Université de Montréal

Dissecting the Effects of Tumor Microenvironment Factors on Cancer Cells to Reveal Novel Targets for Multi-Targeting RNA-Based Therapeutics.

By

Jordan Quenneville

Programme de Biologie Moléculaire, option Biologie des Systèmes

Faculté de Médecine

Thèse présentée en vue de l'obtention du grade de Philosophiae Doctor (Ph. D.)

en Biologie Moléculaire, option Biologie des Systèmes

Août 2022

Université de Montréal

Unité académique : Institut de Recherche en Immunologie et Cancer, Faculté de Médecine

This Thesis is titled

Dissecting the Effects of Tumor Microenvironment Factors on Cancer Cells to Reveal Novel Targets for Multi-Targeting RNA-Based Therapeutics.

Presented by

Jordan Quenneville

Has been evaluated by a jury composed of the following people

Dr. Jean-François Côté
Chairman

Dr. Étienne Gagnon
Director of research

Dr. François Major
Codirector

Dr. Martin Smith
Member of the Jury

Dr. Marc Fabian
External Examinor

Résumé

Il devient de plus en plus clair que pour traiter efficacement les tumeurs solides, nous devons également nous intéresser au microenvironnement tumoral. Physiologiquement, les zones intratumorales peuvent présenter une disponibilité anormale en nutriments, un pH altéré ou encore des niveaux d'oxygène bas (hypoxie). Il est connu que l'adaptation hypoxique engendre des cellules tumorales qui sont plus difficiles à traiter indépendamment de l'approche thérapeutique. De plus, l'adaptation hypoxique est nécessaire pour la progression tumorale puisque cette dernière favorise des processus tels que: la survie cellulaire, la motilité, l'angiogenèse, le métabolisme du glucose, l'immunomodulation ainsi que la résistance aux médicaments. Ces phénotypes passent par la régulation des ARN messager (ARNm) et des micro ARN (miARN). Pour ces raisons, des efforts importants ont été déployés pour comprendre l'adaptation hypoxique et les interventions thérapeutiques potentielles pouvant la contrer. À l'heure actuelle, il y a un manque de cohérence et de variété dans les protocoles de traitement hypoxique *in vitro* qui ne tient pas compte des aspects importants de l'hypoxie *in vivo*, comme la réduction de la disponibilité en éléments nutritifs, la durée de l'exposition hypoxique ainsi que le degré d'hypoxie.

Pour mieux simuler le microenvironnement hypoxique *in vivo*, nous avons développé de nouveaux protocoles hypoxiques *in vitro* qui visent à simuler ces aspects. Tout d'abord, en utilisant une lignée cellulaire B16-HIF1a-eGFP, nous avons optimisé le stress métabolique à court terme en conjonction avec l'hypoxie pour augmenter la stabilisation de l'HIF1a. Pour déterminer comment le programme HIF1 adapte les cellules à ces différentes conditions, nous avons analysé les données de séquencage d'ARN qui démontrent que le stress métabolique induit un programme transcriptionnel HIF1 plus robuste et diversifié dans les cellules hypoxiques, et que ce dernier est représentatif du stress hypoxique *in vivo*. Nous avons également identifié de nouveaux miARN induits par l'hypoxie et démontré que notre protocole d'incubation régule davantage les miRNA associés au pronostic négatif du patient.

Nous avons aussi étudié l'adaptation hypoxique à long terme et extrême *in vitro*. Nous avons observé que l'incubation hypoxique à long terme induit une transition épithéliale à mésenchymateuse (TME), indépendante de l'expression différentielle des facteurs de transcription du TME canonique. Ce changement se produit à des niveaux spécifiques d'oxygène, et nécessite une pré-incubation à des niveaux hypoxiques plus faible. Avec ce protocole, nous avons découvert une nouvelle isoforme de WT1 (tWT1), un moteur potentiel du TME. tWT1 commence la transcription dans l'intron 5 du gène WT1, une région avec plusieurs séquences d'ADN contenant des éléments de réponse à l'hypoxie. La protéine tWT1 a une fonctionnalité limitée : elle est localisée au niveau du noyau, et conserve la liaison de l'ADN aux régions précédemment connues. Nous avons aussi identifié l'expression de tWT1 dans les échantillons de patients atteints de leucémie ainsi qu'une isoforme tWT1 potentiellement plus fonctionnelle grâce à des analyses par kmer.

Pour cibler ces phénotypes identifiés dans nos expériences d'adaptation hypoxiques, nous avons développé une nouvelle catégorie d'ARN intérférent (ARNi) thérapeutique : le microARN synthétique (synmiR). Les synmiR sont des molécules de RNAi avec des multiples cibles. En utilisant des expériences *in vivo*, nous avons établi de nouveaux principes de RNAi qui élargissent considérablement l'espace de conception pour les synmiR. Nous avons mis au point deux algorithmes de conception de synmiR distincts et avons testé leur efficacité dans le contrôle de l'activité transcriptionnelle du génome du VIH *in vivo*.

En conclusion, nous avons montré que l'inclusion de facteurs physiologiques supplémentaires associés à l'hypoxie *in vitro* entraîne un engagement plus robuste de l'adaptation de l'hypoxie. À ce jour, aucun de nos protocoles d'hypoxie n'a été reproduit dans la littérature. Nous contribuons aux connaissances dans le domaine en décrivant les nouveaux ARNm/miARN induits par l'hypoxie, ainsi que la méthode d'induction fiable de l'EMT par l'hypoxie seulement. Nous faisons également état de l'existence de nouveaux isoformes de WT1 et de leurs liens avec le cancer et l'hypoxie. La connaissance de ces isoformes est importante pour l'avenir de la recherche sur WT1, car elle pourrait faire la lumière sur des résultats auparavant inexplicables. Notre travail dans les synmiR ouvre une nouvelle voie d'investigation pour le traitement de certaines maladies, et fournit un mécanisme d'action testable pour les

miRNA endogènes. Une fois suffisamment développés, les synmiR offrent une occasion thérapeutique unique d'exploiter leur multi-ciblage pour avoir un impact spectaculaire sur une seule voie, ou affecter plusieurs voies par le ciblage simultané de gènes clés.

Mots-clés : cancer, hypoxie, HIF1a, RNAseq, microRNA, TME, WT1, ChIP, bioinformatique, RNAi.

Abstract

It is becoming increasingly clear that in order to effectively treat solid tumours, we must also address the tumour microenvironment. Physiologically, intratumoral areas may have abnormal nutrient availability, pH, or lower oxygen levels (hypoxia). It is known that hypoxic adaptation results in tumour cells which are harder to treat regardless of therapeutic approach, and hypoxic adaptation is necessary for disease progression due to the induction of tumour promoting phenotypes such as, but not limited to: cell survival, motility, angiogenesis, glucose metabolism, immunomodulation, and drug resistance. This is accomplished through the regulation of both mRNAs and miRNAs. For these reasons, significant effort has been applied to understanding hypoxic adaptation and potential therapeutic interventions. Currently, there is a lack of consistency and protocol variety in *in vitro* hypoxic treatments that leaves out important aspects of *in vivo* hypoxia, such as reduced nutrient availability, length of hypoxic exposure, and degree of hypoxia.

To better simulate the *in vivo* hypoxic microenvironment, we have developed new *in vitro* hypoxic protocols which aim to simulate these aspects. First, using a B16-HIF1α-eGFP hypoxia reporter cell line, we optimized short-term metabolic stress in conjunction with hypoxia to enhance HIF1α stabilization. To ascertain how the HIF1 program adapts the cells to these different conditions, deep transcriptome profiling were performed and demonstrated metabolic stress induces a more robust and diversified HIF1 transcriptional program in cells under hypoxia, which was more representative of *in vivo* hypoxic stress. We identified novel hypoxia-induced miRNAs as well, and demonstrated our incubation protocol regulated more miRNAs associated with negative patient prognosis.

We also investigated long-term and extreme hypoxic adaptation *in vitro*. Long term hypoxic incubation induced a epithelial to mesenchymal transition (EMT), independent of canonical EMT factor differential expression. This switch occurred at specific oxygen levels, and required pre-incubation at milder hypoxic levels, highlighting the relevance of simulating *in vivo* hypoxia development *in vitro*. Through this protocol, we discovered a novel isoform of WT1 (tWT1), a potential driver of our EMT. tWT1 begins transcription within intron 5 of the WT1

gene, a region with several Hypoxia Response Elements DNA sequences. tWT1 retains limited functionality: it is able to localize to the nucleus, and retains DNA binding to previously known gene promoter regions. We also identified the expression of tWT1 in leukemic patient samples as well as a potentially more functional tWT1 isoform through kmer-based analyses.

To target these multiple phenotypes identified in our hypoxia adaptation experiments, we worked towards developing a new category of RNA-interference (RNAi) therapeutic, the synthetic microRNA (synmiR). SynmiRs are single-sequence, multi-targeted RNAi molecules. Using *in vivo* knock-down experiments, we established new RNAi principles which dramatically expand the design space for synmiRs. We developed two philosophically distinct synmiR design algorithms, and validated their efficacy in controlling HIV genome transcriptional activity *in vivo*.

In conclusion, we have shown the inclusion of additional physiological factors associated with hypoxia *in vitro* results in a more robust engagement of hypoxia adaptation. To date, neither of our hypoxia protocols have been replicated in the literature. We contribute to the literature by describing novel hypoxia induced mRNAs/miRNAs, as well as methods for reliably inducing EMT through hypoxia alone. We also discovered the existence of novel WT1 isoforms and their links to cancer and hypoxia. Knowledge of these isoforms is important for WT1 research moving forward, as it may shed light on previously unexplainable results. Our work in synmiRs opens a new therapeutic avenue for multiple disease states, and provides a testable mechanism of action for endogenous miRNAs. Once sufficiently developed, synmiRs offer a unique therapeutic opportunity to harness their multi-targeting to dramatically impact a single pathway, or affect multiple pathways through simultaneous targeting of key genes.

Keywords : cancer, hypoxia, HIF1a, RNAseq, microRNA, EMT, WT1, ChIP, bioinformatics, RNAi.

Table of Contents

List of Tables

List of Figures

List of Symbols and Abbreviations

A

AML: Acute Myeloid Leukemia

B:

BCAA: Branched Chain Amino Acids

B16-HG: B16 hypoxia adaptation reporter cell line derived from single cell clone transduced with pHAGE-HIF1α-eGFP

C:

CDS: protein Coding Sequence (protein CoDing Sequence)

ChIP/seq: Chromatin Immunoprecipitation / Chromatin Immunoprecipitation Next-Generation Sequencing

D:

DAVID: Database for Annotation, Visualization and Integrated Discovery

DEG/s: Differentially Expressed Gene(s)

deGFP: destabilized enhanced Green Fluorescence Protein

DEmiR/s: Differentially Expressed microRNA(s)

dH_2O: distilled water

DMEM: Dulbecco's Modified Eagle Medium

dNTP: Deoxynucleoside triphosphate. Equimolar mixture of ATP, TTP, CTP, GTP

DTT: 1,4-Dithiothreitol

E:

Exon-6: Truncated WT1 isoform which contains exon 6 of the canonical murine WT1 mRNA

EDTA: Ethylenediaminetetraacetic acid

eGFP: enhanced Green Fluorescence Protein

F:

FA: Fatty Acid

FACS: Fluorescence Activated Cell Sort

FBS: Fetal Bovine Serum

FC: Fold Change

G:

GO: Gene Ontology

GSEA: Gene Set Enrichment Analysis

H:

HEPES: N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid

HIF: Hypoxia Inducible Factor

HRE: Hypoxia Responsive Element

HY: *in vitro* metabolic hypoxia experimental condition, standard hypoxic incubation

I:

ISO: Isothermal buffer

IRIC: Institute for Research in Immunology and Cancer

IRES: Internal Ribosomal Entry Site

J

K:

KD: Knock-down

KTS: WT1 splicing event resulting in insertion of lysine threonine, serine tripeptide between WT1 zinc fingers 3 and 4.

L:

lncRNA: Long Non-Coding RNA

LTHY: Long-Term Hypoxia timecourse

M:

mCherry: monomeric Cherry

MCS: Multiple Cloning Site

mg: milligram

MHY: Metabolic Hypoxia adaptation dataset

miR/miRNA: microRNA

miRNAseq: Next-Generation RNA sequencing of small cellular RNAs. Sequences microRNAs

mL: milliliter

mM: millimolar

mRNA: messenger RNA

mRNAseq: Next-Generation Sequencing of cellular RNA for the quantification longer cellular RNA species (messenger and long non-coding RNAs).

N:

NAD+: Nicotinamide adenine dinucleotide

NADH: nicotinamide adenine dinucleotide (NAD) + hydrogen

NADP+: Nicotinamide Adenine Dinucleotide Phosphate

NADPH: Nicotinamide Adenine Dinucleotide Phosphate + hydrogen

ng: nanogram

nt(s): nucleotide(s)

O:

ODD: Oxygen-Dependant-Degradation

P:

P/S: Penicillin and Streptomycin

P2A: self-cleaving peptide P2A

padj: Benjamini-Hochberg adjusted pvalue

PEG-8000: Polyethylene Glycol 8000

PES: Polyethersulfone

PFA: Paraformaldehyde

PNK: Polynucleotide Kinase

Pol(II/III): Polymerase II / Polymerase III

Poly(A)(U)(T): consecutive sequence of Adenines, Uricils, or Thymines

PTM(s): Post Translational Modification(s)

Puro: Puromycin

Q:

qPCR: quantitative PCR

R:

RHY: *in vitro* metabolic hypoxia experimental condition, metabolically restricted hypoxic incubation

RNAi: Ribonucleic acid interference

RNAse(s): ribonuclease(s)

RNAseq: Next-Generation RNA sequencing

RPM: Revolutions Per Minute

RMPI: Roswell Park Memorial Institute[1]

rRNA: ribosomal RNA

S:

saRNA(s): small artificial RNA(s)

SK: Isoform of truncated WT1 which splices out canonical *WT1* exon 6.

synmiR: synthetic microRNA

T:

TC: *in vitro* metabolic hypoxia experimental condition, standard tissue culture condition

TCA: Tricarboxylic Acid

TF: Transcription Factor

TSS: Transcription Start Site

tWT1: truncated-WT1

U:

ug: microgram

uL: microliter

uM: micromolar

V,W,X,Y,Z

Give me a lever long enough and a fulcrum on which to place it, and I shall move the world.
-Archimedes

Thanks

First and foremost, I would like to thank my supervisor Dr. Étienne Gagnon for letting me join his lab back in 2014 as a summer student. Throughout my 8 years under his mentorship, he's done nothing but support me both professionally and personally. I'd like to give him my deepest thanks for forcing me to see the value of my work, challenging me to think critically, and being an exemplary scientist who I strive to model. After doing my summer rotation, I chose to do my Master's with you because I thought you were a good mentor. After my Master's I knew you were a good mentor, and that opinion hasn't changed now. I'm grateful for the time I've spent in your lab.

I would also like to thank my co-director, Dr. François Major, for always being open for discussions, challenging my preconceptions, and giving me the opportunity to teach. I want to thank him as well for and the multi-hour, mind-expanding, head-ache inducing discussions we'd have after hours. They always pushed my thinking to the limits, and I always enjoyed them.

I would like to thank all of my committee members as well for helping to guide my research over the years, pushing me to develop my professional presentation skills. I would like to thank Dr. Brian Wilhelm (and his optimist's tears mug), and Dr. Jerry Pelletier in particular for challenging me and hammering out any chinks in my ideas during my committee meetings. A special thanks goes to my high school Biology teacher, Ms. Jennifer Hunter, who first exposed me to the world of molecular biology and fed my passion for it. I'd like to thank Julie and Pascale from IRIC academic affairs as well, for helping me navigate something infinitely more complex than cancer biology: UdeM administration.

I want to thank all the past and present members of the Gagnon lab. I couldn't have asked for better lab-mates, and you all made my time in the office fun on the good days, and bearable on the bad. Audrey, thanks for giving me the nickname "hackerman" after I showed you find&replace on excel. Rebecca, thanks for being the glue that kept us all together, looking out for me when I needed it. Margaux, thanks for being a great friend throughout our time here. Having you around during the COVID recovery made the lab a lot more friendly. Your hard work and determination inspires me. I'm proud to be a member of team "laipns". I'd like to thank

Angélique for keeping the lab running smoothly, and making sure I never ran out of olives. Special thanks to Albert for being the best rubber duck a guy could ask for, as well as a good friend. Thanks to Roger, Yahya, and Emma as well for their continual friendship and support over the years.

I don't think I would be where I am now if not for my incredible family. Mom, thank you always supporting me, and nurturing my scientific interest since infancy. Thank you for always believing in me, wanting the best for me, and helping me. And thank you for letting me return the favour. Dad, thank you for being a damn good guiding light, keeping me on the straight and narrow, and supporting my scientific interest. It meant a lot knowing you have my back. Oma, thank you. You're always helping. Les, thank you for being the other academic in the room, and your continual love and support. Ilona, thank you for not snapping me like a chicken, and your running jokes. To Alex and Amelie, thank you for reminding of life beyond the bench.

While a lot of my friendships have slipped over the years, there are a few that are bonded forever. To Khal and Mike, thank you both for being who you are. For always wanting the best for me, and keeping me grounded, and the way we all looked up at each other. For appreciating me being the dumbest guy in the room. You guys gave me some of the biggest laughs of my life, and helped me recharge during the darker days. Mike, thanks for always being the life of the party, and pulling me into a good mood. Rock and stone. Khal, thanks for teaching me about cars, I love to learn and your enthusiasm is infectious. Sand the fridge.

For Jackie, the love of my life. Thank you for being the anchor in the storms. For dragging me out of the lab to live, and keeping my nose to the grindstone when needed. For always being willing and able to listen, and courageous in speaking your mind. Everything I have built in no small part, was thanks to you. For pulling me to be better, for you. With you at my side, what's the worst that can happen?

That might be a lot of people to thank, but it's hard to pick when you're surrounded by good people. But enough of that, on to bigger ponds. Thank you all.

Forward

As a tumour grows the local microenvironment changes and becomes more heterogeneous. Inevitably, the amount of cancer cells exceeds the capacity that the local circulatory system can support. This results in a build up of metabolic waste, an increase in acidity, a drop in nutrient availability, and a drop in local oxygen levels known as hypoxia. To adapt to this physiological stress, cancer cells need to engage in a transcriptional response known as the hypoxia response. This transcriptional response activates adaptations leading to an increase in therapeutic resistance, enhanced and dysregulated angiogenesis, increased metastatic probability, and other tumour promoting phenotypes. This transcriptional response is so beneficial to the tumour that many cancers constitutively express it, regardless of local oxygen status. As a result, hypoxic adaptation has been of intense interest in tumour biology.

In this doctoral thesis, the adaptive capacity of tumour cells in response to hypoxia is further explored through the development and investigation of novel *in vitro* hypoxia incubation protocols. Their purpose was to identify therapeutic targets for a novel RNAi technology which we developed. This thesis is written in the classical style and is broken into six chapters. The first chapter is an introduction to the research done which acts as the foundation for this project: spanning hypoxia research and HIF1-controlled phenotypes, the role of EMT and WT1 in tumour biology, and the role microRNAs in disease and therapeutics. Chapter two contains the materials and methods used to generate the results presented in chapters three to five, which are detailed in the following paragraph. Chapter six contains a larger discussion of the results presented chapters three to five, how the work presented could be progressed, and how they fit within the literature as a whole.

The results are divided into three chapters as follows. Chapter 3 focuses on the generation of a hypoxia-adaptation reporter cell line, its use towards developing a novel *in vitro* hypoxic incubation protocol which focuses on the addition of metabolic stress which hypoxic adaptation, and the description of these adaptations through RNAseq analyses. Chapter 4 focuses on how long-term and deep hypoxic conditions affect cancer cells. This was accomplished through the generation of novel long-term hypoxic incubation protocol, the

description of the transcriptomic changes it induces, and the characterization of a novel WT1 isoform induced specifically by this protocol. Chapter 5 focuses on the generation of a novel kind of RNAi molecule which is designed to be multi-targeting, through the use of *in vivo* experiments and novel design algorithms designed to generate a synthetic microRNA. Results are briefly discussed in each chapter as they arise, and a cohesive conclusion for the chapter is given at the end of each chapter. Chapter six focuses on discussions surrounding potential future experiments for each chapter, how the work presented fits into the literature as a whole, and how this work progresses the scientific knowledge of humanity.

Chapter 1 – Introduction

1.1 Chapter Summary

In this chapter, I will provide the necessary background information needed to understand the framing of this project. The first section of this introduction focuses on hypoxia and its role in the tumour microenvironment. It begins with a brief history of the hypoxia research field, starting with the initial observation made in the 1930s to the description of the molecular pathways governing the cellular response to hypoxia. It then moves onto the common dysregulations seen in cancer leading to the over activation of the hypoxia response, and how the hypoxia response induces so many tumour promoting phenotypes. To cap off this section, there is a brief introduction to the common practices used in the literature to study the hypoxia response *in vitro*.

Following the hypoxia section of the introduction, there is an introduction to the epithelial to mesenchymal transition (EMT). Beginning with a discussion of the different types of EMT found in biology, this section focuses more on the role of EMT in tumour biology and pathology through the upregulation of classical EMT driving transcription factors. Particular focus is given to WT1, a non-canonical EMT driving transcription factor, as it is a major gene of interest in chapter 4.

The final section of the introduction focuses on the role of microRNAs in disease and their potential as therapeutics. An introduction to the biogenesis of microRNAs, their role in homeostasis and disease is provided. The current understanding of the microRNA mechanism of action is discussed as well, as this is of particular importance for chapter 5. The section concludes with a discussion of the potential for RNA therapeutics in treating disease. I conclude this chapter with the hypothesis and objectives of this thesis.

1.2 The influence of hypoxia in cancer progression

1.2.1 The physiological role of oxygen

At a molecular level, oxygen is essential for the homeostasis of our cells. Molecular oxygen (O_2) is required to maintain cellular energy levels, where the vast majority is used to generate cellular ATP through the electron transport chain in the mitochondria[2,3]. Beyond metabolism, O_2 is used for a host of other functions. Oxygen is required for hydroxylation of proline, asparagine, lysine, aspartate, histidine, and other amino acids[4–8]. Other biological reactions require O_2 as an essential component, such as for the generation of nitric oxide in immune cells to fight pathogens[9]. Whenever cells are exposed to oxygen levels lower than ideal, they are said to be experiencing hypoxia[10]. Hypoxia can occur naturally, such as during embryogenesis, where hypoxia occurs naturally as tissues develop and the local vasculature can no longer adequately supply oxygen to the more distal cells[11]. In this scenario, the hypoxic cells adapt by altering their metabolic profile to survive, and signal for increased local vascularization to eventually alleviate the hypoxia. Different tissues have different oxygen demands, and so the threshold for cellular hypoxic signalling is somewhat tissue specific. As such, normal tissue oxygen levels (physoxia) can range from 2-9% O_2, and averages at 5% O_2[12]. Physoxia for certain tissues would be considered hypoxic in others. For example some tissues have particularly low physoxia such as: the skin (1-5.3% O_2)[13–15], kidney medulla (2% O_2)[15–17], thymus (>1% O_2)[15,18,19], and bone marrow (2.7% O_2)[20]. These represent large drops in oxygen content relative to the oxygen content of arterial blood at 15% O_2 [12]. In pathological settings hypoxia can occur during ischaemia, inflammation, wound healing, and cancer; given that these disease states result in reduced blood flow or increased oxygen demand to the local tissue[10,12].

1.2.2 History of hypoxia research in cancer

The first indication of a link between oxygen and cancer was made by Dr. Otto Warburg in the 1920s. In his work, Warburg observed that tumour cells had a predisposition for metabolizing glucose by the oxygen-independent aerobic glycolysis pathway rather than the more energy generative electron transport chain within the mitochondria[21]. This phenomena, termed the Warburg effect, has since been observed in both solid and liquid cancers and

remains as one of the hallmarks of cancer to this day[22]. Shortly after this observation, Gray and colleagues discovered that a complete lack of oxygen in tissue (anoxia) had a protective effect against X-ray radiation therapy, suggesting that tumour cells may have an increased survival chance when in a sufficiently hypoxic environment [23–25]. Additional evidence of this phenomena was provided by Powers & Tolmach as well as Churchill-Davidson, which demonstrated the increased efficacy of radiation therapy after hyperoxgenation of the blood using hyperbaric air in both mice and cancer patients[26,27]. Together, these data suggested a radioprotective effect of hypoxia and anoxia on tumours. However the mechanism of action for this effect remained a mystery.

Before direct detection of the hypoxic environments could be achieved, its existence was inferred using immunohistochemical studies of lung cancer tumours[25]. In that study, Gray and colleagues observed large necrotic sections of the tumour with islands of viable cells surrounding capillaries. These islands extended 85 micrometers from the capillaries, which correlated well with the expected oxygen diffusion distance of 72 micrometers[25]. This suggested that as the tumour grows, it outpaces the oxygen and nutrient supply capacity of the local vasculature (**Figure 1**)[25]. This feature of tumour growth was verified in other solid tumour types, and oxygen measurements of tumours found them to be more hypoxic than their surrounding normal tissue [28–35]. Consumption of nutrients, growth factors and oxygen by the tumour cells outpaces the capacity of the vasculature, which in extreme cases, leads to necrosis of sections of the tumour beyond the oxygen diffusion distance[36]. Therefore, as cancers grow within the body, the growth of the tumour inevitably complexifies the physiological environment in which they inhabit, where one of the major physiological differences is the lack of oxygen.

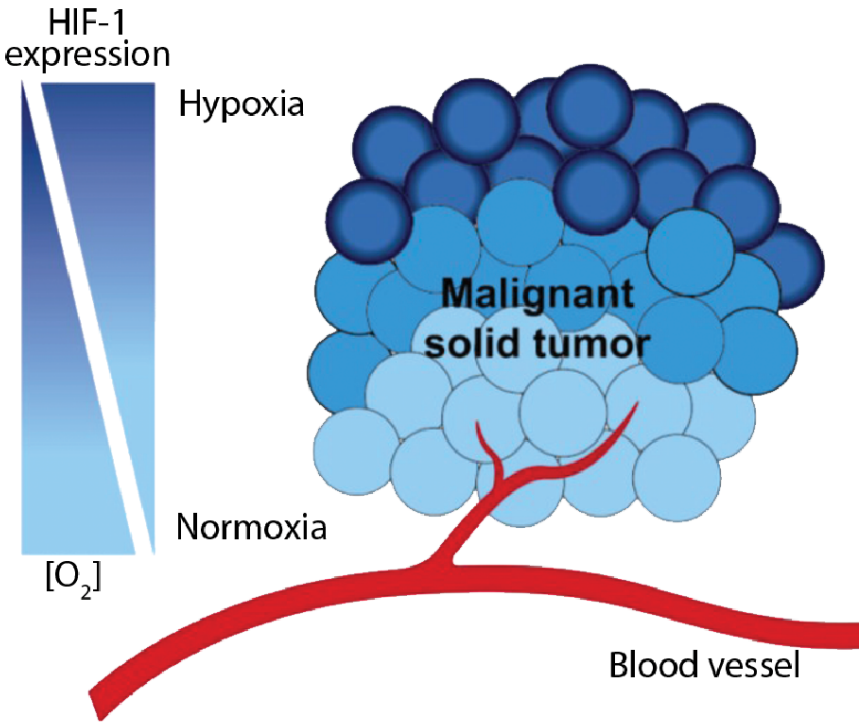

Figure 1: The tumour microenvironment and hypoxia

As tumors grow, O_2 concentrations ([O_2]) drop as the tumor grows beyond the O_2 diffusion limit. As O_2 levels drop, HIF1 protein levels rise. Adapted from REF[758]

1.2.3 Discovery and regulation of the Hypoxia Inducible Factor (HIF) proteins

1.2.3.1 Discovery of HIF1α

As hypoxia was being discovered and described at the level of tissues, research was also being done to determine the effects of hypoxia at the cellular level. Studies found hypoxia capable of inducing several different cellular phenotypic changes, including angiogenesis signalling, increased metastatic potential, DNA replication dysregulation, stress protein synthesis, and changes in glucose metabolism[37–41]. The first transcription factor found to be induced by hypoxia was discovered by Semenza and Wang in 1992, and was termed Hypoxia Inducible Factor 1 (HIF1). They found that HIF1 bound to a specific DNA sequence, termed the Hypoxia Regulatory Element (HRE: 5'-TACGTGCT-3'), which was originally identified as an enhancer of the *EPO* gene, one of the genes identified to be upregulated under hypoxic or

anoxic conditions[42–47]. Through subsequent studies, HIF1 was discovered to be a heterodimer, composed on the HIF1α and HIF1β, also known as ARNT [48]. A few years prior to this discovery, HIF1β was identified as a component of the Ah (dioxin) receptor, which binds to xenobiotic (ie: environmental pollutants such as aromatic hydrocarbons and dioxin) responsive genes, and is constitutively expressed [49,50]. HIF1α was a novel protein at the time, and through studies in HeLa cells was shown to be regulated by hypoxia at the level of protein, with a maximal induction level at 1% O_2 in HeLa cells[51].

1.2.3.2 Discovery of HIF1α paralogs

After the discovery of the HIF1α protein, two other proteins were identified as potential oxygen-dependant HIF1α-like proteins based on their sequence similarity and oxygen-dependant protein stability (**Figure 2**)[52]. Each of these proteins are synthesized from their own genes, and were later identified as paralogs[53]. While the exact origin of the primordial *HIFα* gene is still unknown, the evolutionary history of the three HIFα paralogs found in vertebrates has been investigated. Based on phylogenetic studies performed by Kalle T. Rytkönen and colleagues, the three *HIFα* genes found in vertebrates arose from two successive gene duplication events, with the first duplication eventually giving rise to *HIF3α*, and the second giving rise to *HIF1α* and *HIF2α*[54]. By the time that cartilaginous fish diverged from from tetrapods and teleosts (approximately 420 million years ago[55,56]), both of these gene duplication events had occured[53]. Based on mutation rates between the paralogs, different selection pressures were applied to each paralog, suggesting differing functionality[53].

The HIF1α paralogs are termed HIF2α, also known as endothelial PAS domain protein 1 (EPAS1); and HIF3α, also known as inhibitory PAS domain protein (IPAS)[11,57]. HIF2α was found to have a similar function to HIF1α, it heterodimerizes with HIF1β and controls the expression of HRE-regulated genes[58]. However HIF2α expression was limited to specific cells of tissues, in contrast to the ubiquitous expression of HIF1α[58]. While both HIF1α and HIF2α recognize the same HRE sequence, there are transcriptional targets which are common or unique to each of the HIF proteins[59,60]. In contrast to HIF1α and HIF2α, HIF3α functions as a dominant negative variant of HIF1/2α, sequestering HIF1β from the other HIFα subunits and prevents HIF1/HIF2 transcription. This dominant negative function is the result of a lack of C terminal transactivation

domain, which prevents the recruitment of transcription cofactors necessary for transcription[61]. Its expression appears to be limited to specific tissues, such as the thymus and eye[57,61].

