into a DESeq object using *tximport*[51] and differential expression analysis was performed with standard arguments[52]. All results were filtered to have padj < 0.05. In the case where R could only generate 0.00 for the padj values, they were reset to the lowest non-zero padj value in the data set. Where count data was used, it was normalized across samples using DESeq.

Gene set enrichment analysis

Differentially expressed genes were ranked by the shrunken log2FoldChange values generated by *DESeq2*. Gene sets were acquired using the *R* package *msigdbr (7.4.1)* and filtered to only contain gene sets with 'Hallmark' status. The *R* package *fgsea (1.18.0)* was used to generate Gene Set Enrichment estimates which were filtered to results with adjusted pvalues < 0.05.

UMI De-duplication

Paired end illumina reads were adapter trimmed using FastP[53] with default settings. UMIs were extracted from the read and moved to the readname using Umi tools extract from the UMI-tools package UMI-tools[54] with the

barcode pattern set to "NNNNNNNN". UMI-removed reads were aligned against HG38 using the STAR aligner with the GENCODE v38 annotation set. Aligned reads were deduplicated using "UMI-tools dedup" with default settings.

Discussion

Here we show that oncogenic KRAS(G12C) signaling is required for the expression of specific TE superfamilies, namely LINE and LTR elements, as well as a subset of lncRNAs, further demonstrating how RAS signaling regulates the noncoding transcriptome[8,8,38]. We also used a 3D lung cancer spheroid model for our KRAS(G12C) inhibitor experiments since 3D models have been shown to more faithfully recapitulate the in vivo drug response when compared to 2D culture models[55]. Furthermore, our application of a UMI-based full-length RNA-seq technique[45] allowed us to more accurately capture TE RNA composition and dynamics in our lung cancer spheroids by removing PCR duplicates in our RNA-seq data.

Our findings are consistent with previous studies of KRAS(G12C) inhibition, where IFN alpha and gamma response genes were upregulated in ARS-treated H358 lung cancer cells[56]. Based on the known immunogenic properties of Alu-derived RNAs, our results suggest that the specific upregulation of young AluY elements upon KRAS(G12C) inhibition is at least in part responsible for the strong upregulation of ISGs in ARS-treated lung

cancer spheroids. Notably, the significant enrichment of IFN-related genes in response to KRAS(G12C) inhibitor treatment does not include the further upregulation of RNA sensor ISGs such as MDA-5, RIG-I, or PKR, which initially become upregulated in lung cells in response to oncogenic KRAS signaling[8,37].

We have previously shown that mutant KRAS signaling alone is sufficient to induce TE RNA upregulation in human lung cells that have been transformed in vitro[8,36], and our results described here extend these observations to lung cancer cells with a different activating KRAS mutation. KRAS(G12D) or KRAS(G12V) mutations both induce the significant upregulation of the LTR12C subfamily in transformed lung cells[8], but we did not see significant enrichment of LTR12C-derived TE RNAs in our KRAS(G12C) lung cancer spheroids. Instead, we saw upregulation of LTR51, LTR1B0, LTR14B, and LTR28B, suggesting that diverse gain-of-function mutations in KRAS regulate different aspects of the TE RNA transcriptome.

Our work provides a comprehensive assessment of how the noncoding/TE RNA transcriptome dynamically responds to KRAS(G12C) inhibition. Future studies may provide new insights into the potential roles of noncoding/TE RNAs in mechanisms of KRAS(G12C) inhibitor resistance. Furthermore, TE RNAs that are secreted from cancer cells upon KRAS inhibition may serve as extracellular RNA biomarkers[8,10,17,37,57] of response and/or resistance to KRAS inhibitor therapies [20].

Chapter 5: Extracellular RNA signatures of mutant KRAS(G12C) lung adenocarcinoma cells (Khoja et. al.)

Introduction

Growing evidence suggests extracellular vesicles excreted from cancer cells hold promise for the identification of novel cancer biomarkers[57]. In the context of cancers driven by mutant KRAS however, unraveling the connection between aberrant KRAS signaling and the secretion of exosomes becomes problematic. This is because traditional modes of gene expression interruption rely on near complete ablation of gene targets with the use of gene knockouts or silencing using siRNA[20]. Such methods are a problem for the study of popular oncogenic targets, as often times disrupting these oncogenic functions leads to cell death[58]. Until recently, cancers with mutations for KRAS remained largely undruggable due to the difficulty of the drug design prescribed for RAS GTPase activity. However, a recent class of small molecule inhibitors targeting KRAS with G12C mutations has recently been unveiled. These inhibitors can be fine-tuned to interrupt key cellular signaling pathways while simultaneously preserving cellular viability[16]. The two primary inhibitors used in this study[17] and the previous are AMG-510 and ARS-1620[16,19,56]. This study aims to understand how KRAS G12C signaling ablation tunes the non-coding transcriptome and what if any of these non-

coding RNAs may hold potential for serving as biomarkers. We found that exposing 2D cancer cells to AMG-510 induces a wide transcriptional change in both coding and non-coding RNA. We also found that of the exported non-coding RNAs, transposable elements represent one of the largest classes representing total RNA content. We believe these results are an important first step in the use and identification of novel non-coding transcripts as potential biomarkers.