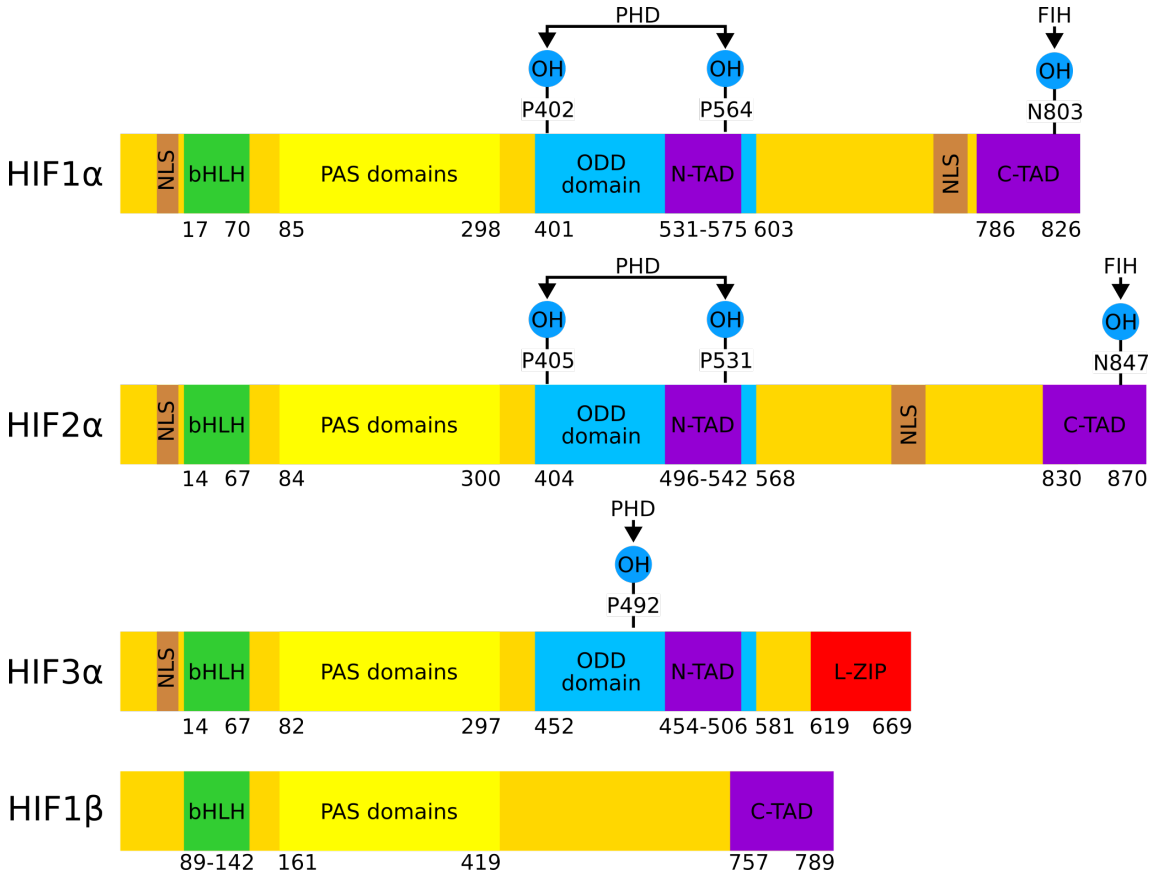

Figure 2: Domain architecture of HIF proteins

Protein domain comparisons between hypoxia inducible factors. NLS: nuclear localization signal. bHLH: Basic helix loop helix domain. Used for DNA binding. PAS: Per-Arnt-Sim (Per: period circadian protein. Arnt: Ah receptor nuclear translocator protein. Sim: single-minded protein) domain[82,758–760]. Used for dimerization of alpha and beta subunits. ODD: Oxygen-Dependant-Degradation. N-TAD: N-terminal transactivation domain. C-TAD: C-terminal transactivation domain. L-ZIP Leucine Zipper domain. PHD: Prolyl Hydroxylases. FIH: Factor Inhibiting HIF1. OH: hydroxylation post-translational modification. PHD: Prolyl hydroxylases.

1.2.3.3 Discovery of HIF oxygen-dependant stabilization mechanism

Following the discovery of the oxygen dependency of the HIF1α protein, work began to determine how this protein's expression was regulated by oxygen levels. Salceda and other groups discovered that HIF1α protein oxygen-dependant stability was regulated by a 200 amino acid long region termed the Oxygen-Dependant-Degradation (ODD) domain, and that HIF1α under normoxic conditions is ubiquitinated and degraded at the proteasome[62–64]. The discovery of the ubiquitin E3 ligase responsible for ubiquitinating HIF1α was made in 1999 by Maxwell, Ratcliffe, and colleagues using renal carcinoma cells[65]. These cells lacked the expression of the von Hippel–Lindau tumour suppressor protein (VHL), a substrate recognition component of the ubiquitin E3 ligase complex[65,66]. In these cells, HIF1α protein was constitutively present under normoxic conditions. Upon exogenous expression of VHL, hypoxic control of HIF1α protein levels was restored[65]. Further studies from the Ratcliffe group and others demonstrated that the HIF1α:VHL interaction required for ubiquitination is dependant on hydroxylation of specific proline residues within the ODD domain of HIF1α, specifically P564[67–69]. This hydroxylation was found to be performed by a small family of constitutively expressed prolyl hydroxylases (PHD1-3), with a fourth vertebrate specific member of the family found later (PHD4)[70–73]. Enzymatic analyses of these proteins found that Fe^{2+}, 2-oxoglutarate, ascorbate, alpha-ketoglutarate, and O_2 was required to induce HIF1α prolyl hydroxylation[74–78].

Similar to the PHDs controlling HIF1α protein stability, HIF1 transcriptional activity is also regulated through oxygen-dependant hydroxylation. The 2-oxoglutarate-dependent oxygenase Factor Inhibiting HIF1 (FIH1) hydroxylates an asparagine residue within the C-terminal transactivation domain of HIF1α (N803)[79,80]. This hydroxylation prevents HIF1 from interacting with transcriptional co-activators, thus preventing HIF1 transcriptional activity[5].

And thus, the canonical regulatory mechanism controlling hypoxia adaptation was established (**Figure 3**, left and middle pathways). HIF1α and HIF1β proteins are constitutively expressed. Under normoxic conditions, HIF1α is subjected to oxygen-dependant degradation via hydroxylation at specific proline residues within the ODD domain by prolyl hydroxylases. This post-translational modification (PTM) is recognized by the ubiquitin E3 ligase VHL, resulting in polyubiquitination of HIF1α, and subsequent proteasome-mediated degradation[63,65]. This

regulatory mechanism ensures that HIF1α protein only has a half-life of 5 minutes under normoxic conditions, and is transcriptionally inactive[5,62,79,81]. Under hypoxic conditions, the lack of oxygen prevents the hydroxylation of HIF1α. This allows HIF1α to escape proteasomal degradation, and heterodimerize with HIF1β. The functional HIF1 protein can then translocate into the nucleus to transcribe genes regulated by HRE elements[82]. Transcription is mediated through HIF1α or HIF2α mediated interactions with transcriptional co-activators such as the p300-CREBBP (CREB-binding protein) complex[83].

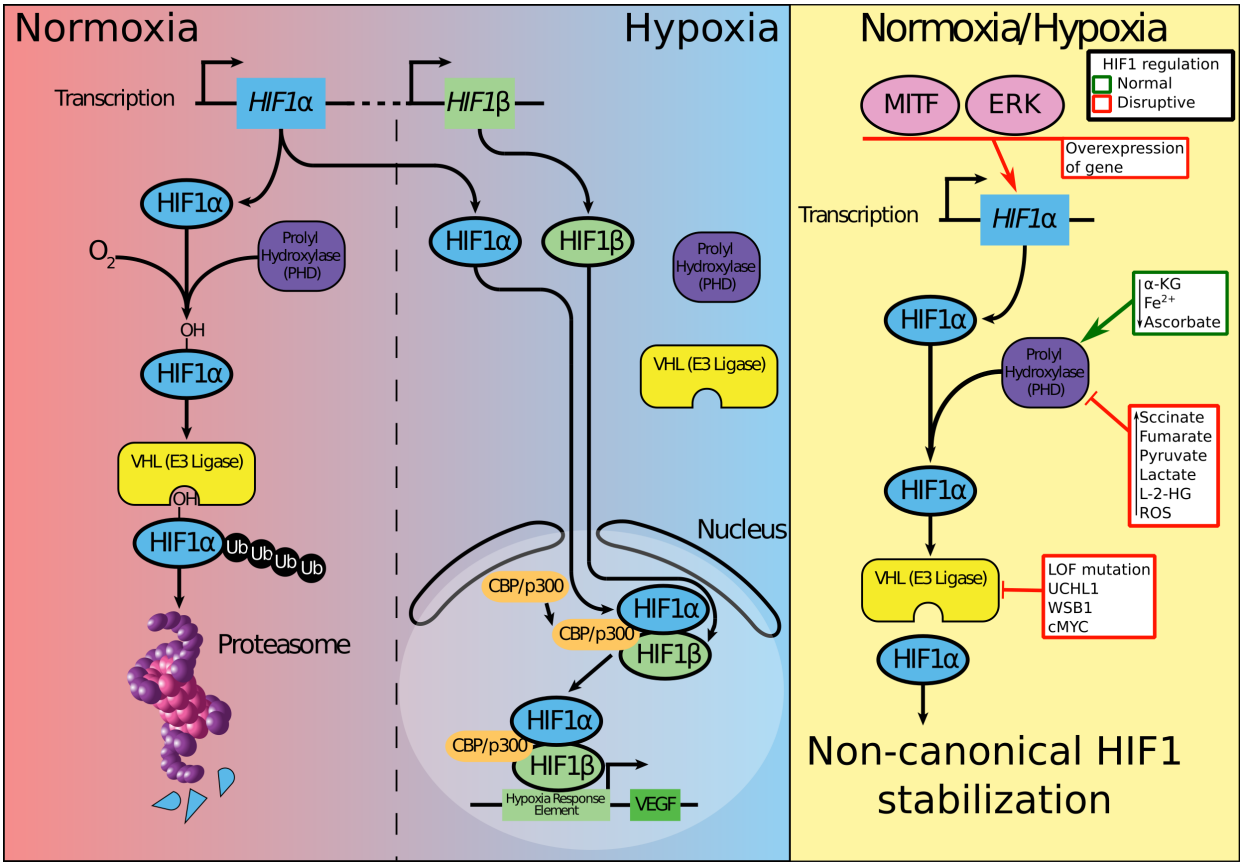

Figure 3: HIF1a canonical and non-canonical regulatory pathways

Overview of HIF1 regulatory mechanisms. ROS: Reactive Oxygen Species. LOF: Loss of function

In addition to hydroxylation, there are other post-translational modifications known to affect HIF1α. HIF1α can be phosphorylated, methylated, acetylated, and sumoylated, on several different residues[84]. Several modifications increase the sensitivity of HIF1α to normoxic oxygen dependant degradation[85–90]. Other modifications increase HIF1 stability or function[91–101]. And

other modifications have been shown to decrease HIF1 functionality, by either influencing HIF1 transcriptional activity or disrupting HIF1α:HIF1β binding[102–104]. In total, 41 different PTMs have been identified on HIF1α, with the conservation of these modifications found in over 200 vertebrate species[102]. As such, HIF1 activity can be finely tuned through post-translational modifications of HIF1α in a cell-environment specific context.

1.2.3.4 HIF1 transcriptional regulation

As mentioned previously, once stabilized HIF translocates to the nucleus to activate its transcriptional program by binding to HRE sequences in the genome[105,106]. Through the analysis of of 107 hypoxia-regulated gene promoters, the original *EPO*-associated HRE (TACGTGCT) could be condensed to a core motif of RCGTG, where "R" represents the presence of a purine nucleobase[107,108]. Should the HRE be preceded by a cytosine, the motif becomes an enhancer-box (E-box) motif, which can be recognized by several other transcription factors, typically with basic helix-loop-helix protein structure, such as MYC[109,110]. Given the presence of the CG dinucleotide in the core motif, it is possible that the HRE sequence could be methylated[111]. Recent work by D'Anna and colleagues have demonstrated that DNA methylation of this dinucleotide can directly negatively HIF binding to HREs[112]. Therefore, which HREs HIF proteins bind to is sensitive to the epigenetic context of the cell.

Chromatin immunoprecipitation sequencing (ChIPseq) analyses performed by Schödel and colleagues in 2011 revealed that while HIF1 and HIF2 do bind to the same HRE consensus sequence and target the same genes, they also bind to phenotypically distinct genes[113]. In their analysis, they demonstrated that the plurality of HIF1-bound HREs were within 2.5kb of the closest annotated Transcription Start Site (TSS)[113]. This was contrasted against the HIF2-bound HREs, which were found to be more distally distributed from known genes[113]. They found that a positive correlation between these HIF-bound genes and hypoxia-upregulation, but no enrichment among hypoxia-downregulated genes. This suggests that HIF1/2 binding to gene promoters results in activation of the gene, and that genes down-regulated by hypoxia are not the result of direct HIF-binding[113]. For HIF-bound genes which did not respond to hypoxia, Schödel and colleagues found a correlation between HIF-binding and expression of the second-nearest transcript, suggesting that HIFs can regulate gene expression across larger distances and

ignoring genomic elements[113]. To validate the long-distance HIF:gene interactions, Schödel and colleages confirmed two of the longest HIF1-bound HRE interaction in their dataset (to *ARRDC3* and *ERRFI1*), which were interactions spanning over 100kb[113]. In summary, Schödel and colleagues demonstrated that the HIF proteins can regulate genes over short or long genomic distances, through upregulation of gene transcription[113].

Later ChIPseq analyses performed by Choudhry and colleagues in 2015 expanded on the work done by Schödel by examining the regulation of non-protein coding RNAs by hypoxia[114]. Through their analyses, they demonstrated that the HIF transcriptional factors regulate all classes of RNAs, reinforcing the finding by Schödel that HIF-binding is associated with transcriptional activation[113,114]. They found that in addition to mRNAs as a class being upregulated by hypoxia, that ~22nt translation-regulating RNAs (termed microRNAs) and >200nt long non-protein RNAs (termed lncRNAs) were also upregulated under hypoxic conditions. In contrast, the other non-protein coding RNA classes (piwi-interacting RNAs, small-nucleolar RNAs, and transfer RNAs) were down-regulated after hypoxic incubation[114]. Choudhry also elucidated a mechanism fo HIF-regulated coding and non-coding gene activation: the release of pre-bound RNA polymerase II[114]. Much like the regulation of the HIFα subunits, their gene regulation mechanism of action is also primed to minimize activation time[62,114]. Among the lncRNAs, Choudhry and colleagues found that MALAT1 and NEAT1 to be the top upregulated lncRNAs after hypoxic incubation. Since their work, several other lncRNAs have been implicated in the transcriptional response to hypoxia. While a cohesive model of their function in hypoxia has not yet formed, recent work has shed light on potential mechanisms[115]. Several of these lncRNAs, including MALAT1 and NEAT1, function in part by sequestering endogenous microRNAs, allowing for increased protein production of oncogenic proteins[115–117]. Following the work of Choudhry, several additional lncRNAs have been associated with the hypoxia response in a multitude of cancers[115]. As more evidence of the role of lncRNAs play in tumorigenic hypoxia adaptation collects, it is now clear that the transcriptional response to hypoxia is is composed of a mixture of mRNA, microRNA, and lncRNA differential regulation.

1.2.4 Non-canonical HIF1 regulation

Since the establishment of the canonical HIF1α oxygen-dependant regulatory pathway, different mechanisms resulting in non-canonical HIF1α stabilization have been discovered at every level of the pathway (**Figure 3**, right pathway)[118]. At the level of *HIF1α* transcription, activation of the PI3K/AKT/mTORC1 pathway leads to increased HIF1α mRNA levels. Additionally, mTORC1 activation increases HIF1α translation rate[119]. The RAS/RAF/MEK/ERK signalling cascades have also been shown to increase HIF1α translation rate[118,120]. In a similar fashion, the microphthalmia-associated transcription factor (MITF) with the E318K mutation can increase *HIF1α* transcription[121]. V-Scr has been shown to increase HIF1α mRNA transcript levels as well[122]. As such, many of the oncogenic signalling pathways overlap with the HIF1α regulatory pathway, at the level of transcription or translation[122–133]. These mechanisms combined result in higher levels of HIF1α protein under normoxic conditions. If expressed at sufficiently high levels, HIF1α protein levels can overwhelm the PDH/VHL/FIH regulatory mechanisms and lead to functional HIF1 complexes under normoxic conditions[121].

HIF1α prolyl hydroxylation can also be inhibited in cancer cells, leading to normoxic stabilization of HIF1α protein. Reduction of the necessary PHD cofactors such as cytosolic Fe^{2+}, alpha-ketoglutarate, or ascorbate can lead to a reduction of hydroxylation of the ODD domain and therefore an accumulation of HIF1α protein[134–139]. In addition to a drop in cofactor levels, PHD activity can be impaired through competitive inhibition. Other divalent metal ions, such as Co^{2+} and Ni^{2+} can displace the Fe^{2+} ion from the PHD catalytic site, inhibiting their activity[140,141]. Displacement of 2-Oxoglutarate by succinate or fumarate can inhibit PHD activity when their cytosolic concentrations increase due to mitochondrial defects[71,142,143]. Elevated levels of pyruvate and lactate exhibit a similar effect on PHD activity[144–146]. Despite being a normal metabolic intermediate from the mitochondria, 2-hydroxyglutarate (2HG) can inhibit PHD function due to its structural similarity to alpha-ketoglutarate, and is often found in tumours at elevated levels as the result of mutant isocitrate dehydrogenase function[147–149]. Intracellular cysteine depletion results in self-inactivation of PHD proteins, leading to increased levels of HIF1α protein[150]. Therefore, PHD functionality can often be disrupted in cancer settings due to

either a lack of cofactors or presence of natural competitive inhibitors, leading to normoxic stabilization of HIF1α.

Beyond PHD activity, HIF1α protein degradation can be inhibited through a loss of functional VHL protein. Loss of *VHL* function can be a predisposition, as with people suffering from Von Hippel-Lindau Disease, who carry germline inactivating mutations in the *VHL* gene and a loss of function in the VHL protein. These people carry an elevated risk of developing hyper-vascularized tumours and most patients die of cancer, typically clear cell renal carcinoma[151]. Due to the lack of *VHL* function, VHL disease patient tumours often exhibit elevated markers of HIF1 transcriptional activity[151,152]. Even without considering VHL disease, the majority of clear cell renal carcinoma tumours contain *VHL* inactivating mutations[153–156]. Alternative mechanisms for VHL loss of function are: upregulation of ubiquitin carboxyl-terminal hydrolase L1 (UCHL1), a deubiquitinating enzyme shown to remove ubiquitin from HIF1α[157]; ubiquitin directed degradation of VHL by WD repeat and SOCS box-containing protein 1 (WSB1)[158]; and MYC regulated VHL degradation[159,160]. Therefore, there are several mechanisms which result in VHL loss of function in tumours, hinting its importance as a tumor suppressor gene.

In addition to mechanisms which circumvent oxygen-dependant HIF1α degradation, there have also been investigations into oxygen-independant HIF1α regulatory mechanisms. One such mechanism involves the protective binding of HSP90 to HIF1α to prevent RACK1 binding and recruitment of ubiquitin E3 ligases[161,162]. A second oxygen-independant pathway involves binding to p53, recruiting the mouse double minute 2 homolog (MDM2) E3 ubiquitin ligase, resulting in ubiquitin-dependant proteasomal degradation[163]. Thus, loss of p53 function can contribute to increase HIF1α protein levels in some cancer settings.

1.2.5 The role of the hypoxic microenvironment and HIF1 expression in cancer

1.2.5.1 Mutational rate of HIF genes in cancer

Through the study of hypoxia adaptation in cancer, it has become clear that it is a driving force for tumour cell adaptation, and a key regulator for this adaptation is HIF1α (**Figure 4**)[164–166]. Hypoxia is a common feature of tumour biology, owing to the hyper proliferation of cancer cells

and dysregulation of angiogenesis leading to ineffective vasculature[123,167,168]. Any tumour greater than 2mm across will contain hypoxic areas for these reasons[169–171]. As a result, tumour hypoxia is found in 90% of patient solid tumours[172–174]. Tumor hypoxia can develop in a stable area, which leads to necrosis and tumour cell death, or can fluctuate as the result of dysregulated vasculature[175]. Due to the ability of HIF1 to induce angiogenesis, HIF1 at least has a role in promoting tumour growth and metastases through the promotion of angiogenesis[176].

Given the role of HIF1 in tumour progression, it could be hypothesized that *HIF1α* is frequently mutated for gain of function in tumours. Surprisingly, mutational analyses of the *HIF1α* locus by Albanese and colleagues using the COSMIC database revealed that *HIF1α* has a low mutational rate of 1.5% in patient tumours, with this mutation rate dropping to <1% when quantifying missense mutations specifically[84,177]. This is in stark contrast to the *TP53* gene, which has a mutation rate of 28% in the same database of patient samples[84]. Nevertheless, HIF1α protein is frequently overexpressed or expressed under normoxic conditions in tumours[168]. Therefore, cancers are frequently dysregulating HIF1α to allow for its function in non-hypoxic settings without directly mutating it. As mentioned previously, there is significant evidence for the dysregulation of the HIF1α regulatory pathway which can lead to non-canonical HIF1α stabilization[178].

1.2.5.2 HIF1-mediated metabolic changes

Due to the significant expression of HIF1α in patient tumours as well as cancer cell lines, investigations have been done to determine the transcriptional targets of HIF1[179]. It was found that HIF1 is capable of regulating a plethora of genes associated with a wide range of phenotypes[82,180]. Due to this wide range of phenotypic control, hypoxia adaptation is now considered one of the most fundamental properties of solid tumours[22,181,182]. One of the largest of these sets of phenotypes is glucose metabolism. As a result of lower oxygen levels, cells convert their energy generation from the oxygen and mitochondria-dependant oxidative phosphorylation pathway to oxygen-independent glycolysis. HIF1 regulates several genes in the glycolysis pathway, from glucose import, glucose metabolism, to metabolic waste export, increasing the metabolic flux of glycolysis to survive hypoxic conditions[82]. HIF1 also upregulates the conversion of pyruvate to lactate and the exporting of lactate into the extracellular space

through the upregulation of lactate dehydrogenase A (*LDHA*) and monocarboxylate transporter 4 (*MCT4*, also known as *SLC16A3*). This allows for the regeneration of NAD+ for glycolysis without the need for oxygen. HIF1α also upregulates pyruvate dehydrogenase kinases, which reduces the entry of pyruvate into the tricarboxic acid (TCA) cycle, and reduced the flux of the oxidative phosphorylation pathway[183].

1.2.5.3 Hypoxic influence on apoptosis

In addition to metabolic changes, HIF1 has also been shown to influence apoptosis. The effect of HIF1α on apoptosis appears to be dependant on cancer type and cancer clone context. In support of the role of HIF1 in inducing apoptosis, HIF1α was shown to directly stabilize p53, resulting in increased rates of hypoxia-induced apoptosis[184]. HIF1 was also shown to induce expression of *BNIP3*, a pro-apoptotic gene[185]. However there is also evidence of HIF1 conferring apoptosis resistance by Akakura et al, showing that HIF1α expression is necessary to reduce hypoxia and starvation-induced apoptosis[179]. HIF1 was also shown to increase levels of VEGF, which among other functions, provides an anti-apoptotic effect in an auto/paracrine fashion[186]. Hypoxia has also been shown to increase levels of the anti-apoptotic protein IAP2, which in turn reduces the functionality of the pro-apoptotic protein BAX[187,188]. These conflicting data resulted in the hypothesis that the hypoxic microenvironment leads to the selection of apoptosis resistant cancer cell clones[189]. Thus hypoxia can influence the genetic landscape of the tumour, and select for clones which are more resistant to the hypoxic microenvironment. By adapting in this fashion, clones also become more aggressive and resistant to therapeutics[189,190].

In line with this hypothesis, hypoxia appears to promote tumour cell genomic instability through multiple mechanisms. First, hypoxia and HIF1 have been shown to down-regulate all DNA repair pathways, resulting in an increased accumulation of mutations[182,191]. Secondly, hypoxia promotes genomic rearrangements and polyploidy, further extending the possible mutational possibility space[192,193]. These mechanisms synergize to increase genomic instability and accumulation of mutations, which are selected for therapeutic and apoptosis resistance by the nature of their environment[194,195].

1.2.5.4 Hypoxic influence on therapeutic resistance

In addition to selecting for apoptosis-resistant cancer clones, hypoxia also contributes to tumour aggressiveness by directly promoting resistance to various therapeutics. Hypoxia has been shown to upregulate the expression of multidrug resistance genes including *MRP1-3* and *MDR1*, which remove chemotherapeutic agents from the cell, reducing their efficacy[196–198]. The hypoxic microenvironment can also reduce cell proliferation, further reducing the efficacy of these therapeutics[199–202]. Also, the nature of the hypoxic microenvironment being difficult for oxygen and nutrients to diffuse into also reduces the efficacy of systemically administered therapeutics[203,204]. The extracellular pH of the hypoxic tumour microenvironment is acidic as a result of lactate export and the activity of hypoxia-induced carbonic anhydrase 9 (CA9 in humans, CAR9 in mice)[205–207]. This increased acidity can influence chemotherapy activity and influence absorption rates, reducing efficacy[208,209].

Hypoxia has also been linked with promoting cancer stem cells. Studies in pediatric neuroblastoma and rhabdomyosarcoma cell lines found that cancer stem cells migrated to hypoxic areas, providing a link between hypoxia and stem cell phenotypes[210]. Further work demonstrated that hypoxia and HIF1α activity motivated cancer stem cell multi-potency and self-renewal[211]. In glioblastoma cells, it was found that hypoxic exposure changed the cell's gene expression pattern to a more premature, and hence stem cell-like, phenotype[212]. Phenotypic properties of cancer stem cells such as the ability to form spheroids and express stem cell marker genes were augmented under hypoxia, along with upregulation of stem cell associated transcription factors OCT4 and SOX2[213]. The maintenance of these phenotypes appears to be regulated more by HIF2 rather than HIF1 [52,214]. These adaptations allow hypoxic tumour cells to more readily repopulate the tumour microenvironment after ineffective treatment, often with more aggressive cancer clones[215].

The hypoxic tumour microenvironment also inhibits immunotherapy through several mechanisms. PD-L1 has been established as a HIF1 target gene, resulting in hypoxic tumour cells being intrinsically more immunosuppressive[216]. Hypoxic shedding of MHC class I chain-related (MIC) protein provides protection of NK cell based killing[217]. Tumor hypoxia also recruits immune suppressor cells (MDSC cells) to the microenvironment and inhibit NK cell function

directly[218]. Tumor hypoxia has also been shown to recruit immunosuppressive T-regulatory cells through the induction of CC-chemokine ligand 28 (CCL28), which promotes tumour immune tolerance[219]. Vascular endothelial growth factor (VEGF) expression by hypoxic tumour cells inhibits dendritic cell maturation and increases PD-L1 expression levels in dendritic cells, resulting in further reduced T cell function[220–222]. As such, the hypoxic tumour microenvironment is generally inhibitory to proper immune cell function.

In summary, the hypoxic microenvironment is associated with therapy resistance with regards to radiotherapy, chemotherapy, targeted therapy, and immunotherapy. This is due in part to the reliance of certain therapeutics on a minimal amount of oxygen to be functional. But it is also due to intrinsic resistances provided by the hypoxia adaptation program.

1.2.5.6 Hypoxia influence on metastatic potential

A tumour phenotype often correlated with tumour hypoxia is an increase in metastases. Indeed, some of the genetic targets of the HIF1 transcriptional program and the hypoxia response more generally increase the likelihood of metastases[182,223,224]. As the natural long-term relief from hypoxia is the synthesis of new blood vessels, a portion of the hypoxia response is to signal and induce angiogenesis towards the hypoxic region. This is accomplished through the expression of various pro-angiogenic factors: primarily VEGF, as well as other factors including angiopoietin-2 (ANG-2), Tie2, platelet-derived growth factor (PDGF), and basic fibroblast growth factor (bFGF), as well as other factors[225–230]. This increase in vasculature does not always alleviate the hypoxia, as tumour vasculature growth is often ineffective and leaky[231]. These leaky blood vessels facilitate extravascular invasion, promoting metastases[224].

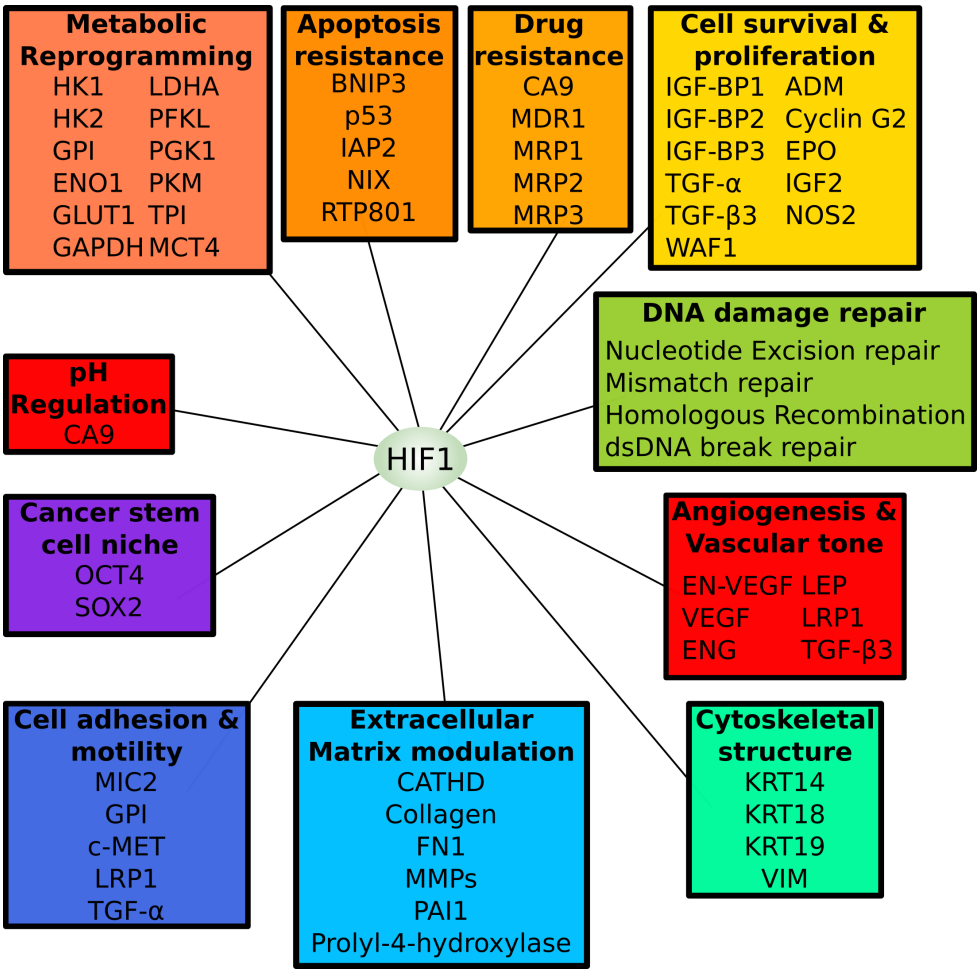

Figure 4: phenotypes controlled by HIF1 and hypoxia

A selection of phenotypes and associated genes regulated by HIF1. Adapted from REF [82]

Additionally, the hypoxia response has been shown to make tumour cells intrinsically more likely to metastasize. In order to metastasize, tumour cells must increase their motility. This occurs through the activation of epithelial to mesenchymal transition (EMT) associated transcription factors, which make the cells more motile and invasive[224]. This necessarily makes them more likely to metastasize. In recent work done by Godet and colleagues, they demonstrated that tumour cells exposed to hypoxia become motile in order to escape the hypoxic microenvironment using a fate-mapping approach[174]. They demonstrated that hypoxic tumour cells move away from the hypoxic microenvironment and towards blood vessel, are

more likely to be circulating tumour cells in the bloodstream, and survive the mechanical and physiological stresses of the bloodstream[174].

At a molecular level, hypoxia has been shown to promote EMT, migration, and metastases in a multitude of cancers through multiple mechanism by increasing the motility of cancer cells and modifying the extracellular matrix[232]. Hypoxia and HIF1 have been shown to upregulate the transcription factors classically associated with EMT, such as, TGFβ, TWIST, SNAI1 [182,223,233–236]. Through the activation of these transcription factors, tumour cells become more motile and likely to metastasize. The exact mechanism of action varies from context to context, but hypoxic adaptation overall is associated with an increase in motility and likelihood to metastasize.

HIF1 has also been shown to increase levels of extracellular matrix modifying genes such as the matrix metalloproteinases[237]. These enzymes are used to remodel the extracellular matrix and allow cells to move more freely[238]. HIF1 also upregulates the expression of prolyl 4-hydroxylases and lysyl oxidases, which modify intracellular collagen fibres, reducing collagen cross-linking and reducing the stiffness of the extracellular matrix, easing cellular motility[239,240]. Therefore, hypoxia and the HIF1 transcriptional program enhance tumour cell motility by upregulating EMT transcription factors, extracellular matrix modulating enzymes, and pro-angiogenic factors; resulting in an increase in metastatic potential, both within the cell and beyond.

Due to this increase in metastatic potential, hypoxia is now considered one of the foundational aspects of the tumour microenvironment[22]. We now know that the majority of primary tumours and metastases over-express HIF1α, and based on the transcriptional program it activates, it is obvious why[168]. The hypoxia adaptation transcriptional program confers several survival advantages to tumour cells which express it (**Figure 4**). These advantages come in the form of metabolic shifts, increased resistance to apoptosis, increased rate of mutation, increased motility, and increased probability to metastasize. As such, tumours which cannot express HIF1α have a dramatically higher patient survival rate, as the tumour cannot engage in the beneficial transcriptional program[241].

Since the majority of patient mortality from cancer is the result of metastases, limiting or preventing metastases could dramatically reduce patient mortality[242]. Toward this end, anti-VEGF immunotherapies have been developed to attempt to reduce angiogenesis and the probability of metastasizing. While the therapeutic works initially, the benefits are temporary, and the tumour escapes the therapeutic by upregulating alternative angiogenesis pathways such as bFGF[229]. This highlights a common theme in cancer therapeutics, where therapeutics are initially beneficial, only for the tumour to escape the therapeutic via an alternative pathway. Due to this, it is becoming increasingly clear that a deeper understanding of the adaptive capacity of cancer is needed. Additionally, a multi-targeted approach is needed to effectively treat cancer to account for the tumour adaptive landscape.

1.3 Current methods & drawbacks for *in vitro* HIF1α stabilization

In the field, *in vitro* HIF1α stabilization is achieved in two main ways: chemical stabilization or *in vitro* hypoxic culture. Since HIF1α protein is constitutively being synthesized, chemical induction of HIF1α needs only to subvert one of the normoxic degradation steps to achieve normoxic HIF1α stabilization. This is done most commonly using divalent metal ions, most commonly $CoCl_2$, iron chelants such as Deferoxamine (DFO), 2-oxoglutarate analogs such as Dimethyloxaloglycine (DMOG), electron transport chain uncouplers such as Dinitrophenol, or protease inhibitors such as MG132[243–247]. Despite $CoCl_2$ being one of the most commonly used chemical stabilizers of HIF1α, the mechanism of action by which it operates is still unknown. It increases both HIF1α and HIF2a protein levels in a dose and time dependant manner, with differing efficacy and toxicity varying from cell line to cell line[248]. As a result, $CoCl_2$ treatments are rarely consistent across experiments and cell lines[141].

Currently, there are three differing hypotheses to explain Co^{2+} induced HIF1α stabilization: the replacement hypothesis, the iron oxidation hypothesis, and the direct binding hypothesis[141]. In the replacement hypothesis, Co^{2+} replaces the Fe^{2+} ion in the PHD active site, inactivating the enzyme. In the iron oxidation hypothesis, intracellular Co^{2+} results in an increase of intracellular reactive oxygen species (ROS), resulting in oxidation of Fe^{2+} to Fe^{3+}, inactivating

PHD enzymes. In the direct binding hypothesis, intracellular Co^{2+} ions bind directly to HIF1/2α or the VHL 2E ubiquitin ligase complex, preventing the ubiquitination of hydroxylated HIF1/2α.

Despite the mechanistic uncertainty behind $CoCl_2$, it has remained as one of the most popular ways to chemically stabilize HIF1α. This is due to the historical precedence of $CoCl_2$ treatment, as well as the evidence showing $CoCl_2$ inducing expression of some of the classical hypoxia-inducible proteins such as VEGF and glycolysis related genes[249,250]. However, comparative analyses between $CoCl_2$-treated and hypoxia-incubated cells confirmed that $CoCl_2$ treatment does not completely recapitulate the hypoxia adaptation response, and regulated certain genes in a $CoCl_2$ specific manner[251,252]. Therefore while $CoCl_2$ treatment and other chemical stabilizers of HIF1/2 are useful tools for normoxic stabilization, they are no substitute for genuine hypoxic incubation when examining the breadth of hypoxic adaptations.

Besides chemical stabilization, the next most frequently used method for HIF1/2a stabilization is *in vitro* hypoxic incubation. Using completely sealed cell culture incubation chambers, researchers are able to expose cells to atmospheres with differing oxygen concentrations. These culture chambers vary from simple heated and humidified boxes to multi-chambered and computer-regulated systems allowing long-term incubation and cell culture work in a completely hypoxic environment[253,254]. These incubation chambers allow for natural HIF1/2a stabilization and can be more representative of *in vivo* stabilization. However much like with $CoCl_2$ treatment, there is inconsistency in hypoxic incubation protocols across the literature. In both severity of hypoxia and length of incubation time there is considerable variation. Incubation protocols can vary from hours to days, and typical oxygen concentrations can vary from 0.1% to 3% O_2[141]. Only recently have hypoxic incubation protocols extended beyond a few days, primarily in the cyclic hypoxia research field[255]. As such, one must be careful to examine the hypoxic incubation on a case by case basis, as a protocol involving direct incubation at 0.1% O_2 will induce different adaptations than incubation at 3% O_2, due to the different oxygen concentration differential.

In addition to hypoxic culture, researchers have also introduced a reduction of FBS in conjunction with hypoxic incubation[256]. The purpose of this is to better mimic the tumour

microenvironment, as *in vivo* hypoxic regions often have a reduction in available nutrients and growth factors due to the same diffusion and consumption principles as oxygen[231]. By removing FBS from the media, there is a reduction of nutrients and growth factors available to the cells. Typically, this kind of treatment is referred to as "serum starvation". However much like *in vitro* hypoxic incubation, there is considerable variability between serum starvation protocols. Serum starvation can range from complete removal of FBS from the media, to a reduction of 50% from standard culture conditions[186,256,257]. In combination with serum starvation, the base media can also be changed to further adjust nutrient availability. For example, work done by Akakura and colleagues generated a cell culture environment lethal to pancreatic cancer cell lines by combining hypoxic culture with serum and glucose free media, which was rendered tolerable through HIF1α overexpression[179]. While it is obvious that serum starvation induces a differential effect on cellular adaptation, serum starvation in conjunction with hypoxic incubation is not a wide-spread practice. Most *in vitro* hypoxic incubation protocols are done using the cell's standard media, and hypoxic exposure is typically at 1% O_2 for no more than 48 hours.

1.4 Regulation of the Epithelial to Mesenchymal Transition (EMT)

As previously discussed, hypoxia adaptation can lead to an increase in metastases, which is in part regulated by an increase in motility of hypoxia-adapted cells. EMT is a reversible cellular phenotype where morphologically epithelial cells lose their normal cuboidal or "cobblestone" shape, and adopt a more spindled and motile mesenchymal morphology[224]. Once in the mesenchymal state, cells can return to their original epithelial morphology through a mesenchymal to epithelial transition (MET). Mechanistically, this is achieved through the differential regulation of cell-cell adhesion complexes and the cytoskeleton, which is orchestrated by specific transcription factors[258–260]. In tumours, cells which undergo EMT are thought to seed metastases through the invasion-metastasis cascade: in which primary tumour cells become mesenchymal, enter the circulatory system through intravasation, travel through the circulatory system, exit circulation through extravasation, and form micrometastases which eventually grow into macroscopic metastases through MET[261–264].

EMT can be broadly categorized in three main types: developmental EMT, tissue regeneration EMT, and cancer-associated EMT[260]. As an example of EMT during development, neural crest cells undergo EMT to migrate to specific areas before giving rise to different derivative tissues such as melanocytes in the skin[260,265,266]. This EMT is primed by the transcription factors SOX, SNAIL, SLUG, and forkhead box D3 (FOXD3); and is later triggered by a combination of several signalling pathways including Wnts, FGFs, BMPs, c-Myb, and msh homeobox 1 (Msx-1) signalling pathways, with BMP contributing the most to the migratory properties[266–272]. This cascade of transcription factor activity results in the loss of expression of E-cadherin and the upregulation of N-cadherin, two cell adhesion genes, which is necessary for the cell's motility[273,274]. Elements of the cytoskeleton, such as the intermediate filament protein vimentin (VIM), are also upregulated during this form of EMT[275].

The second type of EMT is associated with tissue regeneration and wound healing, and has primarily been investigated using kidney fibrosis[276–278]. During this process, cells induce the expression of several motility associated proteins such as S100 calcium binding protein A4 (S100A4, also known as fibroblast specific-protein 1 FSP1), alpha smooth muscle actin (ACTA2), and VIM[260,279–281]. In addition, they also upregulate extracellular matrix remodelling proteins such as collagen (specifically COL1 proteins) and matrix metalloproteases such as MMP1,2,8,13[238,280]. This allows for the localized digestion and remodelling of the extracellular matrix and basement membrane, which is necessary for tissue remodelling and wound healing[282]. Much like with type I EMT, a myriad of signalling pathways and transcription factors regulate it including the fibroblast growth factor-2 (FGF2), PDGF, epidermal growth factor (EGF), and transforming growth factor beta (TGFβ) pathways[282–285]. The expression of this combination of genes allows for cells to simultaneously digest, regenerate, and repopulate areas of tissue which have been damaged.

The third type of EMT is associated with cancer invasion into the surrounding tissue and metastases to distal tissues. Through immunohistological analyses of patient samples, the ability for tumour cells to invade local healthy tissue was the first indication of cancer using EMT to promote disease[260]. The expression of EMT type I and II driver and effector markers are commonly found to be expressed in cancer cells as well. Overexpression of VIM, FN, ACTA2, and

others are often detected at the invasive front of the tumour, and these cells are considered to be the tumour cells which seed metastases through the invasion-metastasis cascade[286]. Despite the commonality of EMT in cancer the exact trigger for inducing EMT is unknown, although signalling from hepatocyte growth factor (HGF), EGF, PDGF, and TGFβ pathways are implicated. While the complete list of type III EMT driving pathways is still unknown, several signalling pathways are commonly found to drive tumour cell EMT such as transcription factors Snail (SNAI1), Slug (SNAI2), ZEB1/2, TWIST1/2, FOXC2, Goosecoid (GSC) GRHL2, and EZH2[287–300]. Generally speaking, the expression of SNAI1/2, TWIST1/2 FOXC2, or EZH2 are sufficient to drive EMT in cancer cells[277,291].

1.5 The Role of Wilms Tumor 1 gene (WT1) in cancer

1.5.1 WT1 effector functions in cancer & development

As mentioned earlier, the list of EMT-driving transcription factors in cancer is not complete. One additional transcription factor of this list is Wilms' tumour protein 1. It was originally discovered in the most common form of pediatric renal cancer (Wilms Tumor), where the gene was often genetically inactivated. This earned *WT1* its name, research focus, and initial classification as a tumour suppressor[301–303]. Much like other EMT driving transcription factors, WT1 has a crucial role in development, particularly in renal development; but also in heart, adrenal gland, and gonad development[304,305]. *WT1* knock-out results in embryonic lethality in mice due to the lack of maintenance of the mesenchyme state of progenital cells causing improper development of the kidneys, ultimately resulting in death[306–308]. During kidney development, the role of WT1 is not limited to maintaining a mesenchymal state in cells. WT1 also induces a MET that is essential for proper nephron development[309–311]. The role of WT1 in development is not limited to the kidneys. WT1 is essential for proper development of the vasculature of the heart as well. In contrast to the role of WT1 in nephron development, the role of WT1 in heart development is to induce EMT[306,312]. WT1 achieves this in part through upregulation of SNAI1 (an EMT driver), as well as through downregulation of migration inhibiting cytokines CXCL10 and CCL5 by downregulating the transcription factor *IRF7*[312,313]. The research done on the role of WT1 in development has concluded that WT1 can have opposing

functions: in one context as a driver of EMT, and in another a driver of MET, through the promotion or repression of transcription[304,314].

As mentioned previously, WT1 was originally classified as a tumour suppressor due to its role in the development of Wilms tumours. However later work found WT1 to be overexpressed in a variety of cancers including colorectal, breast, desmoid, brain, and acute myeloid leukemia (AML)[315–321]. Overexpression of WT1 in these contexts suggests it has a oncogenic rather than tumour suppressive role. Further supporting this hypothesis, it was found that WT1 expression was necessary to prevent apoptosis in breast cancer and leukemia cell lines[129,322,323]. WT1 was found to be a driver of EMT through the upregulation of SNAI1 in VHL-deficient clear cell renal carcinomas[324]. Supporting the role of WT1 as an EMT driver in cancer, knock-down of WT1 expression was shown to induce MET in several human mesothelioma cell lines[325]. Using a combination of Chromatin Immunoprecipitation (ChIP) based experiments and reporter assays, several transcription targets of WT1 have been determined in a cancer context[326]. Thematically, WT1 transcriptional targets can be grouped into the following categories: growth and development, differentiation, cytoskeletal organization, cell adhesion, WNT signalling, MAPK signalling, apoptosis, and epigenetic regulation[326]. As such, it is possible that WT1 could drive a potent tumour progression program through selective regulation of specific target genes. However, exactly how WT1 orchestrates its transcriptional program to control certain genes and not others in specific contexts, is still not completely understood.

1.5.2 WT1 protein structure & isoforms

One mechanism through which WT1 can exhibit differential function is through the expression of different isoforms. Located at chromosome 11p13, the *WT1* locus consists of ten exons (**Figure 5**)[301]. The *WT1* gene spans 47.7kb in humans, with the longest spliced mRNA isoform reaching just over 3kb. The first five and last five exons are separated by a long intron, spanning ~20kb. While not unheard of, this intron length places it within the top 10% longest introns in the human genome[327].

The first *WT1* promoter, which transcribes exon 1 and extends into intron 1 has been shown to be bi-directional and results in the transcription of an anti-sense long non-coding

transcript (lncRNA) known as WT1-AS[328,329]. Transcription from this *WT1* promoter results in expression of both *WT1* and *WT1-AS* transcripts[330–332]. A body of work has shown that WT1-AS lncRNAs have both positive and negative effects on WT1 protein levels and overall tumor pathology through multiple mechanisms[333]. WT1-AS has been shown to interact with the *WT1* promoter TATA-box, down-regulating WT1 mRNA and protein levels, and promoting apoptosis in adeocarcinoma cells[334]. WT1-AS lncRNAs have also been shown to interact directly with WT1 transcripts through their shared *WT1* exon 1 sequence and positively modulates WT1 protein levels[331,332]. The effects of WT1-AS expression is also mixed with regards to tumor pathology. High levels of WT1-AS expression negatively correlates with patient survival in cases of breast and colorectal cancers, while the opposite is the case in cervical, ovarian, liver, stomach, and kidney cancers[333,335].

Much like the gene, the WT1 protein structure can be broadly divided by the first and last 5 exons (**Figure 5**). The N terminal domain contains several domains necessary for various protein-protein interactions such as with BASP1 and HSP70, as well as a dimerization domain[304,336–338]. The N terminal side also contains regions of proline-glutamine rich sequences which are predicted to be involved in RNA binding[339]. The C terminal side consists of four C_2H_2 Krüpple-like zinc finger domains, which are responsible for DNA/RNA binding, as well as nuclear localization[340,341]. There are two nuclear localization sequences within the zinc fingers, one within zinc-finger 1, and the second spanning zinc-fingers 2&3[338,342]. The zinc finger domains also allow for protein-protein interactions with p53, preventing apoptosis; p73, inhibiting WT1 transcription; Creb binding protein CBP, enhancing WT1 transcription; and the prostate apoptosis response factor Par-4, suppressing WT1 transcription[338,343–346]. Whether or not WT1 functions as a transcriptional activator or repressor depends on its binding partners[326].

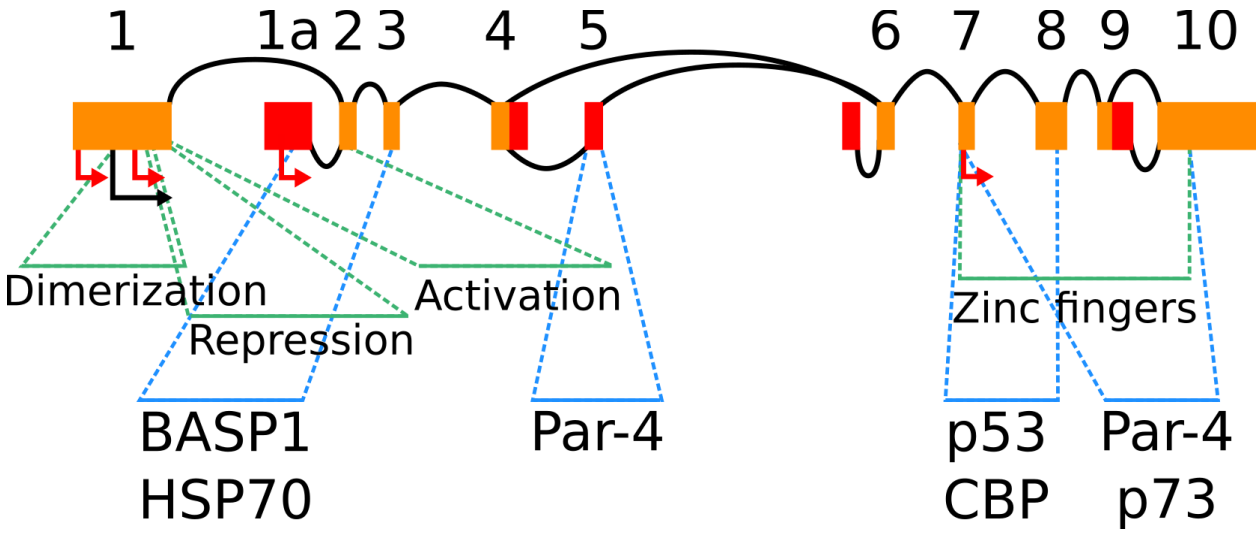

Figure 5: WT1 structure and isoforms

Diagram of WT1 mRNA splicing paths and alternative translation start sites. Top numbers represent exon numbers. Red exons correspond to exons included as a result of an alternative splicing event. Canonical translation start site is denoted by the black arrow, alternative translation start sites are denoted with red arrows. Approximate locations for functional WT1 domains (green) and protein-protein interaction (Blue) sites are listed below.

Through a combination of alternative transcription start sites, splicing paths, and alternative non-canonical translation initiation codons, the total number of potential WT1 protein isoforms has reached 36[304,305]. Currently in the Ensmebl database, there are 22 annotated WT1 mRNA isoforms, of which only ten have evidence of protein-production[347]. Broadly speaking, the mRNA isoforms can be classified based on their TSS. Almost all of the protein coding mRNA isoforms initiate transcription from either exon 1 or 1a. The isoforms with unknown protein production capability initiate transcription much later in the gene, and from canonically "intronic" sequences, especially from within the large intron of the *WT1* gene (**Figure 5, intron 5**). These isoforms were originally identified by the Havana project in RNAseq data, however many of these isoforms remain univestigated[347]. Most work has focused on two different splicing events arising from the exon1/1a class of transcripts: the inclusion or exclusion of exon 5, and the inclusion or exclusion of the lysine-threonine-serine (KTS) motif between exons 9-10 (**Figure 5**)[338,348]. Beyond these splicing paths, there are also alternative transcription start sites within intron 1 and 6 which give rise to alterative WT1 isoforms[305,349,350]. Little is known

about the functional significance of these two alterative transcription start sites and their role in WT1 biology.

The inclusion of exon 5 into the mRNA results in the introduction of a Par-4 binding site as well as alters the functionality of WT1, resulting in a stronger morphological change in the cells[304,338,351]. This isoform also prevents apoptosis in leukemic cells[323]. However deletion of this exon does not result in embryonic lethality[352]. For the KTS motif, having exclusive expression of either isoform results in death soon after birth in mice due to defective kidney function[353]. Mutations in the KTS splice site resulting in altered proportions of KTS+/- result in non-lethal developmental defects classified as Frasier syndrome[354]. The KTS insertion results from a 9 nucleotide extension of exon 9 at the 3' end, and inserts the KTS motif between zinc fingers 3 and 4. Insertion of the KTS motif inhibits zinc-finger 4 from interacting with DNA, resulting in KTS+ WT1 isoforms having a much lower affinity for DNA, although DNA binding is still possible[304,355,356]. However both isoforms are able to bind to RNA equally[341]. Recent work by Nishikawa and colleagues has demonstrated that both KTS+ and KTS- isoforms of the WT1 zinc-finger domains are able to bind to the same 29mer looped RNA structure, and the KTS insertion only affects DNA binding affinities, not RNA[341].

Under physiological conditions, the KTS+ isoform is the dominant form of WT1, at a 2:1 ratio. This suggests that while KTS- WT1 is functioning as a transcription factor, KTS+ WT1 functions more as a RNA binding protein[348]. Through co-immunoprecipitation (Co-IP) experiments, WT1 KTS+ was shown to interact with RNA splicing complexes via protein-protein interactions with the splicing factor U2AF65[357,358]. WT1 KTS+ has also been shown to regulate translation rates of specific mRNAs by directly binding to specific RNA sequences[358]. This RNA binding occurs preferentially in the 3'UTR of target transcripts, and also protects bound targets from degradation[359].

1.6 The biological significance of miRNAs in health, disease, and therapeutics

Within all eukaryotic forms of life, there exists a RNA-interference (RNAi) pathway, where protein output is limited in an mRNA sequence-dependent fashion[360]. This pathway

functions using RNA binding protein (RBP) complexes, which use a small RNA sequence as guide for target sequence inhibition. If the target sequence region hybridizes sufficiently to the guide sequence, mRNA protein output will be limited through differing mechanisms, depending on the degree of hybridization[361]. If there is insufficient hybridization between the guide and target, the target molecule is not acted upon[361]. If there is a high degree of complementarity between the target and guide, the protein complex will cleave the target sequence, resulting in degradation of the target molecule[362,363]. If there is a moderate amount of hybridization, the RBP complex will recruit other RBPs resulting in translational repression and indirect target degradation. It is in this category of moderate hybridization that mircoRNAs typically function.

MicroRNAs (abbreviated to miRs or miRNAs) are short (17-22nt) non-coding RNAs which primarily regulate cellular protein levels through regulation of their target mRNAs[361]. They exert this function by hybridizing to sites on their target sequence, often found in the 3' untranslated region (UTR) of the mRNA[364–366]. Originally discovered as essential genes for the proper development of *C. elegans* in the 1990s, microRNAs have since been found in plants, viruses, and in all bilaterian animals, including mice and humans[366–369].

1.6.1 MicroRNA biogenesis

MicroRNAs are generated through a dedicated biogenesis pathway, which begins at transcription from genomic DNA, generating the primary miRNA transcript (pri-miRNA) (**Figure 6A**). The majority of microRNA genes exist as distinct transcriptional units, similar to protein-coding genes. However they can also be found within the introns of protein-coding genes, and are transcribed independently or when their host-gene is transcribed[370]. The pri-miRNA naturally adopts a secondary structure based on its sequence, forming a hairpin structure. This structure is recognized by the pri-miRNA microprocessor protein complex, composed of DROSHA and DGCR8[371]. The hairpin structure is cleaved from the surrounding sequence by DROSHA, and the remaining hairpin RNA, now termed a pre-miRNA, is exported to the cytoplasm via Exportin-5[372–374]. Once in the cytoplasm, the pre-miRNA is further processed by second protein complex composed of DICER and TRBP[375]. Once transferred to the DICER/TRBP complex, DICER cleaves both arms of the pre-miRNA near the loop, generating a double-

stranded miRNA-duplex[376]. Once processed by DICER, the miR-duplex is loaded into an Argonaute (AGO1-4) protein with the assistance of chaperone proteins HSC70 and HSP90[377]. Once loaded, the Argonaute protein discards one strand of the duplex. The rules governing which strand is loaded and which is discarded are not completely understood, although there is a preference for loading strands with a 5' adenine or uracil base[378,379]. After discarding one strand of the duplex, the AGO protein is properly loaded and is referred to as the RNA-induced silencing complex (RISC), and is now ready to inhibit mRNA protein output in a miR sequence dependant manner[380–382].

1.6.2 Functionality of miR sequence regions

The functionality of any given RISC is dependant on the degree of hybridization to the target sequence region and which AGO isoform is being used. There are four AGO homologs, of which only AGO2 contains RNA cleaving activity[362,383,384]. The other three homologs can only repress mRNA protein output through indirect methods[384]. If the RISC complex is composed of AGO2 and the loaded guide sequence completely hybridizes with the target sequence, the RISC complex cleaves the target sequence[362]. AGO2 cleaves the target RNA through the use of its catalytically active PIWI domain[385,386]. The PIWI domain is structurally similar to RNase H, and contains endonuclease activity[387]. When the guide sequence completely hybridizes with the target sequence, the PIWI domain cleaves the phosphodiester bond on the target sequence directly opposite of the 10^{th} and 11^{th} nucleotides of the guide strand[385,386]. This cleavage of the target RNA results in the degradation of both target RNA fragments by endogenous riboexonucleases[388,389] .

This RNA degradation pathway has been used extensively in research to endogenously reduce the expression level of any given gene without knocking out the gene permanently. As this technique reduces rather than eliminates expression of a gene, it is colloquially referred to as 'knocking-down' a gene. Knock-down experiments can be done through transfection of previously generated RNA duplexes meant to mimic pre-miRNAs called small interfering RNAs (siRNA)[390,391]. Other techniques use transfection or transduction to express short hairpin RNA (shRNA) sequences from a pol II or III promoter, which are designed to mimic the pri-miRNA

hairpin structure[392]. Both siRNAs and shRNAs are designed to target one target site, and fully complement to the intended site. Expression of these RNAi molecules result in downregulation of the target mRNA, and subsequently the encoded protein.

In contrast to shRNAs and siRNAs, microRNAs at first glance share little complementarity with their known sites[393]. In line with their lack of complementarity, miRNA mediated target sequence cleavage is the exception to miR function, not the rule[361,393]. Instead, miR:RISC binding to a target mRNA results in translational repression and mRNA decay through indirect RISC cleavage independent mechanisms. The process begins with the recruitment of the scaffolding protein TRN6 (also known as GW182) to the AGO protein[394]. TRN6 orchestrates mRNA decay by promoting both polyA tail deadenylation and subsequent 5' de-capping[395]. First, TRN6 displaces the polyA-binding-proteins (PABP) from the polyA tail and recruits the CCR4-NOT deadenylation complex to the polyA tail[396–398]. This results in the degradation of the mRNA from the 3' end[399]. Recruitment of the CCR4-NOT complex allows for the further recruitment of the scaffold protein DDX6 and the subsequent recruitment of decapping factors DCP1–DCP2[394,400,401]. Once the mRNA is decapped, it can be targeted by 5' to 3' exonucleases[401,402]. DDX6 also recruits eIF4E-transporter protein (4E-T), which reduces translation of the target mRNA by sequestering the 5' cap from translation initation factors[403]. Thus, while miRNAs may not induce direct cleavage of their target RNA due to reduced guide RNA binding, they nonetheless can reduce target RNA and protein production through deadenylation, decapping, and translational repression.

Through the study of the rules governing the microRNA efficacy on a given site, it became clear that there was significant complexity. While microRNAs didn't have significant hybridization to their target sites overall, the miR almost always had perfect Watson-Crick binding for the first eight nucleotides[361,404,405]. This led to the division of the microRNA sequence into two regions: the seed and the tail. Analysis on the conservation of miR sequences and known target sites narrowed the seed sequence to nucleotides 2-8, with the tail sequence taking up the remaining sequence.

While the seed is classically defined as miR nucleotides 2-8 and miR target sites are thought to have Watson-Crick complementarity to this entire region, it is now known this is not

necessarily the case[361]. There is some flexibility in seed binding profiles to canonical sites, with hybridization ranging from miR nucleotides 2-7 to nucleotides 2-8, with more hybridization correlating with more efficacy[361,404,406]. There are also examples of miRs binding to sites with 6 seed nucleotides, however these sites are always compensated for with nucleotide binding at either end of the 6mer, generating an off-set 7mer[404,407,408]. In addition to these canonical miR binding sites, non-canonical sites are known to exist and display function, although they make up a very small minority of known sites[405]. These sites contain a mismatch within the seed duplex, but compensate for it with significant miR-tail binding beyond the seed[409,410]. As the seed sequence dictates the binding site of a miR, microRNA genes are organized into families based on their seed sequence[361]. Members of a miR family can be found at the same genetic locus, or in a different location in the genome. While some miR families share a common evolutionary ancestry, this is not always the case. The only criteria for miR family membership is the conservation of the seed sequence[361].

As a result of this much shorter targeting sequence in comparison to si/shRNAs, microRNAs are able to target multiple target sites and mRNAs simultaneously. This has lead to an explosion of microRNA target site prediction software, which all use the seed sequence hybridization as a foundation in their analyses[411,412]. More recent software uses the relative abundancy of microRNAs and target sites to better simulate the microRNA targeting landscape[411].