Results

My work for this project consisted of aiding and assisting in the development of wet lab work required to identify an AMG-510 treatment concentration that would both sustain cell viability and diminish KRAS G12C signaling. To do this, I aided in the production of a protocol that utilized time course titrations of AMG-510. I also produced the protocol necessary for assessing KRAS G12C signaling inhibition via a phosphorylated ERK immunoblot. Sustained viability is paramount for downstream RNAseq applications, as gene expression patterns and their interpretation are prone to interrupted oncogene functions resulting in subsequent cell death. Phosphorylated ERK immunoblots are also the chief mode of assessing RAS function, as this step in the pathway is tightly regulated and downstream processes that control cell behavior are executed via this final step.

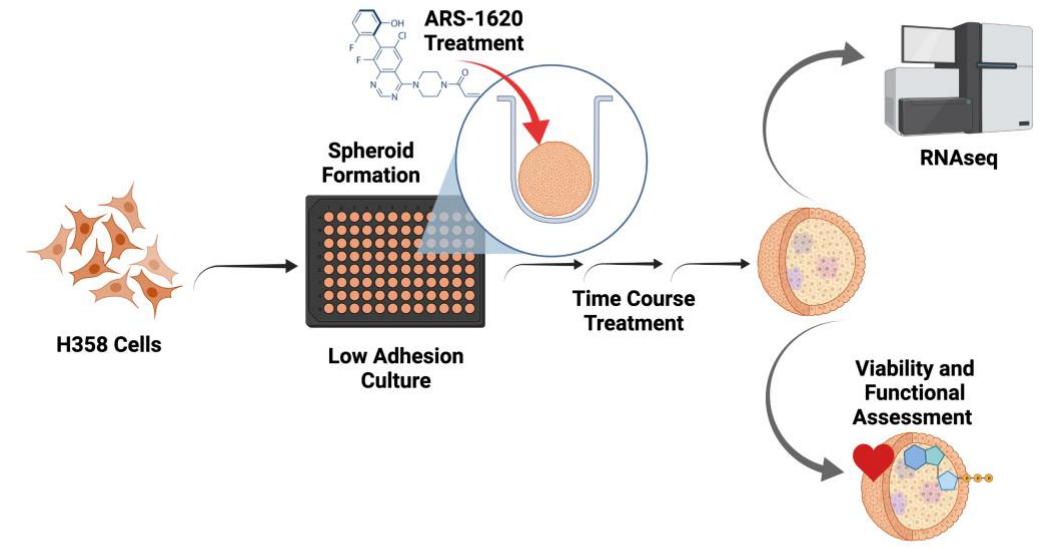

Figure 5: Conceptual diagram of 3D spheroid formation from 2D culture and subsequent experimental pipelines.

Methods and Materials

Cell Lines

H358 lung cancer cell lines containing the KRAS G12C mutation were cultured in RPMI 1640 medium (Invitrogen) supplemented with 10% fetal bovine serum (Sigma) at 37°C, 5% CO2 in a humidified incubator. All cell lines tested negative for mycoplasma. Cell lines were purchased from American Type Culture Collection (ATCC).

Cell viability assays

For spheroid viability assays 10,000 cells/well were seeded in low adhesion round bottom 96-well plates and incubated at 37°C, 5% CO2 for 24 hours. Then serially diluted ARS 1620 and DMSO were added to the cells, and plates were incubated in standard culture conditions for 72 hours, with fresh drug and DMSO media being replaced daily. Cell viability was measured using a Cell Titer-Glo® Luminescent Cell Viability Assay kit (Promega) according to the manufacturer's protocol. The luminescence signal of treated samples was normalized to DMSO control. The luminescence was measured on a SpectraMax iD3 molecular devices.