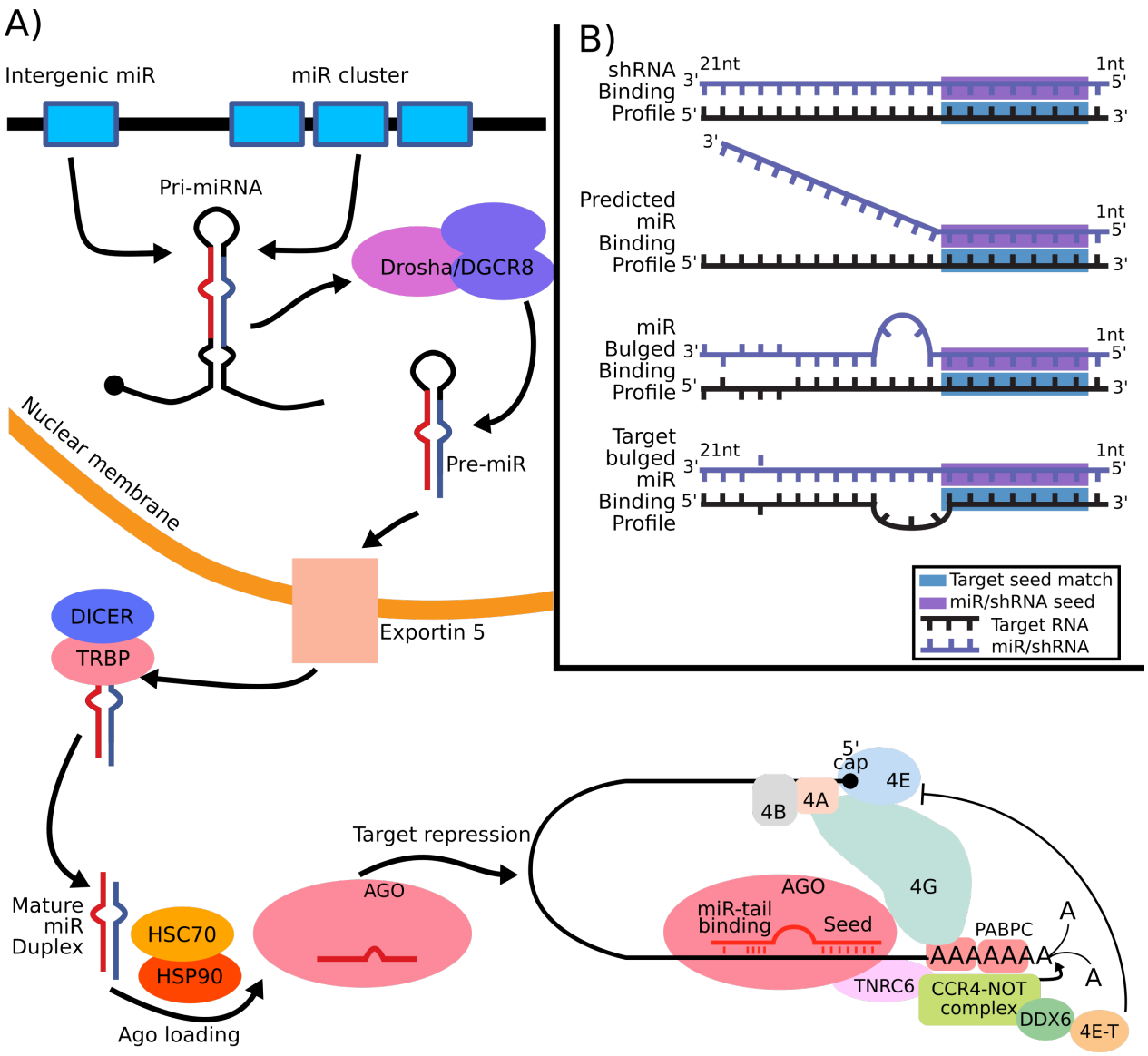

Figure 6: microRNA biogenesis and mechanism of action

A) miR biogenesis and mechanism of reducing protein output. Protein synthesis repression occurs through the recruitment of translation inhibition (4E-T and DDX6) and polyA deadenylation factors (CCR4-NOT complex) by the Trinucleotide repeat containing 6 (TNRC6) adaptor protein. PABPC: Polyadenylate-binding protein, 4E-T: Eukaryotic translation initiation factor 4E transporter . Adapted from REF[449]. **B**) Examples of different RNAi hybridization structures. Top: shRNA binding structure. Second: Original model of miRNA:target binding. Bottom two: New model of miR:target binding, where central sections of the duplex can be unpaired.

While analysis of the functionality of the seed sequence was the initial focus of the microRNA field, recent work has begun to focus on the tail sequence. Research on the tail sequence had been delayed in part due to the higher complexity of the experiments. The increased length of the tail sequence dramatically increased the potential binding modes that are possible. Originally, the tail sequence was thought to have little effect on miR functionality[361,413–415]. However this view has been challenged recently[416]. The advent of advanced high-throughput next-generation sequencing protocols like individual-nucleotide resolution cross-linking immunoprecipitation (iCLIP) and RNA bind-n-seq, has made comprehensive investigation of miR tail binding possible[417,418]. Broughton and colleagues demonstrated that while the target sequence responsible for tail binding was dispensable for AGO2 recruitment to the site, it was nonetheless essential for effective miR targeting *in vivo* using *C. elegans* and chimera PCR[417]. Their group determined that miR-tail binding can have multiple binding patterns, including seed-only binding, weak binding across the entire site, and seed binding with strong tail binding after a small central region of 2-5 nucleotides of non-binding[417]. These results echo previous results presented by Moore et al in 2015, who used an RNA cross-linking and AGO protein pull-down method termed CLEAR-CLIP to sequence local miRNA target sites to describe miRNA:target hybridization architecture in murine brain and human hepatoma cells[416]. While less extensive than the work done by Broughton et al, the work done by Moore demonstrated an enrichment for motifs in the target sequence beyond the seed, demonstrating that miR-tail binding sequences are also conserved[416].

Recent work by Yan et al and Duan et al has further cemented this concept. Through the use of controlled mismatches of nucleotide triplets beyond the seed, Yan et al demonstrated that there was a hierarchy of functionality in an shRNA sequence beyond the seed *in vivo*[419]. This work demonstrated that shRNA nucleotides 12-14 were the most important for functionality, followed by nucleotides 15-17, 9-11, and 18-21[419]. From these results we can conclude that the 12-17 nucleotide region of the shRNA tail is the most critical for tail functionality, a result which re-enforces similar models made in the field[361]. The concept of the importance of this miR region is also supported by recent work done by Duan et al, who demonstrated that mutation of the miR-tail region 11-16 in let-7a had a significant impact on functionality *in vivo*[420]. Their work,

which involved analysis of mutation of both the let-7a miR-tail sequence as well at the known target site on *lin-41*, allowed for the investigation of the mismatch effect on a single nucleotide level. Through this analysis, Duan et al developed a different hierarchy of functional importance across the miR-tail, with an order of 12, 13, 11, 15, 16, 14, which is very close to the order proposed by the work done by Yan et al[419,420]. Through these works, it is now clear the that miR-tail sequence contributes substantial functionality to RNAi as a whole.

Recent work by McGeary et al further expanded on this model using an *in vitro* AGO RNA bind-n-seq approach[418]. They found that miR tail binding appears to adopt two different binding architectures. The first was where the miR tail binds to the target sequence in a similar fashion to siRNA:target duplexes, but with imperfect and non-canonical pairing. The second binding profile was where there was no pairing adjacent to the seed and subsequence binding further away on the target sequence[418]. This second binding profile generates a positive offset between the miR and the target, and often has unpaired miR nucleotides in the central region[418]. This offset structure has been replicated *in vitro* using purified AGO2, miR, and synthetic target sequences by Sheu-Gruttadauria and colleagues, where they determined that binding of miR tail nucleotides 13-16 could occur up to 15 nucleotides away from the seed, provided the tail binding was composed of GC bonds[421].

These findings have resulted in a paradigm shift from the miR-tail sequence being dispensable for miRNA function to miR-tail sequence contributing substantial functionality depending on more complex binding architectures. However, the exact miR-tail region, the importance of nucleotide binding, and the limits of miR-tail binding architectures are all concepts still under investigation[416–420]. There is still much work to be done to develop a predictive model for the functionality of the miR-tail.

1.6.3 The role of miRNAs in homeostasis

Soon after the discovery of the RNAi pathway in mammals, it was quickly discovered how essential the RNAi pathway was to tissue development and general homeostasis. The first microRNA, lin-4 in *C. elegans*, was discovered as a regulator of LIN14 protein levels, which regulated the transition between larval stages of development[367,422]. Bernstein, Wang and

colleagues established that the RNAi pathway was essential for development, as *DICER1 or Dgcr8* null murine embryos failed to develop due to a lack of stem cell development[423,424]. This effect was consistent with both full embryo knock-outs as well as tissue specific knock-outs, highlighting the foundational importance of miRNAs in tissue development[425]. Without these genes, the miRNA biogenesis pathway cannot function, and so these animals have a deficient RNAi pathway which results in embryonic lethality. At the same time, it was discovered that specific miRNAs were found to be expressed in specific tissues, suggesting that specific microRNAs are involved in specific tissue development[426–429]. Curiously, the deletion of tissue-specific miRNAs does not always lead to the loss of tissue development[430]. For example, the heart specific miR-208 is required for heart tissue homeostasis, but *miR-208* null embryos are capable of developing partially functional heart tissue[431,432].

However the deletion of specific miRNAs can lead to a myriad of different developmental defects[361]. In mice, the removal of *miR-1-1/2*, *miR-9-2/3* results in postnatal lethality before weaning, as a result of heart or brain defects[433–435]. The knock-out of other miRs can have less extreme phenotypes than lethality. Loss of miR-22 expression results in an attenuated Th17 response as well as impaired cardiac function[436,437]. In a similar vein, loss of miR-31 results in improved immune cell recovery following an initial viral infection and reduced autoimmune pathology, at the cost of increased susceptibility to chronic infection and tumourigenesis in the colon[361,438–440].

The microRNA-210 also has immunological effects; as its loss of expression results in an increase in spontaneous auto-antibodies, an increase in Th17 T cell development, and an increased pathology in colitis models[441–443]. Beyond the immune system, miR-210 appears to have important roles in the maintenance of retinal cells, sensory hair cells, and the placenta[444–446]. The loss of the miR-200 family and miR-429 results in a loss of female fertility, while the loss of miR-429 alone results in liver dysfunction[447,448]. The abnormal phenotypes caused by miR gene knock-outs is common in mammals, as 77% of the mammalian miRNA families are associated with a abnormal knock-out phenotype[361,449]. It should be noted that these drastic phenotypes are more commonly observed when knocking-out multiple members or an entire family of miRs[361,450]. This suggest that there is functional redundancy within miR families.

Nevertheless, the diverse effects of the loss of these microRNAs reflect their multi-targeting nature.

1.6.3.1 Hypoxia and HIF1 induced microRNA signatures

In addition to the regulation of protein coding genes, hypoxia and HIF1 have been shown to differentially regulate microRNAs. While studies have differing lists of microRNAs regulated by HIF1α, they all contain miR-210-3p, as it is a direct transcriptional target of both HIF1 and HIF2 [451,452]. By the nature of microRNAs, miR-210 targets several genes simultaneously. It has been regarded to help switch metabolism from oxidative phosphorylation to glycolysis through the targeting of iron-sulfur cluster assembly proteins ISCU1/2 mRNAs for repression[453]. It has also been shown to regulate resistance to apoptosis through targeting of BNIP3 and AIFM3, and promote cell proliferation through the repression of MNT, a MYC inhibitor[454–456]. The role of miR-210 appears to be both multifunctional and essential to hypoxic adaptation, highlighting the importance of microRNAs in defining and promoting a disease state[454].

While miR-210 may be the canonical hypoxia-induced miR, there are several other miRs controlled by hypoxia which influence tumour behaviour. Hypoxia has been shown to induce the expression of miR-27a, which in turn promotes chemoresistance in gastric tumour cells[457]. MiR-27a has also been shown to induce angiogenesis through the regulation of semaphorin 6A (SEMA6A)[458]. MiR-211 inhibits hypoxic adaptation by promoting oxidative phosphorylation energy production and represses HIF1α mRNA under hypoxic conditions[459]. MiR-211 is often downregulated in patient melanoma samples, suggesting an over activation of the hypoxia response[459,460]. The catalog of hypoxia regulated miRs is expanding, however the consensus is that differential regulation of hypoxia regulated miRs serves to promote tumour pathology[461].

1.6.4 The impact of miRNAs in disease and cancer

Given the extensive roles of microRNAs in development and tissue homeostasis, it is not surprising that microRNAs play a large role in disease states as well[361,449]. There is extensive evidence for the role of miRNAs as drivers of pathology in metabolic disorders[462]. In lipid homeostasis, a loss of expression of miR-122 in the liver leads to a significant reduction of circulating cholesterol[463–466]. This reduction in circulatory cholesterol levels was shown to be

caused by a downregulation of cholesterol biosynthesis genes and an increase in fatty acid oxidation, likely as a result of increased AMPK activity[463]. Similarly, miR-200a/b/c has been shown to regulate the accumulation of intracellular triglycerides in hepatocytes, through the targeting of JUN mRNAs which control the transcription of lipogenic genes such as sterol regulatory element-binding protein 1 (SREBP1)[467]. Reduction of miR-200a/b/c has also been shown to prevent interleuken-induced insulin resistance in hepatocytes[468]. Other microRNAs have been shown to regulate other shift in metabolism, such as miR-211 regulating the balance between glycolysis and oxidative phosphorylation through the targeting of pyruvate dehydrogenase kinase 4[459]. These miRs are drivers of these disease-states, as returning their expression level to their normal state can reverse the disease[459,469].

The largest body of work describing the links to miRNA dysregulation to a disease state is in cancer. The first link to miRNA dysregulation and cancer was made by Calin and colleagues, when they discovered genomic deletions for miR15 and miR16 in 65% of chronic lymphocytic leukemia, and that samples with intact miR15 and miR16 often expressed low level of miRs[470,471]. It was later found that miR-15 and miR-16 target and repress BCL2, which aides in resisting apoptosis[472]. Therefore downregulation of these miRs represented a new mechanism for cancer cells to resist apoptosis, through the downregulation of tumor suppressive miRNAs.

MiRNAs have also been shown to regulate specific cancer phenotypes, such as EMT. In addition to the previously mentioned lipid metabolic homeostasis functions of the miR-200 family, the family has also been heavily studied in its role in regulating EMT in cancer. The miR-200 family was initially implicated in EMT, when it was found that the whole family was downregulated in cells undergoing EMT[473]. This link was not correlational, as it was shown the miR-200 family negatively regulate the EMT promoting factors ZEB1/2 and SIP, and that overexpression of the family prevented cells from undergoing EMT[473,474]. Blocking of miR-200a/b and miR-205 expression was sufficient to induce EMT in Madin Darby Canine Kidney cells[473]. The same function of the miR-200 family is found in several cancer types as well, including lung and breast cancers[475–477].

In contrast to the functionality of the miR-200 family, there are EMT promoting miRs such as miR-221 and miR-222[478]. The miRs-221/222 are generally considered to be oncogenic miRs due to their overexpression in several cancers cell lines including breast cancer, prostate cancer, bladder cancer, colorectal cancer, gastric cancer, and melanoma[479–484]. Overexpression of miRs-221/222 increase the invasive capacity of prostate and gastric cancer cell lines, suggesting they play a role in driving EMT[480,483]. Recent evidence suggests that miRs-221/222 exert this function through negative regulation of Notch3[485]. As a final indication of the role of miRs-221/222 in promoting EMT, when breast cancer cells are forced to undergo EMT through growth in mammosphere conditions the cells down-regulate the miR-200 family and upregulate the miRs-221/222[486]. Beyond promoting EMT, miRs-221/222 have been implicated in promoting resistance to apoptosis through the targeting of pro-apoptotic gene BMF as well as others, resisting breast cancer chemotherapies[478,487]. MiR-221/222 have also been shown to directly target Dicer1, suggesting that miR-221/222 overexpression can result in a lower overall miRNA abundance[488].

1.6.5 The therapeutic potential of RNAi

As mentioned previously, differential regulation of microRNAs are key components of both the hypoxia adaptation response, EMT, and tumour pathology. Due to this, the RNAi pathway has been investigated as a potential therapeutic avenue. RNAi offers a unique therapeutic advantage as they allow for the downregulation of specific genes using an endogenous pathway. They also offer a way to target undruggable proteins, or proteins with no small molecule therapeutic developed, by targeting the mRNAs for these proteins rather than the proteins themselves. For these reasons, the RNAi pathway has been under investigation for therapeutic potential.

As an example, transthyretin-mediated amyloidosis is a rare life threatening hereditary disease caused by the accumulation of mutant transthyretin protein fibrils[489]. As a treatment for this, an RNAi therapeutic has been recently approved to treat this disease called patisiran. Patisiran is an FDA approved siRNA designed to target both wild-type and mutated transthyretin (TTR) mRNAs[489,490]. By reducing the amount of TTR protein expressed in these patients, patisiran

prevents the disease from progressing[491]. It is well tolerated, and only requires a single intravenous infusion every three weeks[491]. While the treatment does not reverse the disease, it does appear to stop or slow the progression of it. Due to how new this therapeutic is, the long-term side effects are not known. Although if left untreated, median survival is 4-7 years following diagnosis[492].

In a cancer setting, RNAi therapy is also being investigated. However, rather than introducing a synthetic siRNA to repress a target mRNA like in patisirian, miR replacement therapy is being considered. For example, it is known that miR-7 functions generally as a tumour suppressor miR[493,494]. Among other tumour suppressive roles, loss of miR-7 expression is associated with an increase in drug resistance through a loss of repression of MDR1 mRNAs[495]. For these reasons, miR-7 replacement therapy is under investigation for cancer therapy. Other miRs including let-7 and the members of the miR-34 family are also being considered for replacement therapy, and have shown efficacy in preclinical models[496].

As some miRs function as tumour suppressors in a cancer context, others are oncogenic. To reduce the efficacy of the latter, suppressive forms of RNA therapy has been developed, termed antagomirs and miR-sponges[466,497]. Antagomirs are small RNA molecules designed to hybridize to endogenous miRs. Transfection of antagomirs result in specific reduction of the functionality of the targeted miR[466]. A miR-sponge is an RNA containing multiple binding sites to the miR of interest such that when they are sufficiently expressed, they sequester the miR-RISC complexes away from their endogenous targets, derepressing them[498]. There are currently several different sponge sequence designs, ranging from perfect complementarity to the miR sequence, reproducing endogenous sites, or introducing a mismatch between miR nucleotides 9-12 and the target sequence[499–501]. Work is still being done to determine the optimal sponge sequence, which would allow for maximization of target miR binding while minimizing sponge decay. There are limitations to the sponge technique, as certain miRs can be expressed at such high levels that sponging becomes impractical[497]. While little work has been done using synthetic miR sponges *in vivo*, there is evidence of miR-sponging used to repress tumour suppressor miRs endogenously. Through the expression of CDR1αs, an endogenous circular RNA, endogenous miR-7 is sponged away from its target genes[502–505].

In addition to the development of different RNA-based therapeutics, development of RNA delivery technologies have also advanced. With the maturation of lipid-based RNA delivery, local and systemic delivery of RNAs into cells is now possible. For a more precise approach, pH lipid integrating peptides (pHLIPs) can be used to deliver single molecules of linear therapeutic RNA solely in acidic environments, such as the hypoxic microenvironment of a solid tumour[506,507]. Thanks to all of these advancements, the modulation of the transcriptome through RNAi is increasingly becoming a viable therapeutic approach for a multitude of diseases.

1.7 Hypothesis and objectives of PhD project

Hypothesis: Given the correlation between tumour hypoxia and poor prognosis, and the different phenotypes that HIF1 controls, we hypothesize that efficiently countering hypoxia adaptation would increase the efficacy of other anti-cancer therapies by targeting the most resistant fraction of the tumour.

As shown in the literature, the biggest obstacle to cancer therapeutics is the adaptive capacity of the tumour. In order to design better therapeutics to counter the hypoxia adaptation program, we must have a better understanding of the adaptive limits of cancer to hypoxia. Towards this end, we have developed novel *in vitro* hypoxic culture protocols which better recapitulate the *in vivo* hypoxic microenvironment.

As the phenotypes controlled by the hypoxia adaptation program are so diverse, to effectively target the whole of the transcriptional response, a multi-targeted approach is required. To address this, we plan to target these vulnerabilities through a new form of single-sequence, multi-targeting RNA interference (RNAi).

This thesis is divided into three objectives, with each objective covered in their respective chapters.

1. **Objective 1: Characterize the adaptation response to short-term hypoxia in conjunction with metabolic stress *in vitro***
    I. Development of a hypoxia-adaptation reporter cell line.

II. Development of restricted metabolic hypoxia (RHY) incubation protocol: a short-term *in vitro* hypoxic incubation protocol combining metabolic stress and hypoxia to better simulate the *in vivo* hypoxic microenvironment.

III. Analysis of RNAseq data of murine melanoma cells undergoing the RHY incubation protocol.

2. **Objective 2: Characterize the hypoxic adaptation response to long-term and severe hypoxia exposure *in vitro*.**

I. Development of a long-term hypoxic incubation protocol (LTHY): a two week incubation protocol which gets progressively hypoxic to simulate *in vivo* hypoxia development.

II. RNAseq analyses of murine melanoma cells undergoing the LTHY protocol.

III. Identification and characterization of a novel WT1 isoform induced by the LTHY protocol.

3. **Objective 3: Development of synthetic miRNAs for characterization of miR-tail sequence function & therapeutic potential**

I. Characterization of shRNA structures with progressively longer miR-tail binding offsets

II. Development & validation of single-sequence, multi-targeting RNAi molecule design algorithms.

Chapter 2 – Materials & Methods

2.1 Cell Culture

B16 and D4M cells were maintained in DMEM + GlutaMax supplemented with 10% FBS & 1% Penicillin-Streptomycin unless otherwise specified. HEK293T cells were maintained in DMEM + GlutaMax supplemented with 10% FBS & 1% Penicillin-Streptomycin. DU145 cells were maintain in DMEM + GlutaMax supplemented with 10% FBS & 1% Penicillin-Streptomycin. 4T1 cells were maintained in RPMI media + GlutaMax supplemented with 10% FBS & 1% Penicillin-Streptomycin. All cells were routinely checked for mycoplasma contamination and were confirmed to be mycoplasma negative by PCR.

Cells were passaged using FACS buffer-based detachment. Cells were incubated for 5 minutes at 37°C in FACS buffer (PBS; 0.5% FBS; 2mM EDTA; 10mM HEPES pH7.6), after which the cells become detached. Cells are collected in 15mL Falcon tubes and pelleted by centrifugation at 1500 RPM for 3 minutes. Cell pellets are then resuspended in complete media and seeded as appropriate for the context.

2.2 Plasmid construction methods

2.2.1 Oligo annealing & phosphorylation

For the insertion of small DNA sequences into plasmids, a protocol of oligo annealing and phosphorylation followed by restriction ligation was employed. Pairs of DNA oligos are designed to hybridize with each other, and have the appropriate 5’ and 3’ overhangs for the given restriction ligation. Dry oligos are resuspended in distilled water at 1mM concentrations. Oligo pairs are annealed at 100uM in annealing buffer (10mM Tris-HCl pH7.5; 100mM NaCl; 1mM EDTA) by incubating the oligos at 95°C for 10 minutes and allowing the reaction to cool at room temperature for 90 minutes.

After annealing, the hybridized oligos are phosphorylated. 10uM of annealed oligos are phosphorylated using T4 PNK (Invitrogen) as per manufacturers protocol. Once phosphorylated, the oligos are ready for ligation. For ligation into plasmids, 20nM of phosphorylated oligos are

used per ligation. For the generation of the 5xHRE, the entirety of phosphorylation reaction was used to ligate to itself.

2.2.2 Agarose Gel extraction

Size separated DNA fragments were extracted from agarose gel fragments using the Zymoclean Gel DNA Recovery Kit and the manufacturers protocol.

2.2.3 Gibson Assembly

Gibson Assembly master mixes were produced in-house. 5XISO buffer (25% PEG-8000; 500 mM Tris-HCl pH 7.5; 50 mM $MgCl_2$; 50 mM DTT; 1 mM each of the 4 dNTPs; 5 mM NAD) is made in 6mL batches and stored at -20°C. 80 reaction batches of 2x Gibson Assembly mixes are made using the following reagents: 320uL 5XISO buffer; 0.64uL T5 exonuclease 10U/uL (NEB #M0363S); 20uL Phusion 2U/uL (NEB #M0530S); 40uL Taq Ligase 40U/uL (NEB #M0208S); dH_2O 419.36uL. Gibson Assembly reactions are performed as per the original protocols[508].

2.2.4 Single stranded Gibson Assembly

Using the standard Gibson Assembly master mix, single stranded Gibson Assembly was used for the insertion of short DNA sequences in contexts where restriction ligation was not ideal. 100ng of digested plasmid is mixed with oligos designed for single stranded Gibson Assembly at 18.75nM. Using these quantities, the Gibson reaction was performed as previously described[508,509].

2.2.5 Restriction Ligation

Restriction ligation was performed using gel purified restriction digested DNA fragments. A ratio of 1:3 for plasmid:insert was typically used, unless the insert and plasmid were of similar size, in which case a ratio of 1:1 was used. Restriction ligation reactions were performed in 10uL volumes, using T4 ligase (Invitrogen) as per manufacturers specifications. Reactions were incubated for 1 hour at room-temperature. Subsequently, 5uL of the ligation reaction was used for bacterial transformation.

2.3 DNA construct descriptions

2.3.1 pHAGE-HIF1α-eGFP

Human HIF1α cDNA was generated by RT-PCR by Julie Pelloux at the Major Lab and inserted into the pMIN backbone by Gibson Assembly using the oligos in **Table 1**. Using Gibson assembly, the HIF1α stop codon was removed, and a GS linker 10 amino acids in length linked to GFP was added to the 3' end. This construct was then transferred to the pHAGE backbone through a combination of Gibson Assembly and restriction ligation.

Table 1: Oligos used for the construction of pHAGE-HIF1α-eGFP

EG: oligos are from Gagnon lab database. FM: Oligos are from Major lab database. Uppercase: Homology for Gibson Assembly. Lowercase: homology for PCR.

Oligo Ref #	Oligo name	Oligo sequence (5'->3')
FM 854	HIF1a F	TTCTCTAGGCGCCGGAATTAatggagggcgccggc
FM 855	HIF1a R	gttaacttgatccaaag
EG 512	Linker-gib F	TCAGAGCTTTGGATCAAGTTAACgggtccggcggctcc
EG 513	10aa_linker-eGFP	AACGGGTCCGGCGGCTCCGGCGGCTCCGGCTCCatggtgagc aagggcgag
FM 862	mCherry R	TAAAATCTTTTATTTTATCGATttacttgtacagctcgtcca

2.3.2 pHAGE-ODD-eGFP

Using the hypoxia reporter with a promoter swap to EF1α as a starting point, the ODD-eGFP section of the CDS was cloned back into the pHAGE-EF1α backbone by Gibson Assembly using the oligos from **Table 2**. The ODD corresponds to the amino acid range 400-652 for the Uniprot human HIF1α canonical sequence (Q16665-1). This amino acid range corresponds to a merge of the Uniprot ODD range and previously published ODD sequence[510].

Table 2: Oligos used for construction of pHAGE-ODD-eGFP construct

Uppercase nucleotides in oligo 718 denotes the section with homology to HIF1α

Oligo Ref #	Oligo name	Oligo sequence (5'->3')
718	HIF-ODD-gibF	gactgatcgtagcgtagccaccatgGCCCCAGCCGCTGGAG
719	linker-GFP-gibR2	CAGCTCCTCGCCCTTGCTCACGGATCCGCCGGAGCCGCCGGAGCCGCCGG
720	HIF-ODD-gibR1	CCGCCGGAGCCGCCGGACCCAGTAGTTTCTTTATGTATGTGGGTAGGAGATGG

2.3.3 pHAGE-HRE-deGFP

The HRE-deGFP construct was designed entirely in-house. The construction process began by preparing a empty version of the pHAGE backbone. This empty version does not contain any promoter, protein coding sequence, or WPRE and was generated by removing the promoter, ORF, and WPRE in one reaction by restriction digestion, then inserting a MCS containing all the sites needed for the rest of the construction.

The HRE motif used in this construct is derived from the HRE found in the murine *Vegfa* promoter (chr17:46033277-46033312 reverse complement strand). The sequence for this core HRE is TACGTGGG. Hybridizing oligo pairs containing this sequence and cohesive ends were hybridized and phosphorylated, then self-ligated using T4 ligase (NEB) for 1 hour at room temperature (see **Table 3** for oligo sequences). The n-mers were then separated by agrose gel migration (5% agarose; 100V; 55 minutes), and 5mers were purified from the gel and ligated into the empty pHAGE backbone. Successful clones were identified by restriction digest tests, and were sequence verified by Sanger sequencing.

Once the pHAGE-5xHRE was constructed, miniCMV-deGFP gBlocks (IDT) were inserted both upstream and downstream of the 5xHRE insert by Gibson Assembly. The following is the miniCMV-deGFP gBlock sequence:

AGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCAGATCTTTGTCGATCCTACCATCCACTCGACACACCCGCCAGCGGCCGCTGCCAAGCTTCCGAGCTCTCGAATTCAAAGGAGGTACCCACCATGGTGTCTAAGGGCGAAGAACTCTTCACTGGGGTGGTGCCAATTCTCGTTGAACTGGACGGCGACGTGAATGGTCACAAATTTTCTGTCTCTGGGGAAGGCGAAGGGGACGCAACCTATGGGA

AGTTGACCCTTAAATTTATCTGTACAACTGGAAAACTTCCCGTTCCTTGGCCCACTCTGGTGACTACTTTG
ACCTACGGCGTTCAGTGCTTTTCCCGGTATCCAGACCATATGAAGCAGCACGATTTCTTCAAAAGCGCAA
TGCCCGAAGGATATGTCCAGGAACGGACTATTTTTTTCAAAGATGATGGCAACTACAAAACCCGAGCCG
AGGTTAAGTTCGAAGGCGATACATTGGTGAACCGCATTGAGCTTAAAGGCATCGACTTCAAAGAGGACG
GTAATATACTTGGTCATAAACTGGAATATAACTACAATTCTCATAATGTCTATATCATGGCAGACAAGCA
GAAGAACGGAATCAAAGTCAATTTCAAGATCAGACATAATATCGAAGATGGGTCAGTTCAGCTGGCCGA
CCACTACCAGCAAAATACCCCTATTGGTGATGGGCCCGTTTTGCTTCCTGACAACCATTACTTGTCCACCC
AATCCGCTCTTTCTAAAGACCCTAACGAGAAGCGAGATCATATGGTACTGCTTGAATTCGTGACAGCAGC
CGGTATAACTCTGGGTATGGATGAACTGTATAAAAAGCTGTCTCATGGTTTCCCACCCGAGGTTGAAGAA
CAAGATGACGGGACCCTTCCTATGTCCTGTGCTCAAGAAAGCGGCATGGATCGGCATCCAGCCGCATGC
GCTAGTGCTCGAATAAACGTATAA

miniCMV // spacer sequence // deGFP

Full plasmid documentation cen be found in EG plasmid document P357.

Table 3: Oligos used for pHAGE-5xHRE-deGFP construct

Oligo Reference #	Oligo name	Oligo sequence (5->3)
730	pHAGE-HRE-deGFP mcs TOP	CTAGTTAAGCTCGAGCTTAGTCGACGAGCCGC
729	pHAGE-HRE-deGFP mcs BOTTOM	GGCTCGTCGACTAAGCTCGAGCTTAA
728	pHAGE-HRE-deGFP HRE TOP	TCGAGACACAGTGCATACGTGGGTTTCCACAGG TCGTCTCG
727	pHAGE-HRE-deGFP HRE BOTTOM	TCGACGAGACGACCTGTGGAAACCCACGTATGC ACTGTGTC

2.3.4 pLKO-mCherry

The pLKO-mCherry construct was already built in the Gagnon lab. The plasmid was built by swapping Ametrine for mCherry in the Gagnon lab pLKO-Ametrine construct by restriction ligation.

2.3.5 pCW-tWT1 constructs

The truncated WT1 cDNA was generated by PCR from B16-HG LTHY 0.1% O_2 n2 cDNA. Initial construction began with cloning a partial CDS from exon 7 to exon 10 using the following primers for homology to the cDNA: F-tgtcccacttacagatgcatagcc, R-aagcgccagctggagt. This partial CDS was cloning into the Doxycycline inducible pCW backbone with a C terminal SGSGS linker to GFP via Gibson assembly, and called tWT1.1. Subsequently, intron 5 to exon 7 was amplified using the following sequences for homology: F-CTGAGGGTGAATTTTGGGGC, R-cttaaaatatctcttattgcagcctggg. This reaction generated two products, representing intron5-exon6-exon7 and intron5-exon7. Both products were purified by gel extraction and inserted upstream of the tWT1.1 CDS by Gibson assembly. The KTS motif was later removed from this construct via Gibson assembly using the tWT1 primer and the following primers: gaaaagcccttcagctgtcg and cgacagctgaagggcttttcacctgtatgagtcctggtgtg.

2.3.6 sa/shRNA tail spacing constructs

Plasmids to test miR tail binding spacing were built into the Gagnon lab pLKO-AIO backbone (P314). For each shRNA construct, oligos designed to express the various shRNA constructs were cloned into P314 via oligo annealing and restriction ligation.