Western Blot

Approximately 100 spheroids per condition (drug and control) were isolated following the drug treatment time course. Spheroids were then incubated on ice in RIPA buffer supplemented with protease inhibitor for 15 minutes. Lysates were then spun at 10,000 rcf for 10 minutes. Supernatant was then transferred to a new tube for subsequent SDS page sample prep in laemmli buffer, boiled for 5 minutes at 95 degrees Celsius for a final concentration of 1mg/ml. SDS page was performed to separate protein by size followed by subsequent transfer to a PVDF membrane. Membranes were incubated with primary P-ERK (cst) and HSP-90 (cst) antibodies overnight at 4 degrees Celsius. Secondary antibodies (abcam) were then incubated on 3 times TBST washed membranes in blocking buffer. Images were obtained on a biorad gel cabinet.

Discussion

The results of this study indicate a time sensitive manner in which cells harboring KRAS (G12C) mutations must be treated with AMG-510 to provide sustained KRAS signaling ablation and viability. Furthermore, preliminary evidence suggest a adaptation to drug treatment after 5 days of exposure (data not shown). Here I demonstrate a quick and effective mode to identify

the conditions necessary for sustained KRAS signaling interruption and cellular viability using small molecule inhibitors of KRAS G12C. Time course assessments found a sustained effect on KRAS signaling with daily replenishment of the small molecule AMG-510. Further investigation of pathway interruption can be done to substantiate findings indicative of KRAS signaling ablation, such as immunoblots for PI3K, MEK and AKT. This may be of particular interest, as recent findings indicate a fast adaptation to this new class of KRAS G12C inhibitors with regards to these particular pathways[59]. Indeed, identifying points and features of gene expression changes consistent with adaptation to these drugs may also be of interest for filtering of data not related to KRAS signaling alone. Taken together, we find strong potential for an in vitro model of 3D grown cancer spheroids to serve as a tractable and effective means for study of oncogenic pathways using novel cancer therapeutics.

References

1. Cancer Facts & Figures 2023| American Cancer Society. Accessed February 20, 2023. https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/2023-cancer-facts-figures.html

2. Rich JN, Bao S. Chemotherapy and Cancer Stem Cells. *Cell Stem Cell*. 2007;1(4):353-355. doi:10.1016/j.stem.2007.09.011

3. Li S, Feldman MW. Single-cell cloning of colon cancer stem cells reveals a multi-lineage differentiation capacity. 2009;106(23). doi:10.1073/pnas.0904442106

4. Liu A, Yu X, Liu S. Pluripotency transcription factors and cancer stem cells: Small genes make a big difference. *Chin J Cancer*. 2013;32(9):483-487. doi:10.5732/cjc.012.10282

5. Takahashi K, Yamanaka S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*. 2006;126(4):663-676. doi:10.1016/j.cell.2006.07.024

6. Prior IA, Hood FE, Hartley JL. The Frequency of Ras Mutations in Cancer. *Cancer Res*. 2020;80(14):2969-2974. doi:10.1158/0008-5472.CAN-19-3682

7. Moore AR, Rosenberg SC, McCormick F, Malek S. RAS-targeted therapies. *Nat Rev Drug*. Published online 2021.

8. Reggiardo RE, Maroli SV, Halasz H, et al. Mutant KRAS regulates transposable element RNA and innate immunity via KRAB zinc-finger genes. *Cell Rep*. 2022;40(3):111104. doi:10.1016/j.celrep.2022.111104

9. Henry NL, Hayes DF. Cancer biomarkers. *Mol Oncol*. 2012;6(2):140-146. doi:10.1016/j.molonc.2012.01.010

10. Reggiardo RE, Maroli SV, Kim DH. LncRNA Biomarkers of Inflammation and Cancer. *Adv Exp Med Biol*. Published online 2022:121-145 10 1007 978-3-030-92034-0 7.

11. Zeng M, Kikuchi H, Pino MS, Chung DC. Hypoxia activates the K-ras proto-oncogene to stimulate angiogenesis and inhibit apoptosis in colon cancer cells. *PLoS ONE*. 2010;5(6). doi:10.1371/journal.pone.0010966

12. Grimes DR, Kelly C, Bloch K, Partridge M. A method for estimating the oxygen consumption rate in multicellular tumour spheroids. *J R Soc Interface*. 2014;11(92):20131124-20131124. doi:10.1098/rsif.2013.1124

13. Malta TM, Sokolov A, Gentles AJ, et al. Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. *Cell*. 2018;173(2):338-354.e15. doi:10.1016/j.cell.2018.03.034

14. Tirosh I, Venteicher AS, Hebert C, et al. Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature*. 2016;539(7628):309-313. doi:10.1038/nature20123

15. Milanovic M, Fan DNY, Belenki D, et al. Senescence-associated reprogramming promotes cancer stemness. *Nature*. 2018;553(7686):96-100. doi:10.1038/nature25167

16. Janes MR, Zhang J, Li LS, et al. Targeting KRAS Mutant Cancers with a Covalent G12C-Specific Inhibitor. *Cell*. 2018;172(3):578-589.e17. doi:10.1016/j.cell.2018.01.006

17. Khojah R, Reggiardo RE, Ozen M, et al. Extracellular RNA signatures of mutant KRAS(G12C) lung adenocarcinoma cells. *bioRxiv*. Published online January 1, 2022:2022.02.23.481574. doi:10.1101/2022.02.23.481574

18. Campbell JD, Alexandrov A, Kim J, et al. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat Genet*. 2016;48(6):607-616. doi:10.1038/ng.3564

19. Chien S. Targeting the KRAS mutation for more effective cancer treatment. MD Anderson Cancer Center. Accessed February 22, 2023. https://www.mdanderson.org/cancerwise/targeting-the-kras-mutation-for-more-effective-cancer-treatment.h00-159458478.html

20. Carvalho PD, Guimarães CF, Cardoso AP, et al. KRAS oncogenic signaling extends beyond cancer cells to orchestrate the microenvironment. *Cancer Res*. 2018;78(1):7-14. doi:10.1158/0008-5472.CAN-17-2084

21. Significance of Ras Signaling in Cancer and Strategies for its Control. *Oncol Hematol Rev 2015112147–52*. Published online November 23, 2015. Accessed February 22, 2023. https://touchoncology.com/diagnostics-and-screening/journal-articles/significance-of-ras-signaling-in-cancer-and-strategies-for-its-control/

22. Huang X, Xiao R, Pan S, et al. Uncovering the roles of long non-coding RNAs in cancer stem cells. *J Hematol OncolJ Hematol Oncol*. 2017;10(1):62. doi:10.1186/s13045-017-0428-9

23. Glaß M, Michl P, Hüttelmaier S. RNA Binding Proteins as Drivers and Therapeutic Target Candidates in Pancreatic Ductal Adenocarcinoma. *Int J Mol Sci*. 2020;21(11):4190. doi:10.3390/ijms21114190

24. Gutschner T, Hämmerle M, Eißmann M, et al. The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res*. 2013;73(3):1180-1189. doi:10.1158/0008-5472.CAN-12-2850

25. Sun Z, Zhang B, Cui T. Long non-coding RNA XIST exerts oncogenic functions in pancreatic cancer via miR-34a-5p. *Oncol Rep*. Published online 2018:1-10. doi:10.3892/or.2018.6245

26. Jiang P, Wu X, Wang X, Huang W, Feng Q. NEAT1 upregulates EGCG-induced CTR1 to enhance cisplatin sensitivity in lung cancer cells. *Oncotarget*. 2016;7(28):43337-43351. doi:10.18632/oncotarget.9712

27. PD-L1. In: *Wikipedia*. ; 2023. Accessed February 15, 2023. https://en.wikipedia.org/w/index.php?title=PD-L1&oldid=1139304584

28. Sadic M, Schneider WM, Katsara O, et al. DDX60 selectively reduces translation off viral type II internal ribosome entry sites. *EMBO Rep*. 2022;23(12):e55218. doi:10.15252/embr.202255218

29. Mathieu J, Zhang Z, Zhou W, et al. HIF induces human embryonic stem cell markers in cancer cells. *Cancer Res*. 2011;71(13):4640-4652. doi:10.1158/0008-5472.CAN-10-3320

30. Saedeleer CJ de, Copetti T, Porporato PE, Verrax J, Feron O, Sonveaux P. Lactate Activates HIF-1 in Oxidative but Not in Warburg-Phenotype Human Tumor Cells. *PLoS ONE*. 2012;7(10). doi:10.1371/journal.pone.0046571

31. Faubert B, Li KY, Cai L, et al. Lactate Metabolism in Human Lung Tumors. *Cell*. 2017;171(2):358-371.e9. doi:10.1016/j.cell.2017.09.019

32. Petruzzelli R, Christensen DR, Parry KL, Sanchez-Elsner T, Houghton FD. HIF-2α regulates NANOG expression in human embryonic stem cells following hypoxia and reoxygenation through the interaction with an oct-sox cis regulatory element. *PLoS ONE*. 2014;9(10):1-11. doi:10.1371/journal.pone.0108309

33. Ahrendt SA, Decker PA, Alawi EA, et al. Cigarette smoking is strongly associated with mutation of the K-ras gene in patients with primary adenocarcinoma of the lung. *Cancer*. 2001;92(6):1525-1530. doi:10.1002/1097-0142(20010915)92:6<1525::AID-CNCR1478>3.0.CO;2-H

34. Masson N, Ratcliffe PJ. Hypoxia signaling pathways in cancer metabolism: The importance of co-selecting interconnected physiological pathways. *Cancer Metab*. 2014;2(1):1-17. doi:10.1186/2049-3002-2-3

35. Wang Y, Lei F, Rong W, Zeng Q, Sun W. Positive feedback between oncogenic KRAS and HIF-1α confers drug resistance in colorectal cancer. *OncoTargets Ther*. 2015;8:1229-1237. doi:10.2147/OTT.S80017

36. Chiappinelli KB, Strissel PL, Desrichard A, et al. Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. *Cell*. 2015;162:974-986.