Table 4: Tested sa/shRNA sets

Table of all sa/shRNA sequences tested. Sequences are active strands of expressed hairpins, in the 5' -> 3' orientation. Lowercase sequences are Abox sequences, designed for minimal interaction with target sequence.

Site	Structure	Sequence. 5->3
Site 1	5nt bridge	GGGTCGTGtaaGATAACCGGG
	4nt bridge	GGGTCGTGtaaCGATAACCGG
	3nt bridge	GGGTCGTGtaaGCGATAACCG
	shRNA	GGGTCGTGattGCGATAACCG
Site 2	5nt bridge	CGCGCGAAaatTTACGGGGTC
	4nt bridge	CGCGCGAAaatATTACGGGGT
	3nt bridge	CGCGCGAAaaaAATTACGGGG

	shRNA	CGCGCGAAtttAATTACGGGG
Site 3	5nt bridge	GACTGGCTttaGTGAGTAGCG
	4nt bridge	GACTGGCTttaTGTGAGTAGC
	3nt bridge	GACTGGCTttaGTGTGAGTAG
	shRNA	GACTGGCTaacGTGTGAGTAG
Scramble	Non-binding	GTTGGTGCTCTTCATCTTGTT

2.3.7 pLKO-GFPnt2 construct

The GFP-NT2 construct was originally designed by Julie Robitaille from the Fancois Major lab. For use in this work, the GFP-NT2 mRNA sequence was transferred into pLKO-hPGK for steady state expression via Gibson assembly.

Table 5: Oligos used for pLKO-GFP-NTUTR construct

Oligo Ref #	Oligo name	Oligo sequence (5'->3')
827	GFPnt F	CCGACCTCTCTCCCCAGGGGCTAGCATGGTGAGCAAGGGCGAGGA
826	GFPnt R	GCTGCCTTGTAAGTCATTGGTCTTAAAGGTACCTCAGACCGAGCGTAA GTCATTGG

2.3.8 pLKO-APP-synmiRs

synmiRs and control shRNAs for the HIV multi-targeting project were cloned into pLKO-Ametrine-P2A-Puro via oligo annealing and restriction ligation.

Table 6: Oligos used in construction of HIV synmiR constructs

For top oligos, uppercase denotes the active strand of the shRNA

Oligo#	oligo name	Oligo Sequence 5→3
1336	sh1.1_22ntL7_top	ccggcagtgcaggggaaagaatagtatataactTACTATTCTTTCCCCTGCACTGttttttg
1337	sh1.1_22ntL7_btm	AATTCAAAAACAGTGCAGGGGAAAGAATAGTAAGTTATATACTATTCTTTCCCCTGCACTG
1338	sh1.2_22ntL7_top	ccggcttggtaggtttaagaatagtttataactAACTATTCTTAAACCTACCAAGttttttg

1339	sh1.2_22ntL7_btm	AATTCAAAAACTTGGTAGGTTTAAGAATAGTTAGTTATAAACTATTCTTAAACCTACCAAG
1340	Sh1.3_22ntL7_top	ccggagtcaggagctaaagaatagtgtataactCACTATTCTTTAGCTCCTGACTtttttg
1341	Sh1.3_22ntL7_btm	AATTCAAAAAAGTCAGGAGCTAAAGAATAGTGAGTTATACACTATTCTTTAGCTCCTGACT
1347	FM1.1_L7_top	ccggcagccaggtacgagaatagtttataactAACTATTCTCGTACCTGGCTGttttttg
1348	FM1.1_L7_btm	AATTCAAAAACAGCCAGGTACGAGAATAGTTAGTTATAAACTATTCTCGTACCTGGCTG
1349	FM1.2_L7_top	ccggaacattgcagtaagaatagtgtataactCACTATTCTTACTGCAATGTTttttttg
1350	FM1.2_L7_btm	AATTCAAAAAAACATTGCAGTAAGAATAGTGAGTTATACACTATTCTTACTGCAATGTT
1351	JQ1.1_L7_top	ccggatagcaggagaaagaatagtgtataactCACTATTCTTTCTCCTGCTATttttttg
1352	JQ1.1_L7_btm	AATTCAAAAAATAGCAGGAGAAAGAATAGTGAGTTATACACTATTCTTTCTCCTGCTAT

2.3.9 pNL-mCh

The original pNL-mCherry construct was developed by Yifie Yan at the Francois Major lab using the pNL-(R-)-(E-)-luciferase plasmid. The Luciferase CDS was swapped to mCherry via restriction ligation. For insertion of the local synmiR targeting sites, an MCS was inserted upstream of the mCherry start codon via single stranded Gibson Assembly. Once the MCS was inserted, the various local HIV sites were inserted using oligo annealing and restriction ligation.

Table 7: Oligos used in pNL-mCherry constructs

ssGib top: lowercase denotes inserted MCS. MCS structure is XbaI – 6nt spacer – ClaI. Left side uppercase denotes homology with ssGib bottom, right side uppercase denotes Gibson homology with mCherry. SsGib bottom: lowercase denotes homology section with ssGib top, uppercase denotes homology with local pNL sequence. For seed site oligos, uppercase denotes local HIV targeting sequence.

Oligo number	Oligo name	Oligo sequence 5→3
1288	mCherry gateway ssGib top	GGCCGCA tctaga ggaggt atcgat GTGAGCAAGGGCGAGGAGGATAAC
1289	mCherry gateway ssGib	atcgatacctcctctaga TGCGGCCGCGCCACCCATCTTATAGCAAAATCC

	bottom	
1293	seed1 site1 top	ctagcATTGGGGGGTACAGTGCAGGGGAAAGAATAGTAat
1294	seed1 site1 bottom	cgatTACTATTCTTTCCCCTGCACTGTACCCCCCAATg
1295	seed1 site2 top	ctagcATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTat
1296	seed1 site2 bottom	cgatAACTATTCTTAAACCTACCAAGCCTCCTACTATg
1297	seed1 site3 top	ctagcAAATATTGGAGTCAGGAGCTAAAGAATAGTGGAat
1298	seed1 site3 bottom	cgatTCCACTATTCTTTAGCTCCTGACTCCAATATTTg
1299	seed2 site1 top	ctagcATAGTAGAAGAGAAGGCTTTCAGCCCAGAAGGAat
1300	seed2 site1 bottom	cgatTCCTTCTGGGCTGAAAGCCTTCTCTTCTACTATg
1301	seed2 site2 top	ctagcCTGACAGAGGACAGATGGAACAAGCCCCAGAAGat
1302	seed2 site2 bottom	cgatCTTCTGGGGCTTGTTCCATCTGTCCTCTGTCAGg
1303	seed2 site3 top	ctagcTCTTTAAGCAATCCTCAGGAGGGGACCCAGAAAat
1304	seed2 site3 bottom	cgatTTTCTGGGTCCCCTCCTGAGGATTGCTTAAAGAg
1305	NT top	ctagcTCTATCCGGTCCACTTATCTGACTGCGTCCTCTat
1306	NT bottom	cgatAGAGGACGCAGTCAGATAAGTGGACCGGATAGAg

2.4 Lentivirus production

Lentiviral particles were generated through transfection of HEK293T cells with lentiviral packaging plasmids. HEK293T cells were maintained at 60-80% confluency, and split to maintain this confluency percentage the next day. The next day, HEK293T cells are transfected with the following mixes:

Table 8: Transfection reaction mixes for 3rd generation lentivirus production

Material	6 well format	10cm dish format
pMD2.G (aka VSVg)	625ng	3ug
pMDLg (aka pRRE)	312.5ng	1.5ug
pRSV-Rev	312.5ng	1.5ug
pHAGE-HIF1α-eGFP	1.25ug	6ug
Mirus TransIT-LT1	7.5uL	36uL
RPMI media	200uL	1mL

Table 9: Transfection reaction mixes for 2nd generation lentivirus production

Plasmid	Quantity for 10cm	Quantity for 6well	Quantity for 24well
Payload	6ug	1.25ug	500ng
psPAX2	7.2ug	1.5ug	600ng
pMD2G	840ng	175ng	70ng
PBS	1mL	200uL	50uL

Transfections for hypoxia reporters, tWT1 constructs, and miR-tail testing constructs were performed using 3rd generation packaging plasmids, and Mirus-LT1 (Mirus Bio) as the delivery agent. Transfections were performed as per manufacturers' protocol.

16 hours post transfection, the HEK media was changed, and the cells are left to incubate for 48hrs. After this, virus containing media is collected and centrifuged for 5 minutes at 2000 RPM. Viral supernatant is then used for transduction, or aliquoted and snap-frozen on dry-ice and stored at -80°C for later use.

2.5 Lentivirus Transduction

Cells for transduction are seeded 24 hours prior to transduction. At transduction, the cell media is changed to a mixture of cell media and viral supernatant (1:3 virus:media v:v). 8ug/mL of Polybrene is added and the cells are allowed to transduce for 24 hours, after which the media is changed. Cells are examined three days post transduction for evidence of transduction by fluorescence or selection.

2.6 Lentivirus Ultracentrifugation

For lentiviral supernatant concentrated by ultracentrifugation, lentiviral supernatant is generated in 10cm dishes using 3rd generation packaging plasmids. Media is changed 16 hours post transfection. 24 hours post media change, the media is changed and collected. The next day, the media is pooled with the previous collection and used for ultracentrifugation.

The lentiviral supernatant is centrifuged at 2000 RPM for 5 minutes, followed by filtration with a 0.45um Surfactant-Free Cellulose Acetate (SFCA) filter. The filtered media is

then ultracentrifuged at 25,000 RPM for 90 minutes. The original media is then decanted, and the viral pellet is resuspended in 500uL of HEK media through overnight incubation at 4°C. After resuspension, cells are transduced using the entire resuspension in a 1.5mL final volume and 8ug/mL of Polybrene.

2.7 FACS analyses

FACS analyses were done on ZE5 (Bio-Rad), CantoII (BD Biosciences), or an LSRII (BD Biosciecnes). All markers (eGFP, mCherry, Ametrine) were produced endogenously, and did not require antibody labelling. Intracellular Doxycycline levels were quantified using the 405nm laser and 525/50 detection filters[511]. FACS data analyses and figure generation was done using FlowJo V10 (BD Life Sciences).

2.8 Cobalt Chloride treatment

Initially, the stock cobalt chloride was as a solution, at 100mM (Sigma; 15862-1ML-F). Later, a stock solution of 1M was made using Cobalt(II) Chloride (Sigma; 232696-5G). The stock solutions is then filter sterilized using a PES 0.2um syringe filter (Fisher scientific basix) under a tissue culture hood, aliquoted in 500uL aliquots, and stored at -20°C. $CoCl_2$ media supplementation varies for each cell line, and was optimized for each cell line[253]. For HEK293T and D4M cells, media was supplemented with 100uM $CoCl_2$. For B16 cells, media was supplemented with 200uM $CoCl_2$. Cells are treated with $CoCl_2$ for 24 hours. For 4T1 cells, media was supplemented with 250uM $CoCl_2$.

2.9 FACS Cell sorting

Cells were sorted at the IRIC Flow Cytometry platform using the BD FACSAria III sorter. Cells sorted by bulk were sorted into fresh cell media. After sorting, cells were plated into vessels appropriate for the number of cell sorted. For single cell sorting, each positive cell was sorted directly into a well of a 96 well flat bottom adherent plate containing 150uL of conditioned media. Conditioned media made of 45% fresh cell media, 45% media used to grow the same cell line for 24 hours, and 10% FBS.

2.10 Hypoxic Cell culture

Hypoxic incubation was performed in a BioSpherix Xvivo system model X2 closed hypoxic incubation system. O_2 and CO_2 sensors were calibrated prior to experiments as per manufacturers protocol. Relative humidity of the hypoxic incubation chamber was maintained at <= 70%, as per manufacturers instructions. During experiments, O_2 and CO_2 were dynamically controlled and maintained at set levels during the length of the timecourse. 50mL aliquots of cell media (DMEM + GlutaMax, 10% FBS, 1% Penicillin/Streptomycin, 25mM HEPES pH7.6) and FACS buffer (PBS, 2mM EDTA, 0.5% FBS, 10mM HEPES pH 7.6) were kept in the hypoxia chamber with their lids loosened to allow for gas exchange and equilibration to the hypoxic atmosphere for at least 24 hours prior to use. Cells were passed at the end of specific timepoints using hypoxic FACS buffer. Detached cells are transferred into 15mL Falcon tubes and removed from the hypoxic incubation system for centrifugation. Once centrifuged, the Falcons are brought back into the hypoxic incubation system and only then opened. This ensures the cells are not exposed to normoxia during passaging.

2.10.1 MHY protocol

1.25M B16-HG cells were seeded equally into 10cm dishes containing either their standard media (DMEM + glutaMax with 10% FBS &1% P/S) or RPMI media + glutaMax without FBS or P/S. One dish containing cells in standard media was left in the tissue culture incubator, while another and a dish containing cells in RPMI media were moved the hypoxic incubation chamber. The hypoxic incubation chamber was set to 5% CO_2, 1% O_2, and maintained a relative humidity level <=70%.

All cells were left to grow for 48hrs, after which the cells were passaged using FACS buffer, pelleted, and lysed in QaiZol (Qaigen), flash-frozen on dry ice and stored at -80°C. Biological replicates were done serially, and all QaiZol samples were processed at the same time once all biological replicates were completed.

2.10.2 LTHY timecourse protocol

750K B16-HG cells were seeded in 25cm^2 plug seal cap flasks (VWR, cat: 82051-070) at the beginning of the timecourse. The timecourse begins at 5% O_2 for 24hrs, after which flasks are passed, 20% of the cells are re-seeded into new flasks, and the remaining cells are flash-frozen in Qaizol for RNA extraction. After passaging, O_2 is lowered to 1% for 48hrs, after which cells are passed again as previously described. Then O_2 is lowered to 0.5% for 72hrs, and cells are passed, using 30% of the cells for re-seeding. After the 0.5% O_2 timepoint, every 72hrs cells are passed and O_2 is lowered by 0.1%. Once 72hrs have passed at 0.1% O_2, all cells are collected for RNA extraction.

2.11 RNA extraction for RNAseq

Total RNA & miRNAs were extracted using the Qaigen miRNeasy Mini kit, as per manufacturers instructions. RNA quality was validated by the IRIC Genomics Platform and had a RNA Integrity Number (RIN) > 9.5.

2.12 Long RNA Library preparation & run

For the MHY dataset, long RNA library preparation was performed by the IRIC genomics platform. Stranded library preparation was done using the KAPA Biosystems rRNA depletion kit. NGS was performed using 100bp paired end reads on a Illumina HiSeq500.

For the LTHY dataset, long RNA stranded library preparation was performed by the McGill University and Genome Quebec Innovation Center using the KAPA rRNA-depleted (HMR) stranded library preparation for Illumina sequencing. Samples were ran on a Illumina HiSeq4000.

2.13 Small RNA Library preparation & run

For both the MHY and LTHY datasets, small RNA library preparation was performed by the McGill University and Genome Quebec Innovation Center. Library preparation was done using the NEB miRNA library preparation protocol. miRNAseq was performed using single end 50bp reads on a Illumina HiSeq.

2.14 rRNA-depleted mRNAseq data processing

Raw RNAseq reads used quality controlled and trimmed using FASTQC on default settings[512]. Trimmed reads were mapped to the murine genome (UCSC mm10) using Tophat v2.1.1 and bowtie2 v2.2.9. Only reads with a single match to mm10 were kept for further analysis.

Reads are separated into positive and negative stranded reads. Using the mouse genome reference annotation file (Mus_musculus.GRCm38.94.gtf), reads are counted on exons using coverageBed v2.24.0. Exons counts belonging to the same gene are summed using a script written by Zohra Saci at the Francois Major lab. Positive and negative strand exon counts are then merged. Differential gene expression was calculated using SARTools.

2.15 smallRNAseq data processing

For smallRNAseq, raw RNAseq reads are minimally trimmed to maximize genomic mapping. Genomic mapping was done using bowtie1 v1.2 and miRDeep2 v2.0.0.8. Aside from the following options, all settings were set to default: reads shorter than 17nt are discarded, reads can map to up to ten places in the genome, at most 1 mismatch is permitted per read. After genomic mapping, miRs are counted from the surviving reads.

2.16 Quantitative Polymerase Chain-Reaction

Cell pellets were generated by passaging cells using FACS buffer. No less that 100,000 cells were used for RNA extraction. RNA was extracted from cell pellets using Trizol as per manufacturers protocol (Qaigen). RNA pellets were resuspended in Rnase-free dH_2O in a volume sufficient to ensure RNA concentrations above 200ng/uL (20-100uL).

RNA was converted to cDNA using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) and 2ug of RNA. Quantitative PCR (qPCR) was performed using Advanced qPCR MasterMix HI-ROX (Wisent Bioproducts) and 0.75% of the cDNA reaction per qPCR reaction. qPCR reactions were run in 10uL technical triplicates, in either 96 well or 384 well formats. 96 well format plates were run on a StepOnePlus™ Real-Time PCR system (Applied Biosystems), while 384 well plates were run on a QuantStudio™ 7 Flex Real-Time PCR System

(Applied Biosystems). Our qPCR run has the following steps: 1) a denaturation phase of 2 minutes incubation at 50°C, followed by a 3 minute incubation at 95°C. 2) 40 cycles of amplification and signal acquisition starting with 10 seconds at 95°C followed by 30 seconds at 63°C. 3) Melt-curve acquisition was always done after amplification.

qPCR amplification data was analyzed and exported using QuantStudio Real-Time PCR software. Relative gene expression was quantified relative to the used house-keeper gene. Housekeeper genes used were 18s, HPRT, and RPLP0. Relative expression was done calculated using the using the $2^{-\Delta\Delta CT}$ methodology. Error propagation[513] is performed at ΔCT and ΔΔCT calculation steps using the following formulas:

$dCT\_target\_cnd1\_SD = sqrt(CT\_target\_cnd1\_sd^2 + CT\_HK\_cnd1\_sd^2)$

$ddCT\_target\_cnd2\_sd = sqrt(dCT\_target\_cnd1\_sd^2 + dCT\_target\_cnd2\_sd^2)$

Abbreviations: dCT: delta Ct value; ddCT: delta of ΔCT values; cnd: condition; HK: house-keeper; sd: standard deviation.

See **Table 10** for qPCR oligo sequences used.

Table 10: qPCR probes used

Oligo reference number	Oligo name	Oligo sequence (5'->3')
027	18s F	AGGAATTCCCAGTAAGTGCG
028	18s R	GCCTCACTAAACCATCCAA
NA	RPLP0 F	TCCCACTTGCTGAAAAGGTC
NA	RPLP0 R	AGCAGGAGCAGCTGTGGT
514	HPRT F	TCCTCCTCAGACCGCTTTT
515	HPRT R	CCTGGTTCATCATCGCTAATC
1047	Mmu WT1 F	ATCCCAGGCAGGAAAGTGTG
1046	Mmu WT1 R	GTGCTGTCTTGGAAGTCGGA
1097	Hsa WT1 F1 5' probe	GCGATCCTGGACTTCCTCTTG
1098	Hsa WT1 R1 3' probe	CCATTTGCTGCGGCTCAGAC
1099	Hsa WT1 F2 5' probe	AGAAACGCCCCTTCATGTGT

1100	Hsa WT1 R2 3’ probe	CGGGAGAACTTTCGCTGACA
NA	Hsa WT1 F3 5’ probe	AGCTCAAAAGACACCAAAGGAG
NA	Hsa WT1 R3 3’ probe	GGGAGAACTTTCGCTGACAA

2.17 MG132 treatment protocol

HEK cells were treated with 10uM MG132 (Sigma-Aldrich) for four hours. After incubation, cells were brought to the microscope for imaging without changing the media.

2.18 Puromycin Selection

Puromycin stock solution was made in-house from Puromycin-dihydrochloride (Wisent Bioproducts). Puromycin-dihydrochloride is resuspended to a concentration of 1mg/mL in dH_2O. The stock solutions is then filter sterilized using a PES 0.2um syringe filter (Fisher scientific basix) under a tissue culture hood, aliquoted in 500uL aliquots, and stored at -20°C.

Cells are transduced with lentivirus in a 24 well format as previously described. One day post transduction, they are passaged into one well of a 6 well plate. A well of untransduced cells of the same cell line is also seeded in one well of a 6 well format.

Two days post passage, media is replaced with media supplemented with 1ug/mL of Puromycin. Media is changed every other day with Puromycin supplemented media for six days, or until 100% of the untransduced control cells has died. If possible, selection efficacy is validated by FACS.

2.19 Doxycycline induction protocol

Doxycycline (DOX) stock solution was prepared by resuspending Doxycycline Hyclate at a concentration of 10mg/mL. This stock solution was filtered using a PES 0.2um syringe filter (Fisher scientific basix) under a tissue culture hood, aliquoted in 500uL aliquots, and stored at -20°C until used. Media is supplemented at 2ug/mL DOX.

2.20 Microscopy

All fluorescent microscopy images were taken using a LSM-880 (Zeiss). GFP was acquired using an Argon-488nm, and mCherry was acquired using Argon-561nm. Images were processed using ImageJ (US National Institutes of Health). Cell tracing was done using a high brightness/contrast version of the image, and manual tracing in Inkscape. The cell morphology photos (**Figure 24**) were taken using a white light Nikon Eclipse TS100 microscope, the 10x magnification objective lens, and a Nexus 5 smartphone.

2.21 RNAseq analyses

2.21.1 MHY dataset analyses

Gene counts were normalized for sequencing depth and RNA composition using DESeq2's median of ratios method, where raw gene counts are divided by sample-specific size factors to account for sample sequencing depth and RNA composition[514]. Normalized readcounts are then rounded to the nearest integer value. These normalized readcounts (or their average between replicates) were used for all DEG/DEmiR analyses and figures. DESeq2 gene expression tables were saved as static text files and were filtered using Python with no additional libraries.

For GO term analyses, gene expression tables were filtered based on a absolute log2 Fold Change (FC) >= 2 in the comparison the statistics was run on, having at least 100 DESeq2 normalized readcount in any condition, and have a Benjamini-Hochberg adjusted pvalue (padj) less than 0.05 for the comparison the statistics was run on. Fold changes were calculated as the ratio of average normalized DESeq2 readcount between the two conditions under comparison. Padj values were calculated using DESeq2. Filtered gene expression tables were saved to new files, as well as files containing the list of genes that passed filtering.

2.21.2 MHY Venn Diagrams

Venn diagrams were generated using both Python and R. DEG list overlap was calculated in Python, and Venn diagrams were rendered using R and the VennDiagram library.

2.21.3 DEG heatmaps (MHY/LTHY)

Tables of differentially expressed genes were generated as previously described. FC, padj, and minimal read counts were calculated as previously described in Python, with values specified in each heatmap legend section. Using these filtered tables as input, heatmaps were generated in R using the pheatmap library.

When specified, we employed a custom gene expression normalization metric which we call the contribution metric. For a given gene, average normalized gene counts (as calculated by DESeq2) were taken and converted to a fraction of total expression of the gene across all conditions. The contribution formula is as follows: cell_value/row_sum. Once this calculation is complete, any value less than 0.1 is raised to 0.1 to reduce colour bar spread.

For the MHY rRNA-depleted mRNAseq dataset, gene expression was normalized using the contribution metric. Once normalized, the gene expression table is rendered as a heatmap using euclidean distance and the WardD.2 hierarchical linkage metric. The separation of the clustered rows into seven clusters was done using the pheatmap cutree_rows argument. The choice to separate the genes into seven clusters was subjective. Both the rendered heatmap, and the normalized gene expressions with cluster number were saved to files for further analyses.

For the LTHY rRNA-depleted mRNAseq heatmap, gene expression was normalized using the row z-score metric. Gene expression clustering was accomplished using kmeans non-hierarchical cluster. The number of clusters was chosen to be seven based on the elbow method and the within-group sum of squared distances. Heatmap gaps indicate separate kmean clusters. The normalized data used to generated the heatmap, as well as each gene's cluster number, was saved to a separate file for further analyses.

For all rRNA-depleted mRNAseq heatmaps, files listing gene names for each cluster, as well as all possible combinations of clusters, were generated with Python.

For the DEmiR heatmaps, tables of differentially expressed miRs were generated in Python. MiRs were filtered by a padj < 0.05 and minimal expression >= 100 mean DESeq2

normalized reads in any condition comparison using Python. Heatmaps were generated using these filtered tables in R using the pheatmap package. Gene expression was normalized using either the contribution metric for the MHY dataset, or row z-score for the LTHY dataset. The cluster separation method was done with the pheatmap argument cutree_row, with the number of clusters chosen subjectively.

2.21.4 Gene/miR expression histograms

Gene/miR expression histograms were generated using normalized replicate expression values from DESeq2. Histograms were rendered using GraphPad Prism7. Statistical analysis was performed in DESeq2, and added to graphs using Inkscape.

2.21.5 Metabolic network figure generation

Metabolic network figure generation begins with an in-house Python script which generates a Cytoscape compatible network file[515]. This script takes the DESeq2 expression data, and file containing all gene:gene interactions in the desired network, and generates a file containing gene:gene interactions and expression levels for each gene. Expression values could only be one value per condition, so average DESeq2 normalized reads were used. Once imported into Cytoscape, the network is manually arranged and expanded upon as needed. The network is then exported as a PDF, and polished in Inkscape.

2.21.6 GO term analyses for heatmaps

Gene lists for each heatmap cluster were saved as part of the heatmap generation cluster. Using these files as input, GO term enrichment was done using DAVID's online tools[516,517]. No custom settings were used, and GO term enrichment tables were saved to the local disk using curl. GO terms with a p-value < 0.05 were added to the clusters in Inkscape. Complete GO term enrichment tables can be found in supplemental file.

2.21.7 GO term bubble plot rendering

Lists of upregulated DEGs (padj < 0.05, FC >= 2) for HYvsTC or RHYvsTC are passed to DAVID for GO term enrichment[516,517]. GO terms with an FDR < 0.05 are run through REViGO to reduce and summarize the GO terms[518]. These REViGO simplified lists of enriched GO terms are

then visualized as bubbleplots using an in-house Python script which uses numpy, pandas, and matplotlib. Bubbleplots were generated as scatterplots, with each points x-axis position, size, and colour mapped to the relevant data. Y-axis position was dynamically generated to minimize text overlap.

2.21.8 IGV exploration

RNAseq read exploration was performed by loading RNAseq BAM files into the Integrative Genomic Viewer (IGV) v2.9.4 [519,520].

2.21.9 Generation of the WT1 read-coverage plot

The WT1 RNAseq read location plot was generated using the sashimi-ploy.py Python script from the ggsashimi project[521]. Minimum read coverage was set to 3 to remove primary transcript associated reads. Chromosomal range was set to chromosome 2, 105162045-105174815. GRCm38.p6 was used for gene annotation.

2.21.10 kmer analyses of WT1 & Leucegene

Tables of Kmers for all samples in the Leucegene dataset were precomputed by the IRIC Bioinformatics platform using the Jellyfish software[522]. Files containing the isoform specific mRNA junction sequences were used to quantify WT1 isoform level expression within the Leucegene dataset using the *km* software[523]. These junctions correspond to the exon1-exon2 junction for canonical WT1 expression, exon1a-2 for AWT1 expression, and the junction between the first two exons for isoform G expression. Using bash, a script was written to evaluate the expression levels for these three junctions across all Leucegene samples. Then using Python, the analyses were compiled into a single table per analysis. A sample was considered to express an isoform if expression for the isoform specific junction was detected by *km*.

2.21.11 Survival curve generation

Sets of samples expressing various WT1 isoforms were sent to Dr. Hébert and survival curves were generated by Dr. Béliveau using R.

2.21.12 Kozak strength evaluation

Kozak reliability was predicted using the ATGpr algorithm provided by AA Salamov et al (https://atgpr.dbcls.jp/)[524]. Exon6-skip mRNA sequence from beginning to the end of the WT1 CDS was used as input. Kozak reliability scores were rendered as a histogram using GraphPad v7.02.

2.21.13 GSEA analyses

GSEA was run locally for all analyses GSEA for Linux v4.2.2. For the MHY GSEA plots, DESeq2 gene expression tables were re-formatted to be compatible with GSEA. For each DESeq2 comparison, the gene expression table was filtered to genes with a FC >= 1.2 or FC <= 1/1.2 and an padj value < 0.05. Normalized DESeq2 expression values and the OGS for genes passing these filters were saved to a new file in a GSEA compatible format.

GSEA was run using the log2 ratio of classes metric, the weighted scoring scheme, the gene set permutation mode, and 1000 gene set permutations. The chip file used was the GSEA mouse gene symbol to human ortholog file v7.5.1.

Comparative GSEA plots were generated by overlaying GSEA plots in Inkscape. The Y scale was stretched to match both original plots.

2.22 Cell lysis protocols for Western Blots and IP

2.22.1 Minimal Lysis Buffer

Resuspend PBS washed cell pellet in 1mL/10^7 cells of minimal lysis buffer. Add protease inhibitors immediately preceding resuspension. Incubate at 4°C for 1 hour with rotation. Centrifuge for 1 minute at 5000 RPM. Recover supernatant and proceed with protein precipitation protocol.

Minimal Lysis buffer

750uL 2M NaCl	(150mM final)
200uL 1M HEPES	(20mM final)
100uL NP-40	(1% final)
100uL 0.5M EDTA	(5mM final)

ddH_2O up to 10mL

RIPA lysis buffer

Resuspend PBS-washed cell pellet RIPA buffer with freshly added protease inhibitors (Pierce, A32955) at a concentration of 1ml RIPA/10^7 cells. Allow to incubate at 4°C for 30 minutes on a rotator. Centrifuge cell lysate at 10000xg for 15 minutes at 4°C. Recover supernatant and proceed with protein precipitation or IP protocol.

RIPA Buffer

0.60g Tris base
0.88g NaCl
1 ml NP40
0.5g Sodium deoxycholate
0.1g SDS
100mL ddH_2O (final volume)
Adjust to pH 7.6
Store at 4 °C

2.22.2 Wessel-Fluegge protein precipitation for Western Blots

Using 150uL of cell lysate in an aqueous solution, add 600uL of methanol. Then add 15uL of chloroform. Vortex until only one phase is visible. Add 450uL of distilled water and vortex thoroughly. Centrifuge at full speed for 2 minutes. Remove upper aqueous phase without disturbing the interphase. Add 450uL of methanol, vortex thoroughly and centrifuge at full speed for 5 minutes. Remove supernatant, and let air-dry for 10 minutes. Resuspend in 1xLDS (ThermoFisher NP0008) with 100uM DTT. Heat to 95°C for 5 minutes.

2.23 Western Blots

Once proteins were precipitated and resuspended in 1xLDS +10mM DTT, protein samples are loaded into a NuPage 4-12% Bis-Tris gel (Life Technologies) in MES buffer. The gel was run for 1 hour at 120V, or until sufficient migration. Once migrated, proteins were transferred to a methanol-soaked Polyvinylidene Fluoride (PVDF) membrane using a wet transfer box set to 200mA for 30 minutes and transfer buffer. After protein transfer, the whole

PVDF membrane is blocked using TBS + 0.1% Tween-20 (TBST, Bioshop), supplemented with 3% bovine serum albumin (BSA, BioShop) for 1 hour at room-temperature while rocking. After blocking, the whole membrane cut for specific antibody probing. Primary antibody solutions are prepared in TBST+3%BSA supplemented with 0.002% (g/ml) of sodium azide, and primary antibody. PVDF membrane sections are incubated in primary antibody solution at 4°C overnight.

The following day, primary antibody solution is recovered, and membrane sections are washed three times. Primary washes are in 3mL of TBST+3%BSA, rocking at room-temperature for 5 minutes. After primary washes, membrane secondary antibody solutions are applied to the membranes. Secondary antibody solutions are TBST+3%BSA with the appropriate HRP conjugated antibody diluted 1:10,000. Membranes incubate in secondary antibody solution for 1 hour at room-temperature while rocking. After incubation, the secondary antibody solution is removed and the membranes are washed three times in TBST. After the secondary antibody washes, antibody binding is revealed using ECL substrate (Bio-Rad) and chemiluminescence is detected using a ChemiDoc MP (Bio-Rad). Images were processed using ImageJ (US National Institutes of Health).

Table 11: Primary antibodies used for Western Blots

Antibody	Catalog number	Dilution used
goat polyclonal anti-GFP	aum173	1:1000
rabbit monoclonal anti-Calnexin	ADI-SPA-860	1:3000
Mouse monoclonal anti-HIF1α	610958	1:1000

2.24 GFP-IP

Following RIPA cell lysis protocol, immunoprecipitation is performed. Perform all steps on ice. A small sample of lysate is removed and used as the IP input control. The remaining lysate is incubated overnight with 2.5ug of rabbit anti-GFP antibody (Table 12, antibody 1) on a rotor. The next day, 50uL/IP of Protein A agarose beads (Millipore Sigma-Aldrich 16-157) are washed and resuspended in PBS. After washing the beads are added to the antibody + lysate

mixture. The beads are incubated at 4°C for 1 hour on a rotor. Afterwards, centrifuge tubes at 2500xg for 1 minute at 4°C. Remove the supernatant and aliquot 10% of the volume for a post-IP control. Wash the beads in 500uL of ice-cold PBS three times. After the final wash, resuspend the beads in 2xLDS 100mM DTT. The incubate the beads at 95°C for 10 minutes. Proceed with SDS-PAGE.

2.25 Mass spectrometry

HEK cells stably expressing pCW-exon6-skip-tWT1-KTS+ were incubated with 2ug/mL Dox for 48 hours prior to lysis. Cells were lysed using the RIPA method as previously described, and GFP was used to IP the tWT-GFP fusion protein. The IP was validated by running 5% of the IP sample as a Western blot using the goat polyclonal GFP antibody. After validation of the IP, the rest of the IP sample was run on a NuPage 4-12% Bis-Tris gel (Life Technologies) in MES buffer. After sufficient migration, the portion of gel corresponding to the size of the tWT-GFP fusion protein was excised and sent to the IRIC Proteomics platform to be processed for trypsin based Mass spectrometry.

The theoretical tWT1 protein sequence was used to search for peptide coverage in the Mass spectrometry data. These were the search settings used: Search engine: PEAKS Studio v10.5 // fragment tolerance: 10.0 PPM // Fixed modifications: +57 on C (carbamidomethyl) // Variable modifications: +1 on NQ (Deamidated), +16 on M (Oxidation), +42 on n (Acetyl), +80 on STY (Phospho) // Digestion enzyme: Trypsin. Coverage of the tWT1-GFP CDS was calculated using Scaffold v4.8.3.

2.26 ChIP

Cells are grown in 10cm tissue culture dishes for adherent cells. Media in the dish is reduced to 6mL prior to transporting the dish on ice to orbital shaker under a chemical hood. 167uL of 37% formaldehyde solution is added drop-wise to the dish. The dish is left to incubate to 10 minutes at room-temperature while shaking to allow for DNA-protein cross-linking. The cross-linking is stopped by adding 450uL of freshly prepared 2M glycine, and allowing the dish to incubate for another 5 minutes at room-temperature. The media disposed of as liquid PFA

waste. The dish is washed twice with 10mL of cold PBS (2 minutes shaking per wash). Wash liquid is also disposed of in PFA liquid waste.

2.5mL of cold PBS is added back to the dish. Cells are detached using cell scrappers. Cells are collected in 15mL falcon tubes. Dishes are washed twice with 2.5mL of cold PBS, and liquid is added to the falcon tube. Cells are pelleted by centrifugation at 4°C, 1000g, for 5 minutes.

Cell membrane of the pellet is lysed using ChIP Cell Lysis Buffer. The cell pellet is resuspended in 500uL of ChIP Cell Lysis Buffer and incubated on ice for 10 minutes. Nuclei are pelleted by centrifugation at 4°C, 1000g for 5 minutes. The supernatant is discarded, and the nuclei are lysed in 200uL of ChIP Nuclear Lysis Buffer for 10 minutes on ice.

Nuclear lysate is sonicated using a Bioruptor (Diagenode). Sonication settings were 10 minutes total sonication, 30 seconds ON, 30 seconds OFF, medium intensity. After sonication, DNA fragment size is verified prior to the IP. Formaldehyde cross-links are reversed by heating 20uL of sonicate mixed with 20uL of 10% Chelex-100 beads (Bio-Rad, cat. no. 142-1253) to 90°C for 5 minutes. DNA concentration is quantified in duplicates using a NanoPhotometer. 2ug of DNA is run on a 1% agarose gel to determine range of DNA fragment size. For ChIP-qPCR, maximum DNA fragment size should be 300-600bp.

The remaining sonicate is incubated at 65°C for 5 minutes. 50ug of DNA is diluted 1:20 with dilution buffer. 10% of the total volume is taken for ChIP input control. The volume taken is diluted 1:3 with 100% EtOH and stored at -80°C overnight. 2.5ug of antibody is added to the remaining diluted sonicate. IPs incubate at 4°C overnight with rotation.

The next morning, 50uL of Protein A Dynabeads (Invitrogen: 10001D) are washed twice in cold ChIP IP buffer. The IP sample is centrifuged at 12000g for 10 minutes at 4°C, and the supernatant is retained. The washed beads are added to the supernatant and are incubated at 4°C for 2 hours with rotation. After incubation, the beads are separated using a magnetic stand (Invitrogen: 12321D). The supernatant is collected to be used as post-IP control when indicated. The beads are washed twice with 1mL of ChIP Dialysis Buffer. After dialysis, the beads are washed four times with ChIP Wash Buffer.

The input control is removed from -80°C after the overnight incubation. DNA is precipitated by centrifugation at 12000g for 10 minutes. The DNA pellet is washed twice with 70% EtOH.

The elution steps are performed on the IPs and input controls in parallel. IP beads are resuspended in 150uL of ChIP Elution Buffer, IP input control samples are resuspended in 300uL. All samples are incubated at 65°C and 1000rpm using a Eppendorf Thermomixer (model 5355). IP input control samples incubate for 30 minutes. IP samples incubate for 15 minutes, after which the supernatant is collected and the beads are resuspended in another 150uL of ChIP elution buffer and allowed to incubate for another 15 minutes. After this second incubation, the volumes are combined.

DNA binding protein is removed from the samples by incubating with 2.4uL of Proteinase K (stock concentration 20mg/mL; ThermoFisher EO0491) for 1 hour at 55°C with 1000rpm of rotation on the thermomixer. After, DNA is purified using a DNA gel recovery kit (Zymo Research: D4001) into 50uL of ChIP elution buffer. DNA is then quantified using a nanophotometer, and stored at -20°C.

Table 12: Antibodies used for ChIP

Antibody; species; clonality	Manufacturer	Cat. Number
Anti-GFP; Rabbit; Polyclonal	Invitrogen	#A6455
Anti-HIF1α; Rabbit; Polyclonal	Novus Bio	nb100-134

2.26.1 ChIP Cell Lysis Buffer

100mL (store at 4˚C)

1mL	HEPES 500mM pH7.5	(= 5mM HEPES final)
4.250mL	KCl 2M	(= 85mM KCl final)
2.5mL	NP40 20%	(= 0.5% Np40 final)
	H2O milliQ	

Before lysis, add to buffer:

Aprotinin	1µL/mL
Leupeptin	1µL/mL

Pepstatin	2µL/mL	
PMSF	10µL/mL	
DTT	1µL/mL	

2.26.2 ChIP Nuclear Lysis Buffer

100mL (store at room temp.)

5mL	Tris 1M pH8.1	(= 50mM Tris final)
2mL	EDTA 500mM pH8	(= 10mM EDTA final)
10mL	SDS 10%	(=1% SDS final)
	H2O milliQ	

Before lysis, add to buffer:

Aprotinin	1µL/mL
Leupeptin	1µL/mL
Pepstatin	2µL/mL
PMSF	10µL/mL
DTT	1µL/mL

2.26.3 ChIP Dilution buffer

500mL (store at 4˚C)

500 µL SDS 10%		(=0.01% final)
27.5 mL	Triton 100X 20%	(= 1.1% final)
1200 µL	EDTA 500mM pH8	(=1.2mM final)
8,35mL	Tris 1M pH8.1	(=16.7mM final)
16.70mL	NaCl 5M	(=167mM final)
	H2O milliQ	

2.26.4 ChIP IP buffer

10mL nuclear lysis buffer
90mL dilution buffer

2.26.5 ChIP Dialysis buffer

200mL (store at 4˚C)

800µL	EDTA 500mM pH8	(=2mM final)
9,99mL	Tris 1M pH 8.1	(= 50mM final)
2mL	Sarkosyl 20%	(= 0.2% final)
	H2O milliQ	

2.26.6 ChIP Wash buffer

500ml (store at 4˚C)

50 ml	LiCl 5M	(=0.5M final)
25mL	NP40 20%	(=1% final)
50mL	Na-deoxycholate 10%	(= 1% final = 5g for 500mL)
16.6ml	Tris 1M pH 8.1	(=33.2mM final)
	H2O milliQ	

2.27 ChIP-qPCR

Following ChIP protocol as previously described, ChIP input and elute samples are used for the ChIP-qPCR run. The qPCR mixes, acquisition machine, and run settings were the same as for a regular cDNA qPCR run. Quantification was done using either the percent input or fold-enrichment methods[525].

Table 13: ChIP-qPCR probe set list

Kmd3a, Bmp7, Fyn, and Fyn_neg probes were previously validated by Kann and colleagues[526]*. The Vegf probe set was previously validated as well*[527]*. All other probe sets were designed in-house using Pimer-BLAST*[528].

Probe name	F Primer sequence (5'->3')	R Primer sequence (5'->3')
Kmd3a	GCAGCTCCATTCTTCCATTT	GCTCATGATCCTGGGTCTC
Bmp7	TACAAGTCCGGAGAGCGAGT	AGATCGGAAAGGGGTTTGTT
Fyn	TGCCGGGTGGAAAATAGGAAT	GCAGCACGGGGCATCAATTA
Fyn_neg	ACAGGGGACTTAAAGGTGAGAGA	AGGCAAGCCATGCTTTGTTCTA
Car9	TCTACAAAAGGGCACTGTGAGT	CTGTGGACGGGCTGTACG
Vegfa	CGAGGGTTGGCGGCAGGAC	CAGTGGCGGGGAGTGAGACG
WT1 int5 HREs 3.5	CGCACTCCTTTACACGTCCT	TGATGGCCTGTCAATATGTGT
WT1 int5 3.5F2	TGACTGTTCGGTTTTAAGGTCAGA	TGGTACCAGTTTCCTTTTACCATCT
WT1 int5 7361	CATACCAGACACCAGGCACA	AATTAAACACGTGCGCCACC
WT1 int5 7995	GTGTTCACGTGTGACAGCTCA	ATCGAGTCCTAGGGACTGAAGTTG
WT1 int5 12.5k	TTTGAGTTCGTGGCTCTGCT	GGTTTCCAGAATGCAGTGGC
WT1 int5 16kb	GAGAACCACACTGGGAAGCA	CAGTTGGCCCCTGCCTATTT
WT1 HREs 18kb	TGGATGAGAGGAGCCACAGA	AGAATGCTGCAGGCTTCGAG

2.28 ChIPseq

ChIPseq sample preparation was performed using the ChIP-qPCR sample preparation protocol with the following adjustment. Nuclear lysate Bioruptor sonication settings are increased to 1x12 minutes 30 seconds ON, 30seconds OFF, Medium intensity to increase DNA fragmentation to 200-600bp. ChIP sample validation was performed using a ChIP-qPCR run of *Kmd3a*. After validation, ChIP samples were submitted to the IRIC Genomics platform for library preparation using the KAPA library preparation kit and NGS. Libraries were ran on an Illumina NextSeq 500. Raw NGS data was analyzed by the IRIC Bioinformatics platform. Raw reads trimmed using Trimmomatic, mapped to the murine genome (mm10) using BWA, and aligned reads were filtered using SAMtools (MAPQ > 20 & samflag 4)[529–531]. Filtered read analysis was performed using MACS for peak calls and HOMER for functional annotation, known motif analyses, and de novo motif analysis[532,533].

Mapping of the WT1 binding site to called peaks was done manually in Python using an in-house script. Source code can be found in the supplemental material. Generation of gene coverage figures was done using the Spark python tool, using the GENCODE annotation file, the auto-scale and smoothen functions (-gs, -sm 10), and polished in Inkscape[534]. Generation of the annotation cluster bubbleplots were generated performing functional annotation enrichment using DAVID and the ChIPseq peak list, and rendering that functional annotation data using a similar method to the GO term bubble plot script, and polished in Inkscape[516,517].

2.29 PEI transfection

Transfections of pNL and pLKO-APP-synmiR constructs was performed using PEI. Wells of a 24 well plate were equally seeded with WT HEK293T cells. In the evening of the following day, cells were transfected with 750ng of pLKO-APP-synmiR construct. The following morning, the media is changed and cells were transfected with 750ng of pNL-mCh construct. The following morning, the media is changed again cells are left to incubate for two days. Following this incubation, mCherry and Ametrine levels are quantified via FACS using the ZE5 (Bio-Rad).

PEI transfection protocol:

In a 1.5mL eppendorf add 200uL of room-temperature PBS. Add 750ng of plasmid DNA. Add 30uL of PEI (1:4 DNA:PEI ratio). Flick tube to mix, let stand for 20 minutes at room-temperature. After incubation, add mixture to cells drop-wise.

2.30 Knock-down calculations by FACS fluorescence

For saRNA knock-down efficacy calculations, FACS samples were processed to remove apoptotic cells and doublets using size exclusion methods. Once processed, the geometric mean of GFP fluorescence was calculated using the remaining cells. The knock-down was calculated as a ratio of DOX+/DOX- for each construct. This ratio was then normalized against the same ratio for the Scr construct.

For synmiR knock-down calculations, HIV transcriptional activity was measured through mCherry MFI. After processing FACS samples for apoptotic cells and doublets, the geometric mean of mCherry was calculated for the following sets of cells: doubly transfected cells (Ame+/mCh+, i.e. synmiR:HIV); mCherry only expressing cells (Ame-/mCh+, i.e. HIV expressing only). A knock-down value is calculated using a ratio between doubly transfected cells and mCherry only transfected cells for all samples. This is then normalized to a average knock-down ratio calculated from Scr samples.

2.31 Coding for synthetic miRNA work

2.31.1 saRNA seed site evaluation script

See section 5.3.1.2 for a comprehensive description of the seed site evaluation script. Source code for this script can be found in the supplemental material section.

2.31.2 synmiR design algorithms

See sections 5.3.2.1 and 5.3.2.2 for comprehensive descriptions of the generation of the minimal-mutation and ration-design synmiR design algorithms. Source code for the ration-design algorithm can be found in the supplemental material section.

2.31.3 pNL-NT design script

The purpose of the NT sequence was to be as non-binding to the synmiRs and control shRNAs as possible, while still maintaining a valid open reading frame, as the NT sequence was to be inserted into a protein coding sequence. To design this, a Python script was written which decomposed the synmiR and control shRNAs into the set of all nucleotide triplets found within the sequences. This is done using a sliding window and storing all triplets into a list, then removing duplicates by casting the list to a set and then back to a list. This list of triplets is then used to remove codons from the set of mammalian codons. The remaining codons are then used to generate an 33 nucleotide (11 amino acid) long segment. This length was chosen to match the insertion lengths for the endogenous site insertions. The exact amino acid composition was decided manually by Jordan Quenneville and approved by Dr. Gagnon.

Chapter 3 – Hypoxia in conjunction with metabolic stress leads to distinct HIF1α levels and transcriptomic adaptations

Jordan Quenneville[1,2], Julie Pelloux[1,2], Angélique Bellemare-Pelletier[1], Zohra Saci[1,3], François Major[1,3], Étienne Gagnon[1,4]

1. Institute for Research in Immunology and Cancer
2. Department of Molecular Biology, University of Montreal
3. Department of Computer Science and Operations Research, Faculty of Arts and Sciences, Université de Montréal
4. Department of Microbiology, Infectology, and Immunology, Faculty of Medicine, Université de Montréal

3.1 Authors & contributions

Jordan Quenneville: Conception, planning, execution, and analysis of experiments. Generation of hypoxia reporter plasmids. Maintenance and generation of cell lines. $CoCl_2$ treatment and hypoxia incubation optimization. FACS experiment acquisition and analysis. Performed Western blot. Performed all bioinformatic analyses downstream of DESeq2 table generation (heatmaps, GSEA, Venn diagrams, etc…).

Julie Pelloux: Planned original retroviral hypoxia reporter construct. Generated first plasmid containing human HIF1α CDS.

Angélique Bellemare-Pelletier: Conception, planning, and troubleshooting of hypoxia reporter constructs. Performed microscopy experiment.

Zohra Saci: Processed raw RNAseq data to DESeq2 tables of differentially expressed genes.

François Major: Conception of project, analysis and interpretation of results, figure preparation.

Étienne Gagnon: Conception of project, analysis and interpretation of results, final figure preparation.

3.2 Executive Summary

Given the varied regulatory mechanisms governing HIF1, with their myriad of exceptions, establishing an accurate HIF1 activation reporter construct is essential to investigating hypoxic adaptation. Because of these exceptions, the first objective of this project was to establish a hypoxia adaptation reporter cell line. We built several hypoxia-adaptation reporter constructs, which measured HIF1α protein stability or HIF1 transcriptional activity. After validating several different hypoxia-reporter constructs, a B16 hypoxia-reporter cell line was established using a lentiviral HIF1α-eGFP fusion protein construct.

Once the B16-HIF1α-eGFP hypoxia adaptation reporter cell line was constructed, it was used to explore how different hypoxic environments lead to different adaptation programs. After determining that metabolic stress in conjunction with hypoxia lead to increased HIF1α stabilization levels, we investigated how different or similar the adaptation programs were through an in-depth analysis of transcriptomic data from each condition. We found that metabolic stress in conjunction with hypoxic incubation activated a vastly different transcriptional program, resulting in a hypoxia adaptation program with additional emphasis on cellular motility, angiogenesis, and tissues remodelling. Metabolic pathways were also differentially regulated with condition specific glycolytic enzyme specificity, as well as differential regulation in a host of other pathways.

In addition to differentially expressed genes, microRNAs were also significantly differentially regulated across hypoxic incubation conditions. While certain canonical hypoxia induced miRs like miR-210 were equally upregulated across hypoxic conditions, other miRs were specifically regulated under different hypoxia conditions. Metabolic stress in conjunction with hypoxia induced differential expression in a high number of miRs, and more differential expressions associated with negative patient outcomes *in vivo*. For these reasons, we conclude that integrating metabolic stress in conjunction with hypoxic incubation generates a hypoxia adaptation program more representative of *in vivo* hypoxic adaptation.

3.3 Results & Discussion

3.3.1 Development and characterization of hypoxia adaptation reporter constructs

3.3.1.1 Retroviral hypoxia reporter construct

In order to accurately follow HIF1 activation, we began by developing a hypoxia-adaptation reporter construct. This was originally attempted in a retroviral backbone, where psi-gag would drive constitutive expression of a single polycistronic mRNA. This mRNA expresses two separate proteins: human HIF1α linked to eGFP at the C terminus through a 10 amino acid long Glycine-Serine linker, and mCherry. The HIF1α-eGFP fusion protein occurs first in the transcript, and mCherry expression is driven by an IRES after the HIF1α-eGFP CDS. (**Figure 7A**). This cassette allows for the simultaneous expression of a fluorescent marker of HIF1A stability through the HIF1α-eGFP fusion protein, as well a marker for retroviral mRNA production through the constitutive expression of mCherry.

Under standard tissue culture conditions, we would expect cells expressing this cassette to be GFP-/mCherry+. Relative to untransduced cells, transduced cells may have a slightly higher basal GFP signal due to the protein turnover of HIF1α-eGFP. Under any condition resulting in HIF1α stabilization such as hypoxia, dysregulation of the PHD proteins or VHL, or proteosomal blockage, cells would become GFP+/mCherry+. To validate if the cassette functioned as expected, the construct was transfected into HEK293T cells. Under standard tissue culture conditions, HEK cells were mCherry+ but GFP negative. When HIF1α was allowed to accumulate under normoxic conditions by preventing its degradation through MG132 mediated proteasomal blockage, mCherry+ cells became eGFP+ exclusively in the nucleus (**Figure 7B**). This suggested that the HIF1α-eGFP fusion protein was being translated properly, and degraded under normoxic conditions as expected. Additionally, this result demonstrated that the C terminal eGFP fusion protein did not grossly interfere with HIF1α function, as GFP signal was found exclusively in the nucleus.

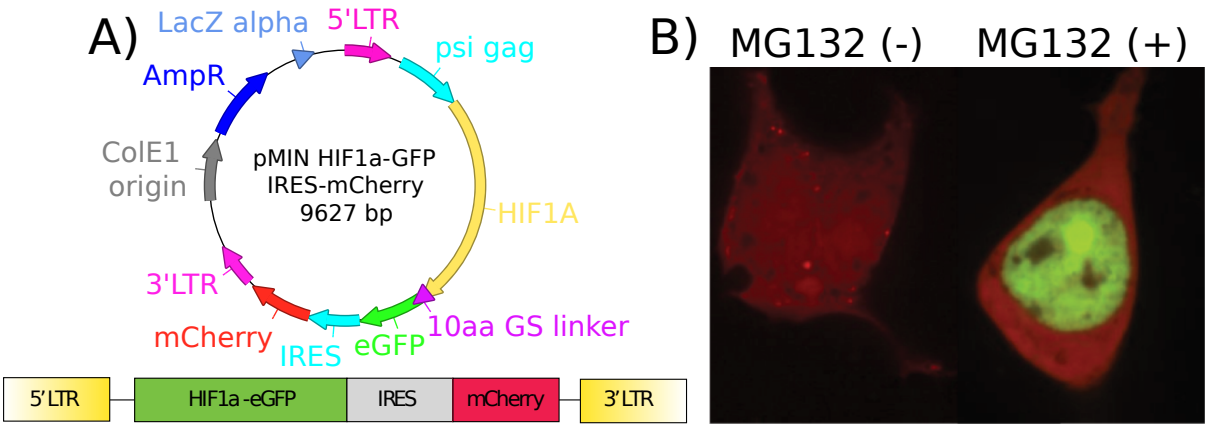

Figure 7: Retroviral hypoxia reporter construct & validation

A) Top: Retroviral hypoxia reporter construct plasmid map. Bottom: Retroviral hypoxia reporter retroviral payload. **B**) Left: HEK cells transfected with the retroviral hypoxia reporter, 40hrs post transfection. Right: HEK cells transfected with the retroviral hypoxia reporter treated with MG132 for four hours.

3.3.1.2 Generation of lentiviral hypoxia adaptation reporter constructs

Due to issues with generating infective retroviral particles capable of transducing our tumour cell lines, this polycistronic hypoxia reporter mRNA payload was restructured and split into two lentiviral constructs: the hypoxia-reporter section and the steady-state mCherry section (**Figure 8H**). With these lentiviral constructs cells are expected to constitutively be mCherry+, while only becoming GFP+ during hypoxic adaptation. Given the initial success of the retroviral reporter construct, the HIF1α-eGFP fusion protein was cloned into the 2nd generation lentiviral backbone pHAGE, which will be referred to as pHAGE-HIF1α-eGFP (**Figure 8A**). This construct allows for the most accurate understanding of HIF1α stability in the cell line because we are expressing the entire protein.

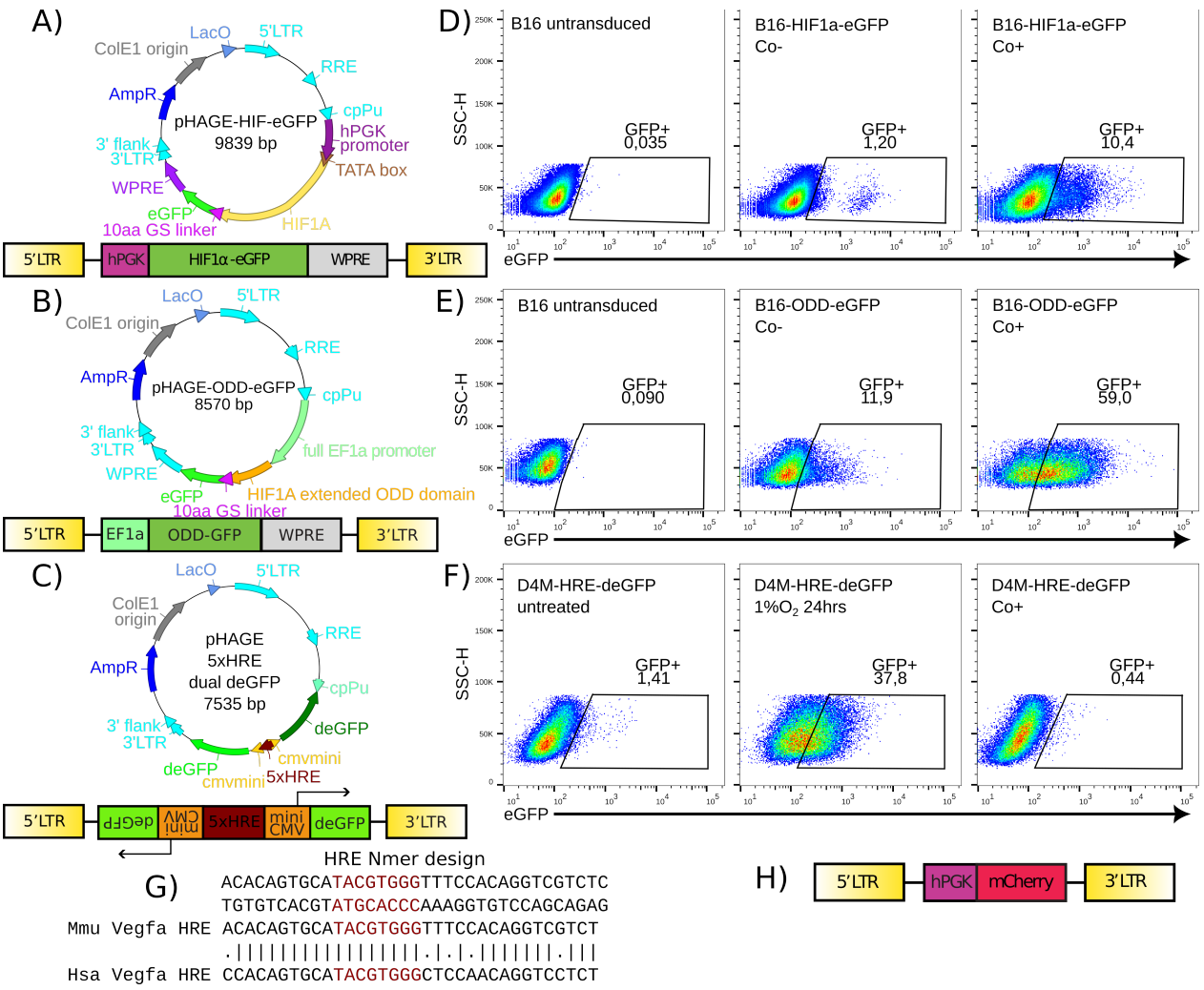

Figure 8: Lentiviral Hypoxia Reporter Constructs & Validation

A-C) Construct descriptions for pHAGE-HIF1a-eGFP (**A**), pHAGE-ODD-eGFP (**B**), and pHAGE-HRE-deGFP (**C**). Top: Plasmid map. Bottom: Lentiviral payload. **D&E**) Functional Validation for pHAGE-HIF1a-eGFP (**D**) and pHAGE-ODD-eGFP (**E**). Left panels: untransduced B16 cells. Middle panels: Unsorted B16-pHAGE-HIF1a-eGFP (**D**) or B16-pHAGE-ODD-eGFP (**E**). Right panels: Unsorted B16-pHAGE-HIF1a-eGFP (**D**) or B16-pHAGE-ODD-eGFP (**E**) cells 24hrs after addition of 200uM $CoCl_2$ to cell media. **F**) Functional validation of pHAGE-HRE-deGFP construct. Left panel: Unsorted D4M-pHAGE-HRE-deGFP. Middle panel: Unsorted D4M-pHAGE-HRE-deGFP after hypoxic incubation at 1% O_2 for 24hrs. Right panel: Unsorted D4M-pHAGE-HRE-deGFP cells 24hrs post addition of 200uM $CoCl_2$ to cell media. **G**) Description of HRE n-mer design and genomic conservation between mouse (Mmu) and human (Hsa). **H**) Lentiviral payload for steady-state mCherry expression.

In order to potentially boost GFP signal dynamic range and further reduce the lentiviral mRNA length, a shorter version of the HIF1α-eGFP fusion protein was built. In this fusion protein based reporter only the oxygen-dependant degradation (ODD) domain of HIF1α is expressed, fused through the same 10a.a. long Glycine-Serine linker as the original HIF1α-eGFP construct (**Figure 8B**). It is expected that cells transduced with this reporter will become GFP+ under conditions which prevent ODD-dependant degradation pathways, such as the canonical HIF1α oxygen dependant degradation pathway. This reporter would also be insensitive to regulatory mechanisms acting beyond this domain, such as the p53 or FIH-1 HIF1α regulatory pathways[535].

The previous lentiviral constructs track HIF1α stabilization. However, to track HIF1 transcriptional activity requires a different approach to the reporter. To capture HIF1 transcriptional activity, a fluorescent reporter construct based on the Hypoxia Responsive Element (HRE) DNA sequence was developed. This reporter's core principle was to use a well established HRE to drive the expression of a destabilized eGFP protein (deGFP) to provide an accurate readout of HIF1 transcriptional activity (**Figure 8C**). Due to the similarities between the HIF1 and HIF2 core HRE motif, any HRE-based reporter can be induced by either HIF[113]. Therefore, this reporter signals based on HIF activity, not specifically HIF1 or HIF2. We chose to use the human/mouse conserved HRE found in the promoter of *Vegfa*, a canonical hypoxia induced gene[536] (**Figure 8G**). Once an array of five HRE sequences was inserted into an empty version of the pHAGE backbone, a miniCMV promoter[537] and the deGFP CDS was inserted both downstream of the HREs and upstream on the reverse complement strand. This reporter becomes GFP+ whenever HIF1 or HIF2 is able to drive gene transcription, regardless of cellular oxygen status.

3.3.1.5 Establishing hypoxia adaptation reporter cell lines

Once these hypoxia adaptation reporter constructs were generated, they were validated in a stable expression context. Stable cell lines were tested using *in vitro* hypoxic incubation or through chemical stabilization of HIF1α via $CoCl_2$ treatment. Given the loss of a steady-state marker of transduction, transduction efficacy could only be determined through induction via *in vitro* hypoxic incubation or hypoxia mimetics. Lentiviral production was significantly reduced when generating pHAGE-HIF1α-eGFP particles, and required ultracentrifugation to generate

sufficient titers for transduction and cell sorting (**Figure 8D, right panel**). Upon transducing B16 cells with pHAGE-HIF1α-eGFP or pHAGE-ODD-eGFP constructs, there were several cells which were GFP+ at steady state (**Figure 8D&E, middle panels**). These cells represent cancer clones which naturally stabilize HIF1α through some sort of genetic mechanism. After incubation with $CoCl_2$, there was a clear increase in the percentage of GFP+ cells revealing the true transduction efficacy for these constructs (**Figure 8D&E right panels**).