37. Reggiardo RE, Maroli SV, Halasz H, et al. Epigenomic reprogramming of repetitive noncoding RNAs and IFN-stimulated genes by mutant KRAS. *bioRxiv*. Published online 2020.

38. Kim DH, Marinov GK, Pepke S, et al. Single-cell transcriptome analysis reveals dynamic changes in lncRNA expression during reprogramming. *Cell Stem Cell*. 2015;16:88-101 10 1016 2014 11 005.

39. Schmitt AM, Chang HY. Long Noncoding RNAs in Cancer Pathways. *Cancer Cell*. 2016;29:452-463.

40. Rinn JL, Chang HY. Long Noncoding RNAs: Molecular Modalities to Organismal Functions. *Annu Rev Biochem*. 2020;89:283-308 10 1146-062917-012708.

41. Slack FJ, Chinnaiyan AM. The Role of Non-coding RNAs in Oncology. *Cell*. 2019;179:1033-1055. doi:10.1016/j.cell.2019.10.017.

42. Roulois D, Loo Yau H, Singhania R, et al. DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts. *Cell*. 2015;162:961-973.

43. Mehdipour P, Marhon SA, Ettayebi I, et al. Epigenetic therapy induces transcription of inverted SINEs and ADAR1 dependency. *Nature*. 2020;588:169-173 10 1038 41586-020-2844-1.

44. Ostrem JM, Peters U, Sos ML, Wells JA, Shokat KM. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature*. 2013;503:548-551 10 1038 12796.

45. Hagemann-Jensen M, Ziegenhain C, Chen P, et al. Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nat Biotechnol*. 2020;38(6):708-714. doi:10.1038/s41587-020-0497-0

46. He X, Ashbrook AW, Du Y, et al. RTP4 inhibits IFN-I response and enhances experimental cerebral malaria and neuropathology. *Proc Natl Acad Sci U A*. 2020;117:19465-19474.

47. Pereira C, Gimenez-Xavier P, Pros E, et al. Genomic Profiling of Patient-Derived Xenografts for Lung Cancer Identifies B2M Inactivation Impairing Immunorecognition. *Clin Cancer Res*. 2017;23:3203-3213 10 1158 1078-0432-16-1946.

48. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114-2120.

49. Brown J, Pirrung M, McCue LA. FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics*. 2017;10(1093/bioinformatics/btx373).

50. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017;14:417-419 10 1038 4197.

51. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res*. 2015;4:1521 10 12688 1000 7563 2.

52. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15, 550:10 1186 13059-014-0550-0558.

53. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34(17):i884-i890. doi:10.1093/bioinformatics/bty560

54. UMI-tools: Modelling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Accessed February 24, 2023. https://genome.cshlp.org/content/early/2017/01/18/gr.209601.116.abstract

55. Sen C, Freund D, Gomperts BN. Three-dimensional models of the lung: past, present and future: a mini review. *Biochem Soc Trans*. 2022;50:1045-1056. doi:10.1042/BST20190569.

56. Mugarza E, Maldegem F, Boumelha J, et al. Therapeutic KRAS(G12C) inhibition drives effective interferon-mediated antitumor immunity in immunogenic lung cancers. *Sci Adv*. 2022;8:8780 10 1126 8780.

57. Wang J, Ma P, Kim DH, Liu BF, Demirci U. Towards Microfluidic-Based Exosome Isolation and Detection for Tumor Therapy. *Nano Today*. 2021;37:10 1016 2020 101066.

58. Singh A, Settleman J. Oncogenic K-ras "addiction" and synthetic lethality. *Cell Cycle Georget Tex*. 2009;8(17):2676-2677. doi:10.4161/cc.8.17.9336

59. Awad MM, Liu S, Rybkin II, et al. Acquired Resistance to KRAS ^{G12C} Inhibition in Cancer. *N Engl J Med*. 2021;384(25):2382-2393. doi:10.1056/NEJMoa2105281