Interestingly, D4M murine melanoma cells transduced with pHAGE-HRE-deGFP were only dynamic when subjected to hypoxic incubation at 1% O_2. $CoCl_2$ treatment did not produce detectable eGFP signal by FACS (**Figure 8F**). This suggests that while $CoCl_2$ treatment does stabilize HIF1α protein, it does not allow for transcriptional activation of our *VEGFA* HRE sequence. These results oppose previous reports demonstrating that $CoCl_2$ treatment does induce a HIF1 transcriptional response, including VEGFA, in as little as 4 hours[538,539]. This maybe due to HIF1 acting as an enhancer through the binding of more distal HREs in the *VEGFA* promoter, and activating *VEGFA* expression indirectly[540]. Therefore, we could say that our construct is a reporter for direct HIF1 transcriptional activity and not HIF1 enhancer functions.

Once we had evidence of transduction as measured through increased eGFP signal following $CoCl_2$ treatment, eGFP+ cells were sorted using single-cell sorting to generate clonally pure cell lines. This allowed us to maximize signal dynamics, completely remove constitutively active signalling, and remove untransduced cells. Once sufficiently grown, each single-cell derived cell line was treated with $CoCl_2$ to validate the clone's dynamic range.

While single-cell sorts of these constructs did generate clones with dynamic GFP signal after *in vitro* hypoxic incubation or hypoxia mimetics, the most dynamic clones arose from B16s transduced with the pHAGE-HIF1α-eGFP (**Figure 9**). Therefore, we chose to focus on the pHAGE-HIF1a-eGFP derived clones. Several pHAGE-HIF1a-eGFP clones (D2, E11, G12, G5, C3, E6) showed dynamics in eGFP signal after $CoCl_2$ treatment ranging from a 2 to 9 fold increase in eGFP signal (**Figure 9B**). Some these clones also appeared to have partially lost expression of the HIF1α-eGFP reporter as proportions of the population would not respond to $CoCl_2$ treatment (**Figure 9B; clones G12, G5, C3, E6**). There are two possible explanations for these non-dynamic

cells within a clonal population. The first being they are the result of improper single cell sorting, where a transduced and a non-transduced cell are sorted into the same well. The other possibility is the cell sorting was performed properly, but the construct became epigenetically silenced in a subset of cells as the clone multiplies from a single cell to a population. Investigating these different possibilities would require sorting of the non-dynamic cells are probing for the integration of the lentiviral construct through genomic PCRs of lentiviral-specific sequences. As the purpose of this sorting was to acquire a optimal hypoxia-reporter cell line, the nature of these non-dynamic cells was not investigated.

Additionally, even without $CoCl_2$ treatment, there appeared to be a minimal GFP signal in comparison to B16 WT cells (**Figure 9A**). Given the canonical HIF1α protein regulatory mechanism, in which there is a constant basal level of inactive protein being generated and degraded[63,64], it was expected that HIF1α-eGFP clones have a minimally different GFP signal in comparison to WT cells. Additionally, there are several in-frame alternative start codons past the ODD hydroxylation sites, as well as the GFP start codon. Translation initation from any of these start codons would result in non oxygen sensitive protein. An anti-GFP Western blot examining the entirety of the protein size spectrum would determine what mechanism is at play.

Other sorted clones (H10, F11, H12, F6) were constitutively eGFP+ (**Figure 9C**). $CoCl_2$ treatment on these clones resulted in minimal increases in eGFP signal, making them useless for generating a hypoxia reporter cell line. They would be useful in studying the properties of constitutive HIF1α stabilization in a cancer setting. Understanding how these clones maintain constitutive HIF1α stabilization and how this property affects their ability to form or exacerbate disease was decided to be beyond the scope of this project however.

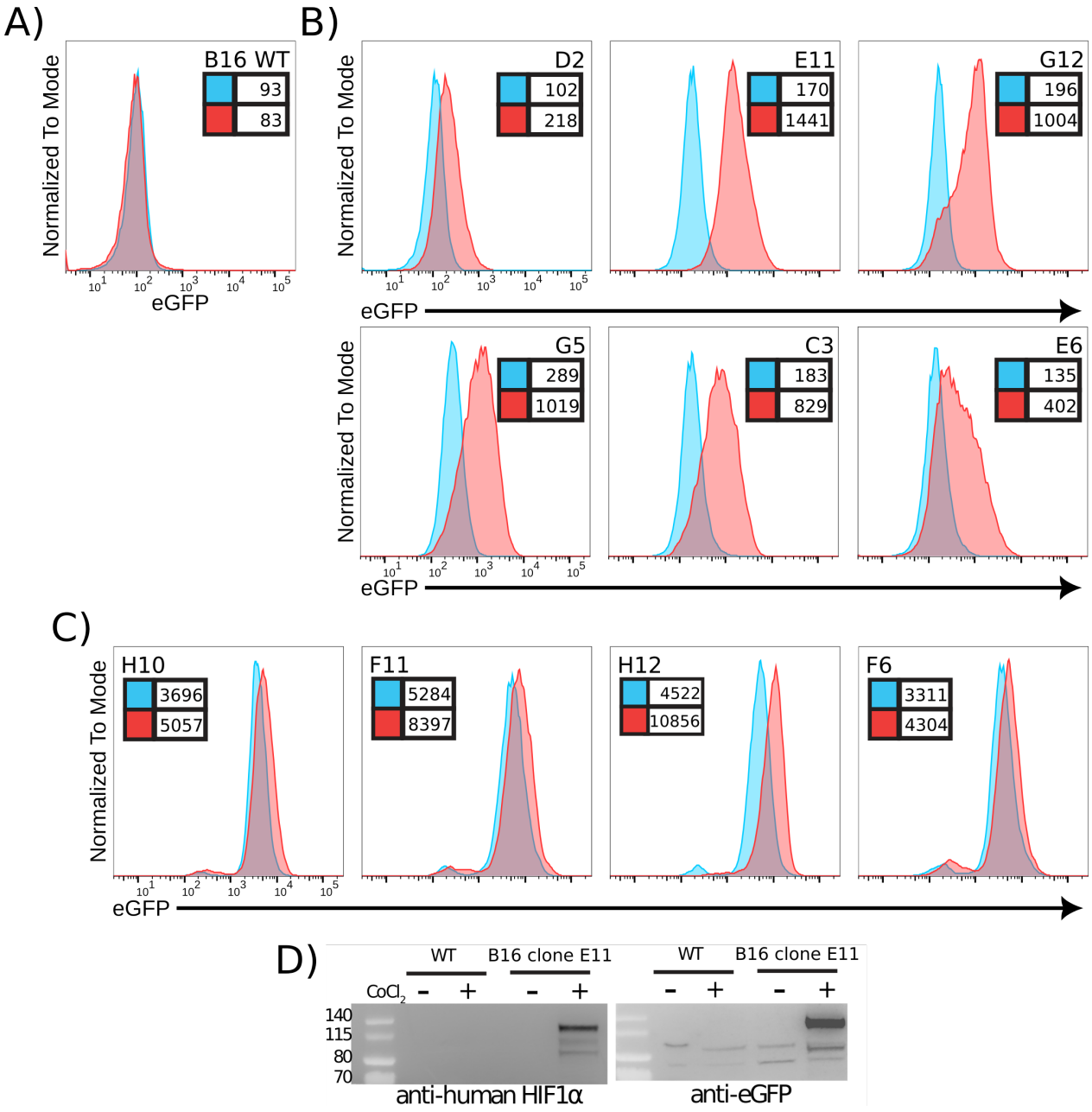

Figure 9: B16-pHAGE-HIF1a-eGFP clone validation

A) B16-wild-type cells undergoing $CoCl_2$ treatment. **B**) Examples of GFP dynamic clones. **C**) Examples of constitutively GFP+ B16-pHAGE-HIF1a-eGFP clones. **D**) Western blot confirming expression of HIF1a-eGFP fusion protein in clone E11. All $CoCl_2$ treatments were 200uM for 24hrs. Values inside legend boxes denote geometric Mean Fluorescent Intensity (geoMFI). Blue: $CoCl_2$ untreated; Red: $CoCl_2$ treated.

For the purpose of generating the hypoxia-reporter cell line, we chose the most dynamic clone (E11) as the basis. To verify that the GFP signal observed was the result of expression of the HIF1α-eGFP fusion protein, a Western Blot was performed on WT B16s and B16-pHAGE-HIF1α-eGFP clone E11 under normal tissue culture conditions or after $CoCl_2$ treatment (**Figure 9D**). The HIF1α-eGFP fusion protein was probed for using both a human specific monoclonal anti-HIF1α antibody and a polyclonal anti-GFP antibody. The predicted size of the HIF1α-eGFP fusion protein is 120kDa. Given the major bands for both antibodies are approximately at this height, it can be assumed that the increase in GFP signal by FACS is in fact due to expression of the fusion protein. To complete construction of the hypoxia-reporter cell line, B16-pHAGE-eGFP clone E11 was transduced with pLKO-mCherry, to act as a steady state marker of our cell line. After transduction with pLKO-mCherry, cells were sorted for purity of mCherry signal, this new cell line was named B16-HG (Hypoxia GFP).

3.3.1.6 Optimization of *in vitro* hypoxic incubation & metabolic stress using the B16-HG cell line

Having established the B16-HG cell line, we could begin experiments to optimize *in vitro* hypoxic incubation. These optimization experiments would allow us to determine which hypoxic incubation conditions would lead to maximal HIF1 engagement, as measured by HIF1α-eGFP levels. To verify that this reporter construct did not stabilize HIF1α at oxygen levels not considered hypoxic, B16-HG cells were incubated at 10% O_2 for 48 hours, and HIF1α-eGFP levels were measured by FACS analyses (**Figure 10, far left**). The cells showed no significant increase in eGFP signal, and therefore no significant increase in HIF1α protein levels.

Due to the variability in literature reported *in vitro* hypoxic incubation time lengths, we set out to determine the optimal *in vitro* hypoxia incubation time which leads to maximal HIF1α-eGFP signal. To accomplish this, B16-HG cells were incubated at 1% O_2 for 24 or 48hrs, after which HIF1α protein levels were measured by eGFP FACS levels (**Figure 10, middle left**). We found that HIF1α levels were highest after 48 hours of hypoxic incubation, increasing by ~3 fold from normoxic conditions. Based on this experiment, we set 48 hours at 1% O_2 as our standard for *in vitro* hypoxic incubation.

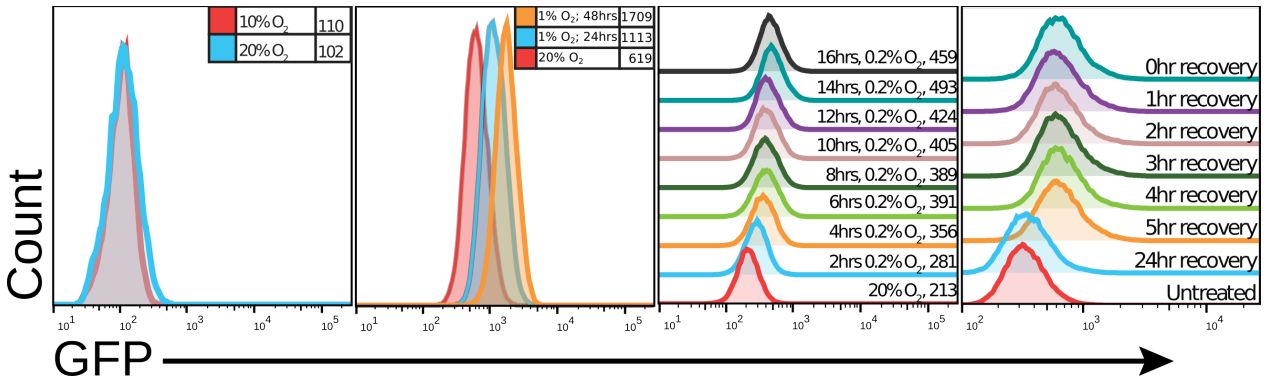

Figure 10: In vitro Hypoxic incubation optimization using the B16-HG cell line

All values are geometric means of GFP signal. **Left panel**: B16-HG cells were incubated at 20% O_2 or 10% O_2 for 48hrs. **Middle-left panel**: B16-HG cells were subjected to standard tissue culture conditions (red), or *in vitro* hypoxic incubation at 1% O_2 for 24 (blue) or 48 (orange) hours. **Middle-right panel**: B16-HG GFP signal induction kinetics via *in vitro* hypoxia incubation at 0.2% O_2 for indicated time. **Far-right panel**: B16-HG GFP signal degradation following re-exposure to normoxia after 48hrs incubation at 1% O_2.

Given that we are using our exogenous HIF1α-eGFP signal to track endogenous HIF1α stabilization levels, it is important to verify that the HIF1α-eGFP protein had similar induction and degradation rates as endogenous HIF1α[541,542]. To determine the rate of HIF1α-eGFP stabilization, B16-HG cells were incubated at 0.2% O_2 for increasing amounts of time, and HIF1α levels were quantified by eGFP FACS levels. This extreme hypoxic condition was used to maximize oxygen diffusion out of the media and into the air while still being hypoxic. Through this extreme hypoxic incubation, we found slight HIF1α-eGFP stabilization after as little as two hours (**Figure 10, middle right**). This correlates with previously published rates of HIF1α induction[542,543]. These results also highlight a benefit of using HIF1α for the reporter read-out, as HIF1α protein levels will logically accumulate faster than the genes and proteins it induces, making it the fastest endogenous protein-based reporter for cellular hypoxia. Therefore using a HIF1a protein-based reporter should have faster induction kinetics than HRE-based reporters.

HIF1α-eGFP protein turnover rate following re-oxygenation was measured by first incubating B16-HG cells at 1% O_2 for 48 hours, then allowing the cells to recover under normoxic conditions for a set amount of time. eGFP levels were stable for up to 5 hours of normoxic

recovery, and eGFP did return back to baseline levels after 24 hours of normoxic incubation (**Figure 10, far right**). This is a significantly longer amount of time than previously published exogenous HIF1α-eGFP expression constructs, which showed 90% degradation after 20 minutes of re-oxygenation, as well a previously published rates of degradation of endogenous HIF1α[541,542]. This may be caused by the GFP tag causing slower degradation times, incomplete degradation or release of the GFP at the proteasome, or hypoxic upregulation of oxygen insensitive alternative start codons present in the mRNA[544]. To address the first hypothesis, generation of a murine cell line expression unlabelled human HIF1a could be generated, and protein degradation rates would be determined by Western blot or intracellular staining for FACS analysis. By completely removing the exogenous label, we remove any artifact in protein degradation introduced by labelling the protein. The last two hypotheses could be addressed with a anti-GFP Western blot of the B16-HG cell line after hypoxic incubation, probing the entire length of the membrane. This would allow for the detection of partial degradation products from the HIF1α-eGFP fusion protein, or proteins derived from alternative start codons.

Having established a baseline for the dynamics of our B16-HG GFP signal in an *in vitro* hypoxic incubation setting, we next sought to investigate how the addition of other physiological stresses which correlate with *in vivo* hypoxia might affect HIF1α-eGFP levels. We began by attempting to simulate metabolic stress in conjunction with hypoxia, given their known correlation[231]. We attempted to simulate this aspect though *in vitro* hypoxic incubation and modification of the cellular media. We reduced the general metabolite concentration by changing the media from DMEM to RPMI media. Doing this halves the amount of glucose available to cells and reduces the availability of several amino acids. To further reduce the amount of available metabolites and growth signals, we also modulated the FBS concentration from 10-0%. After 48 hours of incubation in this modified media under normoxia or 1% O_2, the B16-HG cells were analyzed by FACS for HIF1α-eGFP protein levels.

From this experiment an initial finding was that under normoxia, changing the media from DMEM + 10% FBS to RPMI media+ 10% FBS did not affect HIF1α levels (**Figure 11A**). When the FBS was fully removed in specifically RPMI media under normoxia, HIF1α levels reached those seen in standard B16 media under hypoxia (**Figure 11B**). Interestingly, when B16-HG cells

were exposed to 1% O_2 for 48 hours in RPMI media devoid of FBS, HIF1α levels reached about double the level seen when cells were incubated in hypoxia in the standard media (**Figure 11C**). This protocol suggested that we were inducing significantly higher levels of HIF1α by specifically incubating cells in RPMI media without FBS. Additionally, the HIF1α protein levels reached through hypoxic incubation in RPMI 0% FBS were approaching the levels observed through $CoCl_2$ treatment, suggesting we were close to reaching HIF1α protein saturation levels (**Figure 11D**). Moving forward, this more restrictive hypoxic incubation protocol will be referred to as RHY. Please note that FACS voltage settings were lowered for panel Figure 11D, which is why MFI values are lower than in panels A-D.

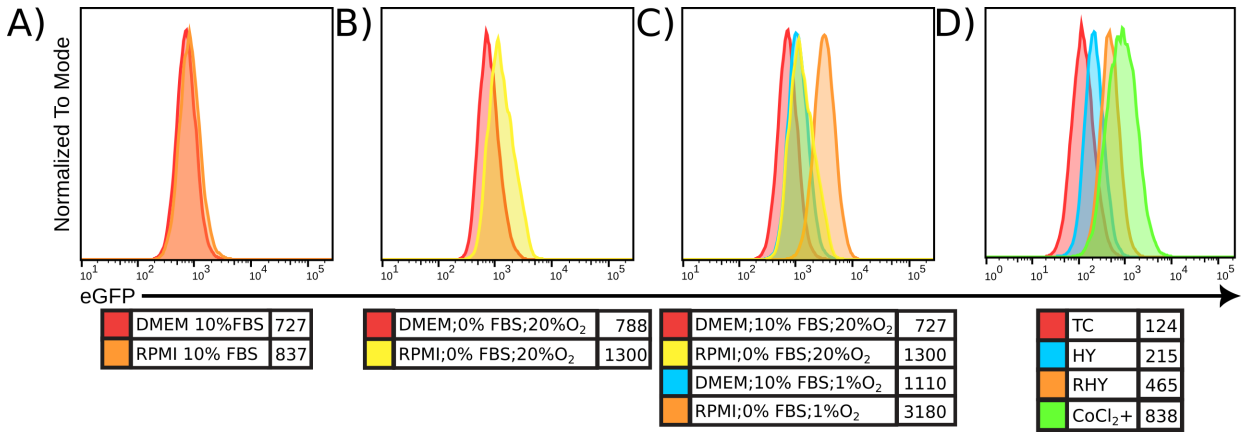

Figure 11: Optimization of metabolic stress in conjunction with hypoxic incubation

A-D) B16-HG cells GFP levels after 48hr incubation in indicated media. Numbers in legends are the geometric mean of GFP signal.

3.3.2 Transcriptomic investigation into the effect of nutrient deprivation on hypoxia adaptation in tumour cells *in vitro*

Given the increased HIF1α protein levels observed through the RHY protocol, we reasoned this protocol would induce a hypoxia adaptation program more representative of *in vivo* adaptation. To investigate this, we performed deep transcriptome profiling on B16-HG cells exposed to one of three culture settings. For 48 hours, B16-HG cells were grown in either standard tissue culture conditions (referred to as "TC"), 1% O_2 in standard B16 media (referred to as HY), or 1% O_2 in RPMI 0% FBS (referred to as RHY) (**Figure 12**). After incubation, cells were

harvested and processed for rRNA-depleted mRNAseq and smallRNAseq. This experiment was performed using serial biological duplicates. The total amount of reads acquired and the total mapped reads per sample are listed in **Table 14**. This experiment which focuses on the additive effects of metabolic stress in conjunction with hypoxic incubation, will be referred to as the MHY experiment, or MHY dataset.

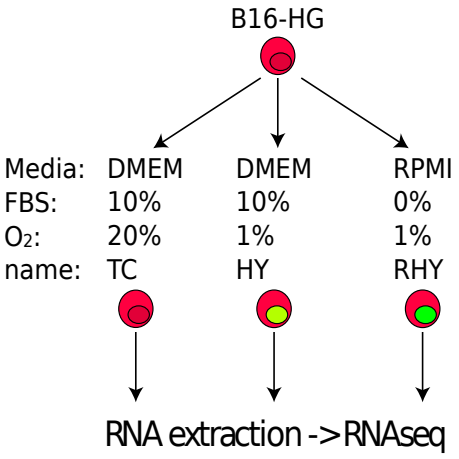

Figure 12: MHY treatment schema

Table 14: MHY mRNAseq read mapping values

Condition	N	Total number of reads acquired	Total number of reads after filtering	Total number of filtered reads mapped
TC	1	80,508,218	62,197,134	50,961,940
	2	110,792,244	83,697,368	69,204,828
HY	1	118,252,446	89,569,204	70,714,298
	2	115,113,564	89,700,110	73,780,158
RHY	1	98,863,414	74,871,082	60,893,060
	2	120,916,614	91,282,010	76,078,124

3.3.2.2 Metabolic stress is an additive transcriptomic adaptation to hypoxia adaptation

Both rRNA-depleted mRNAseq and miRNAseq datasets showed proper sample clustering by PCA analyses (**Figure 13A**). Differentially expressed gene (DEG) analyses show that HY treatment caused the differential expression of 1705 genes, whereas RHY resulted in 4474 by the same thresholds (padj < 0.05) (**Table 15**). Additionally, the RHY DEG list encompasses most of the HY DEG list (86%), suggesting that the RHY protocol induced more of an additive effect to hypoxic adaptation rather than a unique effect on the transcriptome (**Figure 13B**).

Table 15: MHY dataset summary of differentially expressed genes

Comparison	Total	Up-regulated	Down-regulated
HYvsTC	1705	943	762
RHYvsTC	4474	2334	2140
RHYvsHY	989	531	458

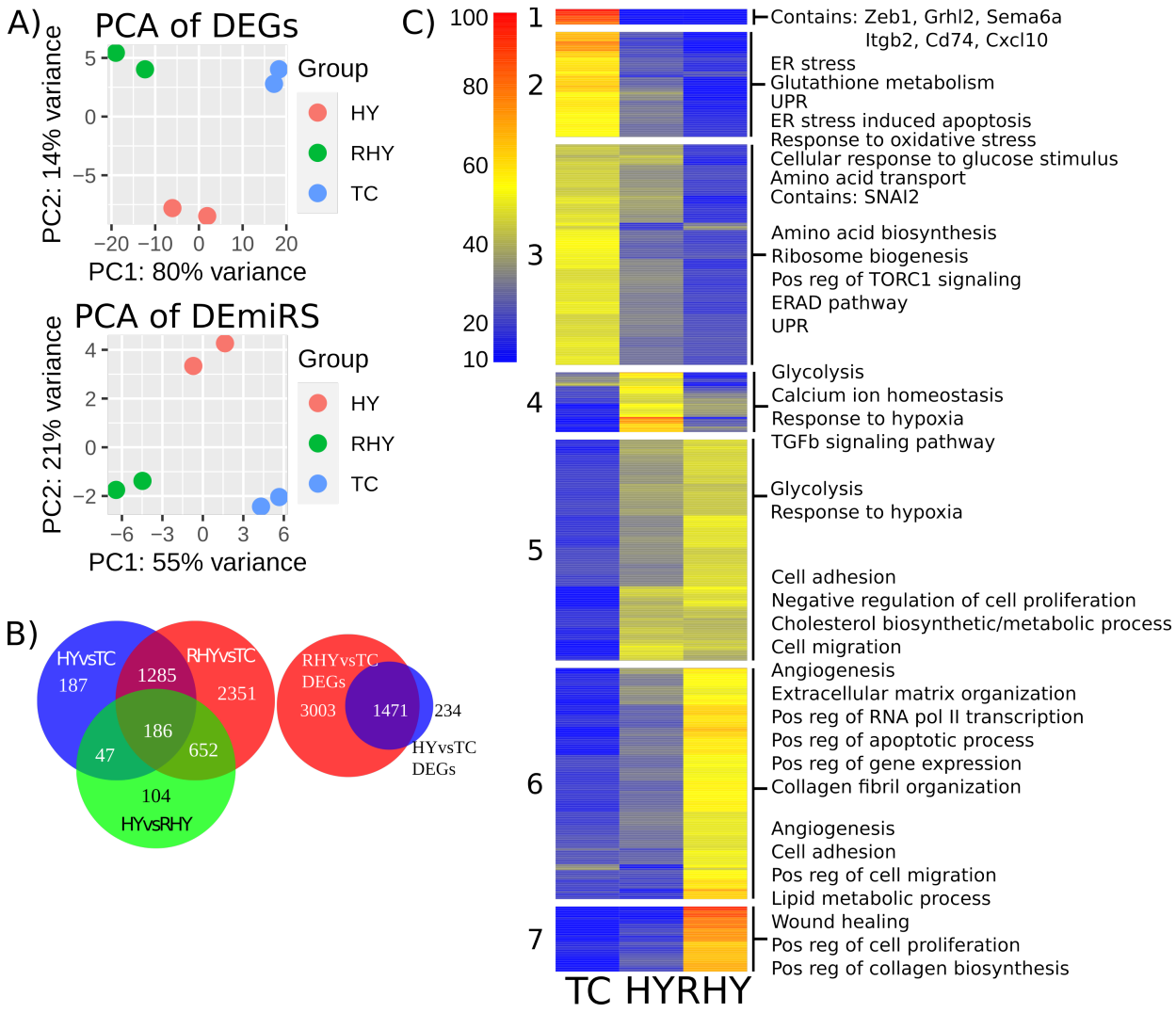

Figure 13: MHY RNAseq data overview

A) PCA analyses of MHY RNAseq data. Top: PCA plot of top 500 variable genes in MHY dataset. Bottom: PCA plot of top 500 variable miRs in MHY dataset. **B**) Venn diagrams of MHY conditions. Left: Unweighted three-way Venn diagram of differentially expressed genes (padj < 0.05) for displayed comparisons. Right: Weighted two-way Venn diagram of DEGs (padj < 0.05) relative to TC. **C**) Hierarchically clustered heatmap of DEGs (padj<0.05 & FC>=2 & minExp>=100) from all condition comparisons. Gene expression is normalized by contribution to total expression. Clustering was done using the Ward.D2 metric. GO term enrichment was performed using DAVID, and listed GO terms are significantly enriched (p<0.05).

3.3.2.3 RHY prioritizes motility-based hypoxic adaptation over metabolic-based adaptation

In order to obtain a more global understanding of the transcriptomic changes induced by the MHY protocol, DEGs from all condition comparisons were rendered as a hierarchically clustered heatmap. Gene clusters were re-ordered without breaking the clustering linkage for aesthetic purposes. The original heatmap, with dendrogram, can be found in the Supplemental Material of this thesis (**Figure 58A, supplemental material**). Seven clusters were chosen as an intuitive number based on clustering dendrogram. Rendering the heatmap using DESeq2 normalized readcounts for replicates rather than averaged values recovered similar clustering patterns (**Figure 58B, supplemental material**). Therefore, we focused the analysis on the heatmap generated with averaged values for simplicity.

We can see that most of the genes are being induced under either of the hypoxic conditions, with about half of the hypoxia-upregulated genes having higher expression levels under RHY than HY (**Figure 13C**). Moving down the heatmap, we can see that the top three clusters represent genes being down-regulated under increasingly stressful conditions. The forth cluster contains genes upregulated exclusively under HY conditions. The fifth cluster contains genes that are relatively equally upregulated under either hypoxic condition. The bottom two clusters contain genes which are increasingly expressed as cells are exposed to increasing stressful conditions.

To determine enriched functionalities within these clusters of similarly regulated genes, GO term analyses were run using DAVID. The top three clusters represent genes who's expression drops significantly in either hypoxic incubation, with specifically low expression under RHY. Some of the phenotypes enriched in these clusters are involved in amino acid transport and glutathione metabolism. This is likely reflective of the relatively low concentrations for all amino acids in RPMI compared to DMEM. Within these clusters, there is also downregulation of EMT regulators ZEB1 and GRHL2, as well as Slug (SNAI2) in cluster two. Classically, ZEB1 induces EMT whereas GRHL2 inhibits EMT, while Slug is an essential transcription factor partner for other EMT driving transcription factors like TWIST1[545,546]. This suggests that hypoxic incubation results in a re-wiring of EMT regulators.

Specifically within cluster two, there is a down-regulation in the cellular response to glucose stimulation. Despite the upregulation of glycolysis genes in later clusters, this indicates that certain aspects of glucose response are down-regulated. This drop in gene expression may also be an effect of the metabolic restrictions imposed by RHY, as RPMI contains less glucose than DMEM, and no additional glucose is supplemented through FBS. Also within cluster two there are genes associated with ER stress-induced apoptosis, suggesting a resistance to this apoptosis avenue through down-regulation of key genes. In line with this lack of nutrients in RHY, there is enrichment for genes associated with amino acid and ribosome synthesis in cluster three, where genes are progressively shut down from TC to HY to RHY. This suggests that due to the reduced nutrient availability, cells are reducing the generation of biomass needed for cell division. This is in line with the observation of reduced cell number after RHY incubation in comparison to HY incubation.

Cluster four represents genes which induced under HY conditions, yet drop in conjunction with metabolic stress. Within this cluster are genes associated with the TGFβ signalling pathway. HIF1 has been shown to activate TGFβ signalling in the literature[547], yet these genes drop in expression under RHY conditions. This suggests that anti-TGFβ therapeutics would not be effective on tumour cells exposed to hypoxia and nutrient starvation, as these cell may have disengaged TGFβ signalling[548].

Cluster five contains genes that are generally upregulated equally across the HY and RHY conditions. Within this cluster, enriched phenotypes are centred on glycolysis and genes classically associated with the cellular response to hypoxia, such as CAR9, VHL, and MPP14.

Clusters six and seven contain genes specifically upregulated under RHY conditions. These clusters are heavily enriched for phenotypes associated with cell motility and secondary responses to hypoxia *in vivo* like angiogenesis, wound healing, and extracellular matrix organization. There is also enrichment for transcription associated genes, potentially explaining the increased number of DEGs RHY (**Figure 13B, right**). There is also an enrichment for genes associated with lipid and cholesterol biosynthesis and metabolism. Given the lack of FBS in RHY, this increase in cholesterol biosynthesis is necessary, as there is no cholesterol present in RPMI.

In a similar vein, lipid metabolism is likely upregulated due to the sever lack of nutrients present in RHY.

While GO term analyses can provide insight into enriched phenotypes in sets of genes, it cannot provide an understanding of how a given phenotype is enriched across the entire transcriptome. To address this, GSEA analyses were run on lists of DEGs relative to TC using the Hallmarks gene set list (**Figure 14**). Hallmark sets for glycolysis, hypoxia, and EMT were all enriched in both HY and RHY datasets. While the hypoxia and glycolysis hallmarks were both more enriched in the HY dataset (**Figure 14A&B**), RHY had higher enrichment for EMT than HY (**Figure 14C**). This suggests that the combination of hypoxic and metabolic stress is sufficient to drive B16 cells to metastasize, as the RHY condition dramatically increases enrichment for EMT associated genes.

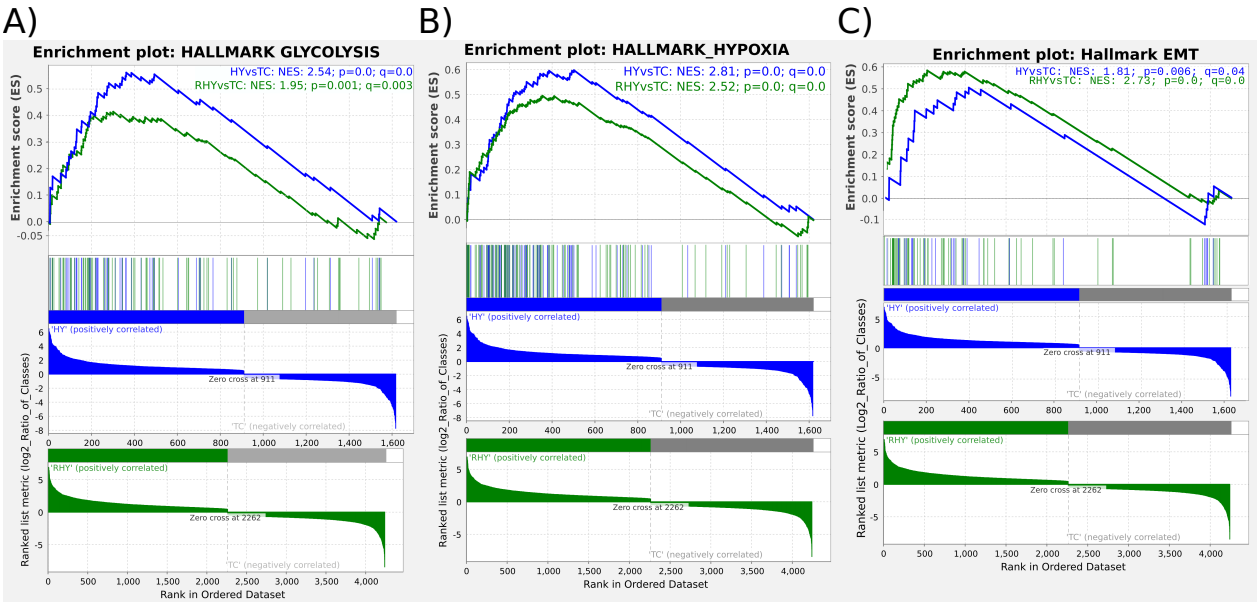

Figure 14: MHY comparative GSEA analyses

Comparative GSEA curves for HYvsTC (blue) and RHYvsTC (green) for GSEA Hallmark genesets for **A**) Glycolysis, **B**) Hypoxia, and **C**) EMT

To gain a deeper understanding of the phenotypes enriched across the transcriptome of either hypoxic conditions, GO term analyses were performed on the upregulated DEG list for each hypoxic condition in comparison to TC. Using DAVID for initial GO term enrichment

calculations, GO term redundancy was reduced using REViGO, and all indispensable GO terms were rendered in bubble plots (**Figures 15&16**). These plots show that the HY condition primarily activated classical hypoxia-associated phenotypes; such as response to hypoxia, angiogenesis, and especially glycolysis (**Figure 15**). Whereas for the RHY condition, there is a much larger number and diversity of phenotypes, as well as more genes associated per phenotype (**Figure 16**). In the RHY GO term bubbleplot, there are several phenotypes associated with EMT that are either unique to RHY or have a higher gene count than HY. These include cell adhesion, positive regulation of cell migration, extracellular matrix organization, positive regulation of collagen biosynthesis, and establishment or maintenance of cell polarity.

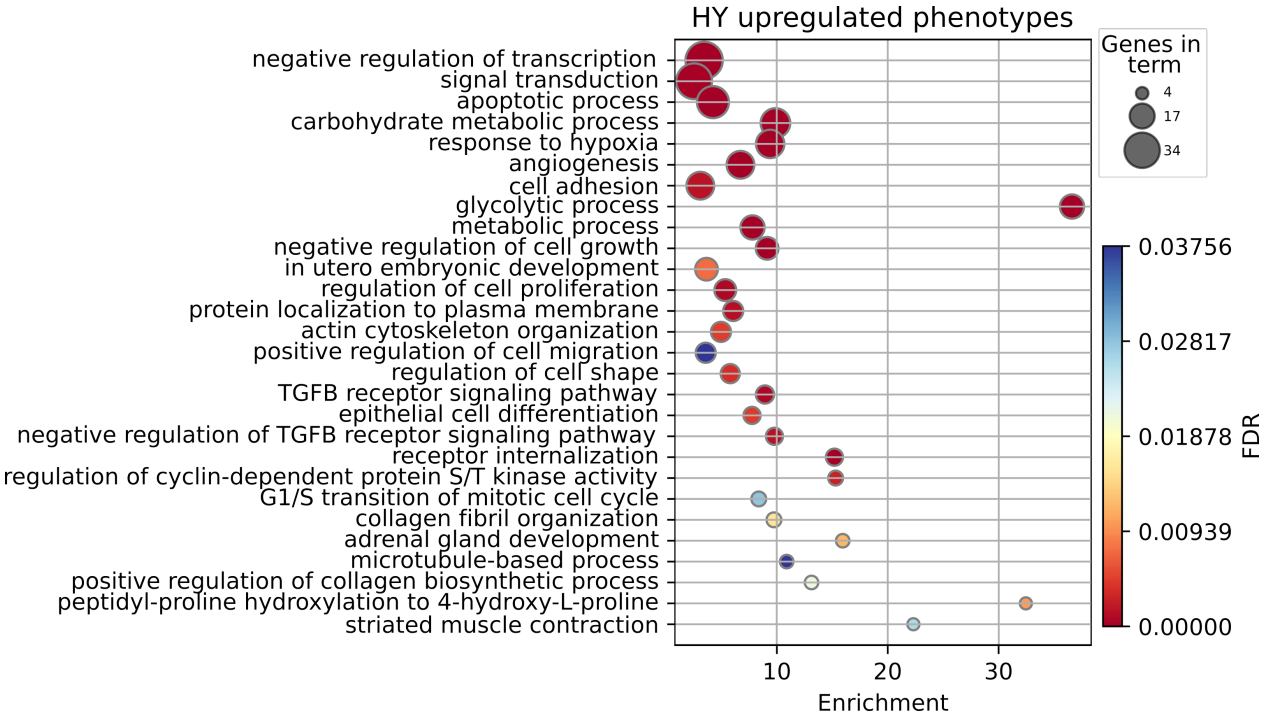

Figure 15: HY upregulated phenotypes bubbleplot visualization

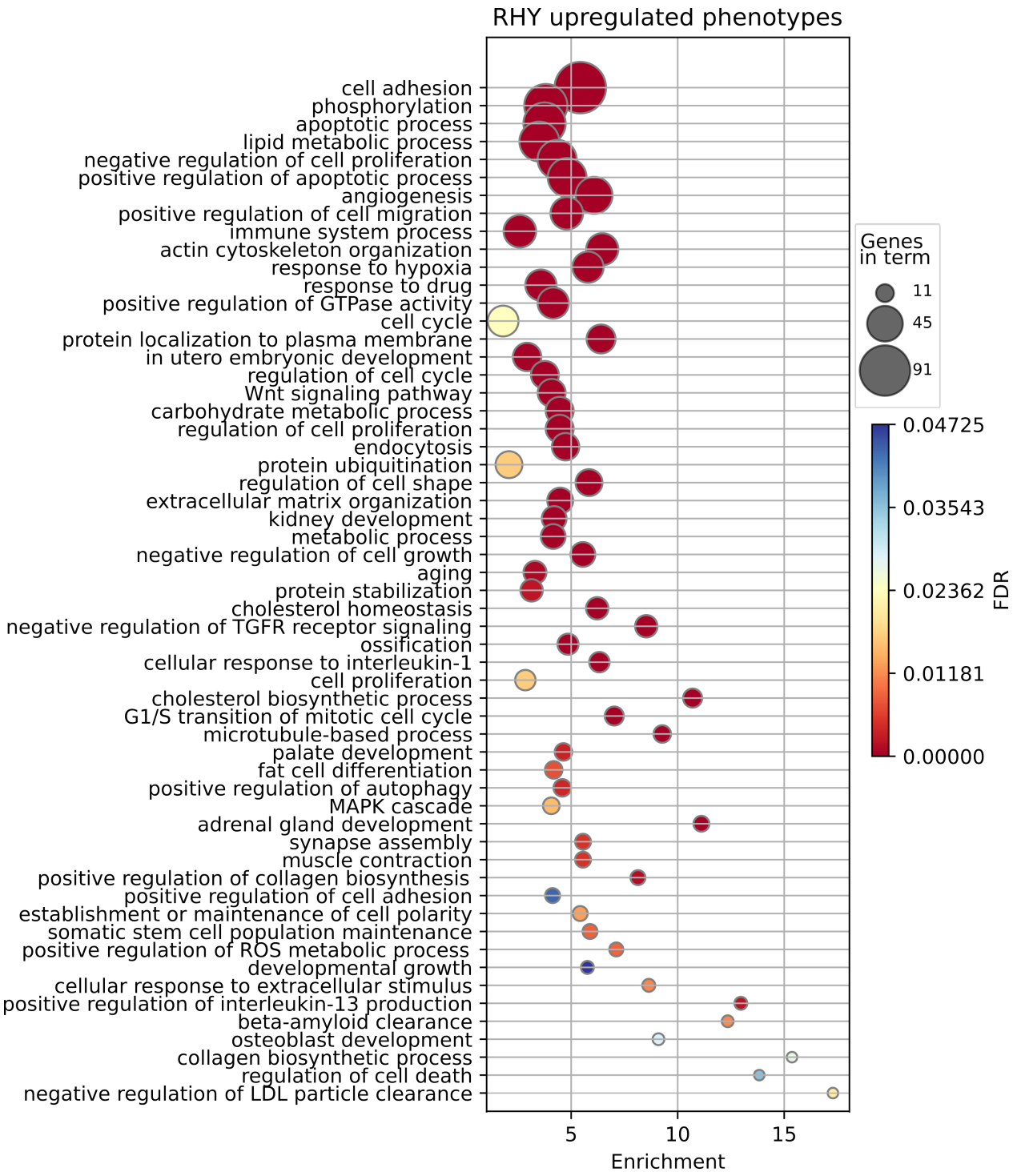

Figure 16: RHY upregulated phenotypes bubbleplot visualization

3.3.2.4 HY & RHY induce condition specific changes in glycolytic gene variants

To investigate the differential regulation of specific pathways, gene interaction networks were generated in Cytoscape[515] with the relative gene expression pattern mapped to each node. Given the known role of glycolysis in the hypoxia literature, as well as the differential enrichment in HY and RHY, we investigated the glycolysis pathway (**Figure 17**). Among the glucose importers, SLC2A3 and SLC5A1 were both significantly down-regulated in either hypoxic condition, with lowest expression being in RHY. While SLC2A3 is known to be upregulated under hypoxia, SLC5A1 has no links to hypoxic regulation[549]. As if to compensate, SLC2A1, SLC5A3, and SLC5A10 all showed up-regulation under hypoxic conditions. Both hexokinase variants were upregulated relatively equally under either hypoxic condition.

As if to aide in elevating glucose-6-phosphate levels in the cell, galactose metabolism through the Leloir pathway is also enriched under hypoxic conditions (**Figure 17, purple left side**). Galactose Kinase 1 (GALK1) is upregulated under hypoxic conditions, generating elevated levels of Galactose-1-Phosphate. There is also an variant preference switch for Phosphoglucomutase (PGM1/2), where PGM1 is moderately down-regulated under hypoxic conditions, PGM2 is significantly upregulated under hypoxic conditions. This drop in PGM1 expression is unexpected, as PGM1 is a known HIF1α target gene[550].

Continuing through the glycolysis pathway, the majority of genes are upregulated under either hypoxic condition, with some notable caveats. Among the phosphofructokinase variants, there is a switch between PFKP to PFKL being the major variant in HY and RHY respectively. While ALDOA is the major variant expressed in the dataset, there is a different expression pattern for ALDOC, with significant upregulation under RHY. ALDOC has been associated with tumour progression in melanoma brain metasteses[551]. Curiously, there is upregulation of Bisphosphoglycerate Mutase (BPGM) and Multiple Inositol-Polyphosphate Phosphatase 1 (MINPP1) under RHY conditions (**Figure 17, purple right side**). These genes make up the Rapoport-Luebering Shunt, a metabolic shunt that by-passes the first ATP generating step of glycolysis[552–554]. The upregulation of this shunt may be an attempt to lower the activity of the AMPK pathway by lowering intracellular levels of 3-phosphoglycerate levels[555].

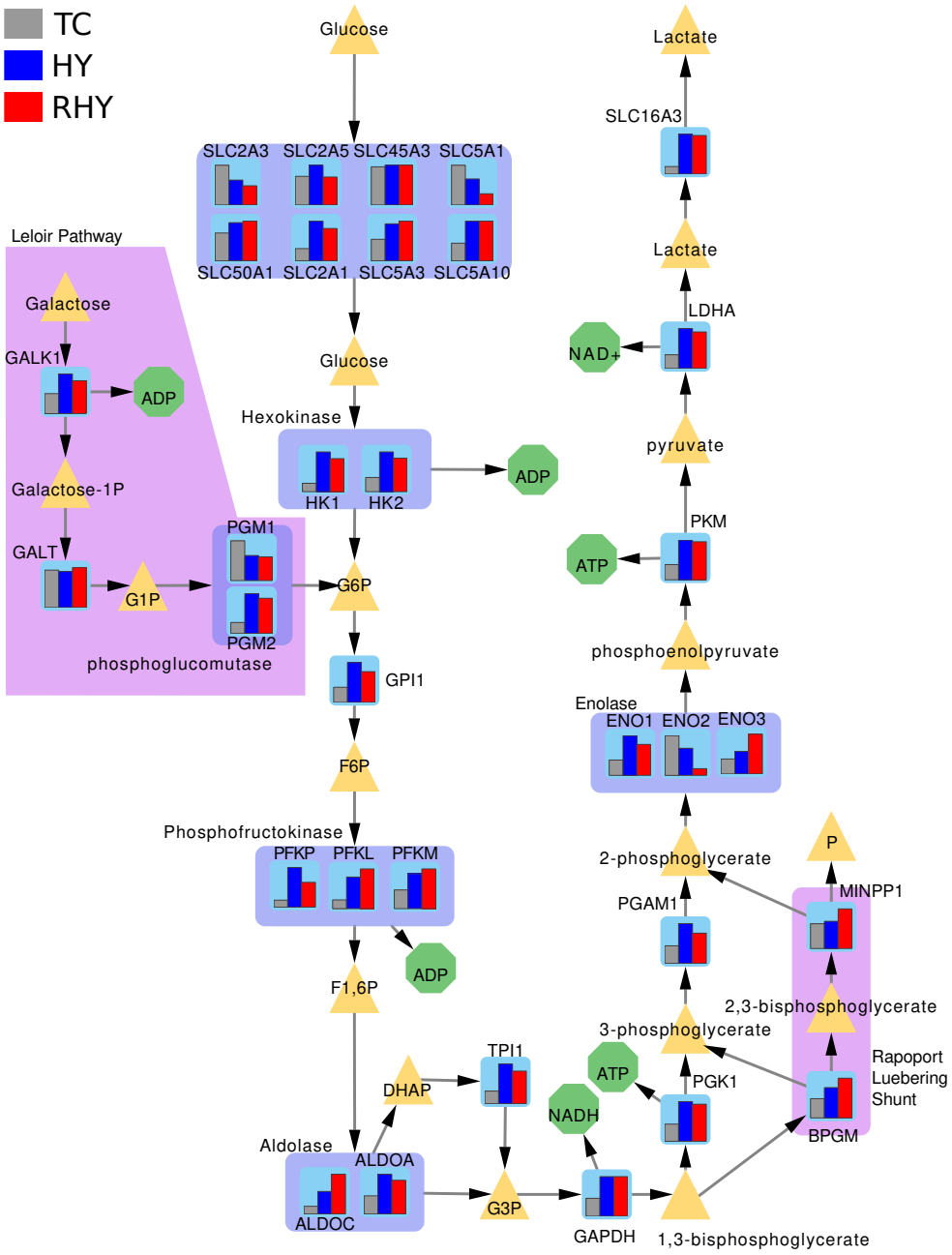

Figure 17: MHY glycolysis gene expression network visualization

Expression of glycolysis genes in MHY dataset. Genes are blue squares, with relative MHY RNAseq expression values mapped to grey (TC), blue (HY), and red (RHY). Expression is averaged DESeq2 normalized reads. Genes with similar function are grouped together. Intermediate metabolites are symbolized as yellow triangles. Energy carrying molecules are symbolized as green octagons. Sub-pathways are denoted with a purple background.

Past the phosphoglycerate metabolizing genes, the Enolase variants are differentially regulated under different culture conditions. Although ENO1 is the major variant expressed in these cells, it appears that there is a variant expression preference for each cell-state: ENO1 is preferred in HY, ENO2 is preferred in TC, and ENO3 is preferred in RHY. Beyond the Enolase variants, the remaining cytosolic glycolysis genes are equivalently upregulated under both hypoxic conditions, with the terminal lactate being exported to the extracellular space by MCT4 (SLC16A3).

3.3.2.5 MHY promotes BCAA import and reduces BCAA metabolism

Amino acid metabolism is also heavily modulated in the MHY dataset. While few transporters maintain stable expression (SLC6A17, SLC7A6, SLC6A15, SLC36A4, SLC38A10) (**Figure 18A**) most expressed transporters are significantly down-regulated (**Figure 18B**). SLC7A11 shows significant down-regulation in the MHY dataset, with this loss of expression being exacerbated under RHY. SLC7A11 (also known as xCT) is a cysteine-glutamate antiporter at the plasma membrane, increasing cytosolic cysteine levels[556]. In triple-negative breast cancer, functional loss of SLC7A11 led to a drop in intracellular cysteine levels resulting in functional inhibition of PHD2, allowing for HIF1α stabilization[150]. The SLC381/2 glutamine transporters are down-regulated with only moderate insignificant upregulation of SLC38A6 under RHY (**Figure 18C**). SLC38A1,2,7 are all down-regulated under HY, with expression falling further under RHY (**Figure 18B**). SLC1A4/5 are neutral amino acid exchangers, often associated with glutamine uptake in cancers[557]. While these transporters are expressed in the B16-HG cell line, their expressions drop under hypoxic conditions, suggesting a reduction in glutamine import into the cells (**Figure 18B**).

The top amino acid transporters that are upregulated under MHY conditions are in the L-type Amino acid Transporters (LAT1-4) family, which transport large neutral amino acids across the plasma membrane[558]. While LAT2 is not expressed in our cell line, LAT1,3,4 show differential expression under MHY conditions (**Figure 18D**). LAT1 (SLC7A1) is significantly down-regulated under hypoxic conditions. Additionally 4F2 cell-surface antigen (SLC3A2), a common component of LAT heterodimeric complexes[559,560], is also significantly down-regulated (**Figure 18C**).

LAT3 (SLC43A1) is up-regulated two-fold in either of the hypoxic conditions. LAT4 (SLC43A2) is only significantly upregulated under RHY conditions. These transporters are mainly known for transporting large neutral amino acids across the plasma membrane. Despite the up-regulation of these transporters the expression of BCAT1, the subsequent enzyme in branched-chain amino acids metabolism[561], is significantly down-regulated under hypoxic conditions (**Figure 18D**). This is in contradiction with a previously published study showing BCAT1 as a direct target of HIF1 in glioblastoma, and that BCAT1 is overexpressed in melanoma[561,562]. A possible explanation for this dependency is the *BCAT1* HRE identified in human glioblastoma is not present in the murine genome, or it is non-functional in mice. Alternatively, the HRE could be mutated specifically in the B16-HG cell line, either as a feature of the parental cell line, or as a result of our single-cell sorting. Assuming the murine/B16-HG *BCAT1* HRE is functional, it may be that hypoxic downregulation of *BCAT1* is a melanoma specific adaption. This would be counter-intuitive, as BCAT1 is generally upregulated in melanoma, and positively contributes towards proliferation and migration[563,564]. What resolves this is the BCAT1 implication in proliferation and migration is through oxidative phosphorylation under normoxic conditions[564]. As oxidative phosphorylation is down-regulated under hypoxia, it follows that *BCAT1* may be down-regulated as well.

Additionally, KLF15, a master regulator of glycemic, lipid, and amino acid metabolisms[565] is significantly down-regulated under hypoxic conditions (**Figure 18D**). KLF15 controls the expression of *BCAT2*[566] which maintains expression under all MHY conditions, suggesting compensatory expression by another transcription factor. Aside from these changes the main metabolic pathways for valine, leucine, and isoleucine are expressed with moderate regulation across the conditions. This suggests that BCAA catabolism could occur, but is dependant on the limited activity of BCAT1 and SLC2A44 to generate and import the needed intermediates to drive the rest of the pathway.

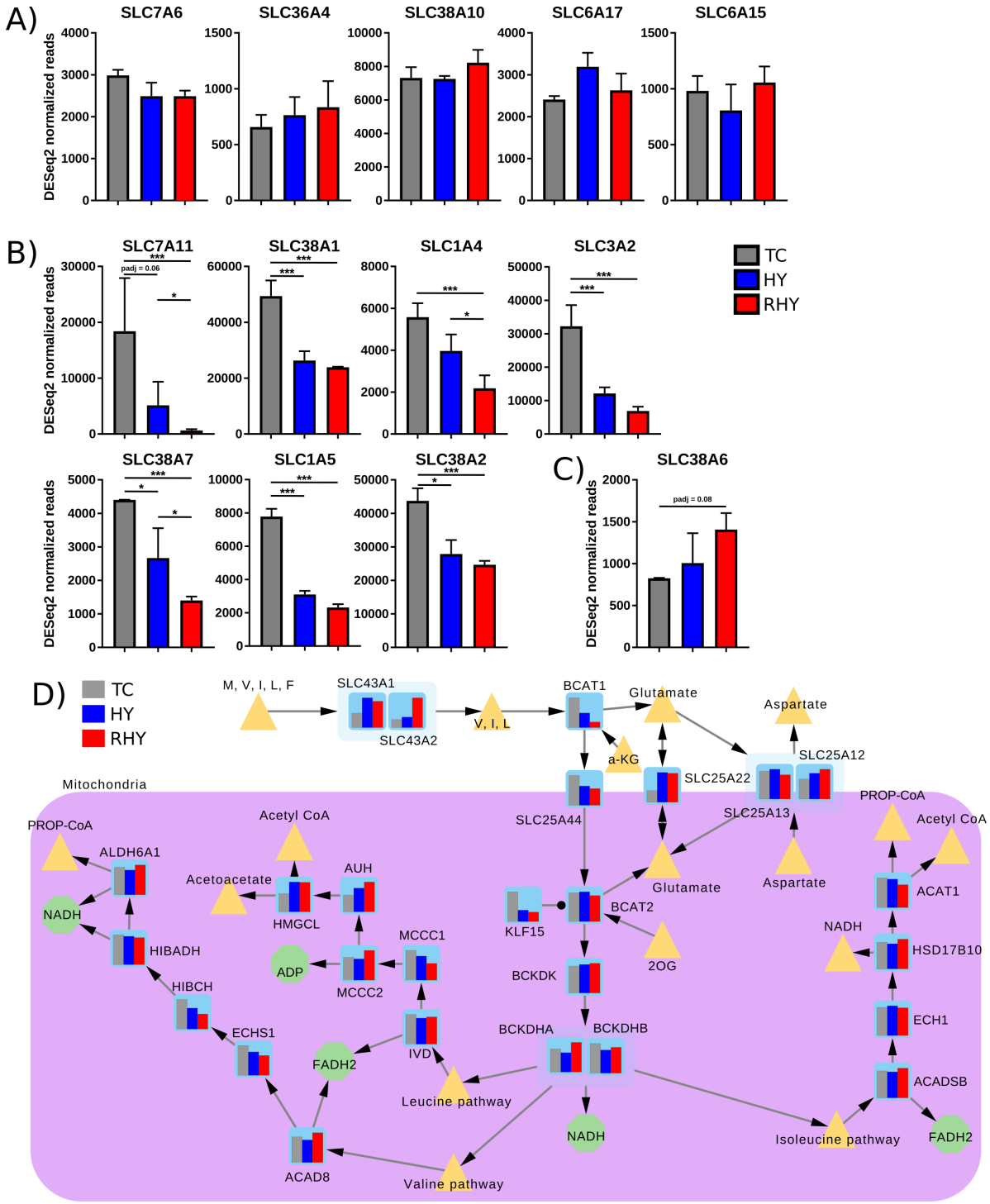
A)
SLC7A6
SLC36A4
SLC38A10
SLC6A17
SLC6A15
DESeq2 normalized reads
B)
SLC7A11
SLC38A1
SLC1A4
SLC3A2
SLC38A7
SLC1A5
SLC38A2
padj = 0.06
TC
HY
RHY
C)
SLC38A6
padj = 0.08
D)
M, V, I, L, F
SLC43A1
SLC43A2
V, I, L
BCAT1
Glutamate
Aspartate
a-KG
SLC25A12
SLC25A22
SLC25A44
SLC25A13
Mitochondria
PROP-CoA
Acetyl CoA
ALDH6A1
NADH
Acetoacetate
AUH
HMGCL
HIBADH
HIBCH
ECHS1
ADP
MCCC2
MCCC1
FADH2
IVD
ACAD8
Leucine pathway
Valine pathway
KLF15
BCAT2
2OG
BCKDK
BCKDHA
BCKDHB
NADH
Isoleucine pathway
ACADSB
FADH2
ECH1
HSD17B10
ACAT1
NADH
PROP-CoA
Acetyl CoA

Figure 18: MHY BCAA metabolism gene expression network

A-C) Gene expression histograms of stably expressed (**A**); down-regulated (**B**); and up-regulated (**C**) amino acid transporters across MHY conditions. Expression is normalized DEseq2 reads, error bars are SD. * denotes relative significance as calculated by DESeq2 Benjamini-Hochberg adjusted p-value (padj). *: padj < 0.05, **: padj < 0.01, ***: padj < 0.001. **D**) BCAA metabolism pathway with MHY gene expression mapping. Genes are blue squares, with relative MHY RNAseq expression values mapped to grey (TC), blue (HY), and red (RHY). Expression is averaged DESeq2 normalized reads. Genes with similar function are grouped together. Intermediate metabolites are symbolized as yellow triangles, not all intermediates shown. Energy carrying molecules are symbolized as green octagons.

3.3.2.6 Hypoxic incubation reduces glutamine import and promotes metabolism

Despite the prevalence of glutamine metabolism in hypoxic adaptation[567–570], there is a drop in glutamine importers SLC38A1/2 and SLC1A5 in the MHY dataset (**Figure 19**). This opposes recent work done by Morotti et al demonstrating upregulation of SLC38A2 under hypoxic conditions in several human breast cancer cell lines[568]. This difference may be caused by Morotti et al observing this effect under endocrine therapy conditions. This combination of cellular stresses may explain our differing results. SLC1A5 has been shown to be upregulated under hypoxic conditions, however this regulation was under the control of HIF2α (EPAS1)[570]. Based on the RNAseq data from the B16-HG cell line HIF2α is not significantly expressed under any condition, giving a mechanism for the lack of SLC1A5 expression and perhaps other transporters (**Figure 19**). The glutamine-cystiene antiporter SLC7A11[571] is also significantly down-regulated under RHY. There is moderate upregulation of SLC7A7, however the overall expression level of this transporter is the lowest of the glutamine importers. Based on the expression patterns of these transporters, there is a significant reduction in glutamine import into the cell under HY and especially under RHY conditions.

Once inside the cell, glutamine is converted to glutamate by Glutaminase (GLS1/2)[569,572]. GLS1 mRNA is moderately upregulated specifically under RHY conditions, while GLS2 is upregulated under HY and loses this upregulation under RHY (**Figure 19**). This suggests that glutamate synthesis is occurring during hypoxic incubation. Currently in the literature, there is

an focus on mitochondrially localized GLS[569,572,573]. Based on the work of Yoo et al, it is believed that cytosolic glutamine is transported to the mitochondrial matrix via an alternatively spliced isoform of SLC1A5 controlled by HIF2α[570]. In addition to the drop on SLC1A5 expression, there is no evidence for transcription initiating from the alternative transcription start site to generate this alternative isoform (data not shown). Despite the literature's focus on mitochondrial GLS, there are previous immunofluorescence and organelle isolation experiments that have shown GLS isoforms can localize to the cytosol[572]. Given these findings, it is possible that cytosolic glutamine is being converted to glutamate via cytosol localized GLS.

In order to be further metabolized, cytosolic glutamate must be transported to the mitochondrial matrix. This can be accomplished by transporters such as SLC25A22, which is upregulated equally under either hypoxic condition, indicating an increased exchange of glutamate across the inner mitochondrial membrane (**Figure 19**)[574,575]. Once inside the mitochondrial matrix, glutamate can be converted to α-ketoglutarate via glutamate dehydrogenase (GLUD1), which shows moderate but insignificant upregulation under RHY conditions specifically. GLUD1 generates α-ketoglutarate, which can then be used in the TCA cycle, used as a cofactor in a myriad of metabolic reactions, or be transported by SLC25A12/13 back into the cytosol.

3.3.2.7 RHY exacerbates TCA cycle reprogramming

Given the necessity of pyruvate in hypoxic glycolysis to regenerate NAD+, the TCA cycle is often differentially regulated under hypoxia[549,576]. The metabolic link between glycolysis and the TCA cycle is through pyruvate. Pyruvate is transported from the cytosol to the mitochondrial intermembrane space via VDAC1, and is then imported into the mitochondrial matrix via the Mitochondrial Pyruvate Carrier, a heterodimer of MPC1/2[577,578]. VDAC1 is significantly upregulated in both hypoxic conditions, suggesting greater metabolite flow between the cytosol and the mitochondrial intermembrane space (**Figure 19**). MPC1/2 show moderate down-regulation under RHY, with only MPC1 having significant down-regulation. While MPC2 has been known to homodimerize, it appears that MPC1 is necessary for pyruvate import[579,580]. Therefore, it can be assumed that there is a reduction in pyruvate import into the mitochondrial matrix.

An alternative source for mitochondrial pyruvate is the conversion of malate to pyruvate via the activity of Malic Enzymes[581]. In both HY and RHY, the mitochondrial isoforms of Malic Enzyme (ME2/3) are significantly upregulated, suggesting that mitochondrial malate is being removed from the TCA cycle and being converted back into pyruvate (**Figure 19**). This pyruvate could then be reintroduced back into the TCA cycle, or exported back to the cytosol via MPC1/2 and VDAC1 and removed from the cells via the glycolytic conversion to lactate and exported via MCT4[576,582]. Alternatively, mitochondrial pyruvate can be used to synthesize alanine via Alanine Transaminases (GPT & GPT2)[556], however both of these genes are down-regulated under RHY (**Figure 59A&B, supplemental material**). Cells may be using the malic enzyme reaction to create more NADPH, for use in in FA synthesis, cholesterol synthesis, or to combat ROS damage[581,583]. NADPH can also be converted to NADH via nicotinamide nucleotide transhydrogenase (NNT), which is moderately upregulated under RHY (**Figure 59C, supplemental material**)[583].

Once inside the mitochondria, pyruvate has two main ways it can enter the TCA cycle. The first is via dehydrogenation to acetyl-CoA through the pyruvate dehydrogenase complex. Aside from PDHA1 which has stable expression across conditions, all genes associated with the pyruvate dehydrogenase complex are down-regulated under hypoxic conditions (PDHB, PDHX, DLAT, and DLD) (**Figure 19**). In order to function, the pyruvate dehydrogenase complex requires sufficient expression of the various components including the non-enzymatic PDHX and DLAT/DLD. Likely even the catalytic subunit of the complex, E1 which is a heterodimer of PDHA1 and PDHB, is functionally down-regulated due to the reduction in PDHB levels[582]. Additionally, PDHA1 protein activity can be inhibited by phosphorylation through Pyruvate Dehydrogenase Kinases (PDK1/2/3)[584,585]. All of these kinases are upregulated under hypoxic conditions, in line with the literature (**Figure 19**)[585]. Given the drop in expression of these various components, it can be assumed that metabolic flux through this complex is also reduced. The second access point to the TCA cycle is through carboxylation and the conversion to oxaloacetate via pyruvate carboxylase (PCX) activity[582,586]. PCX is significantly down-regulated under RHY, suggesting pyruvate is not being used to regenerate oxaloacetate (**Figure 19**).

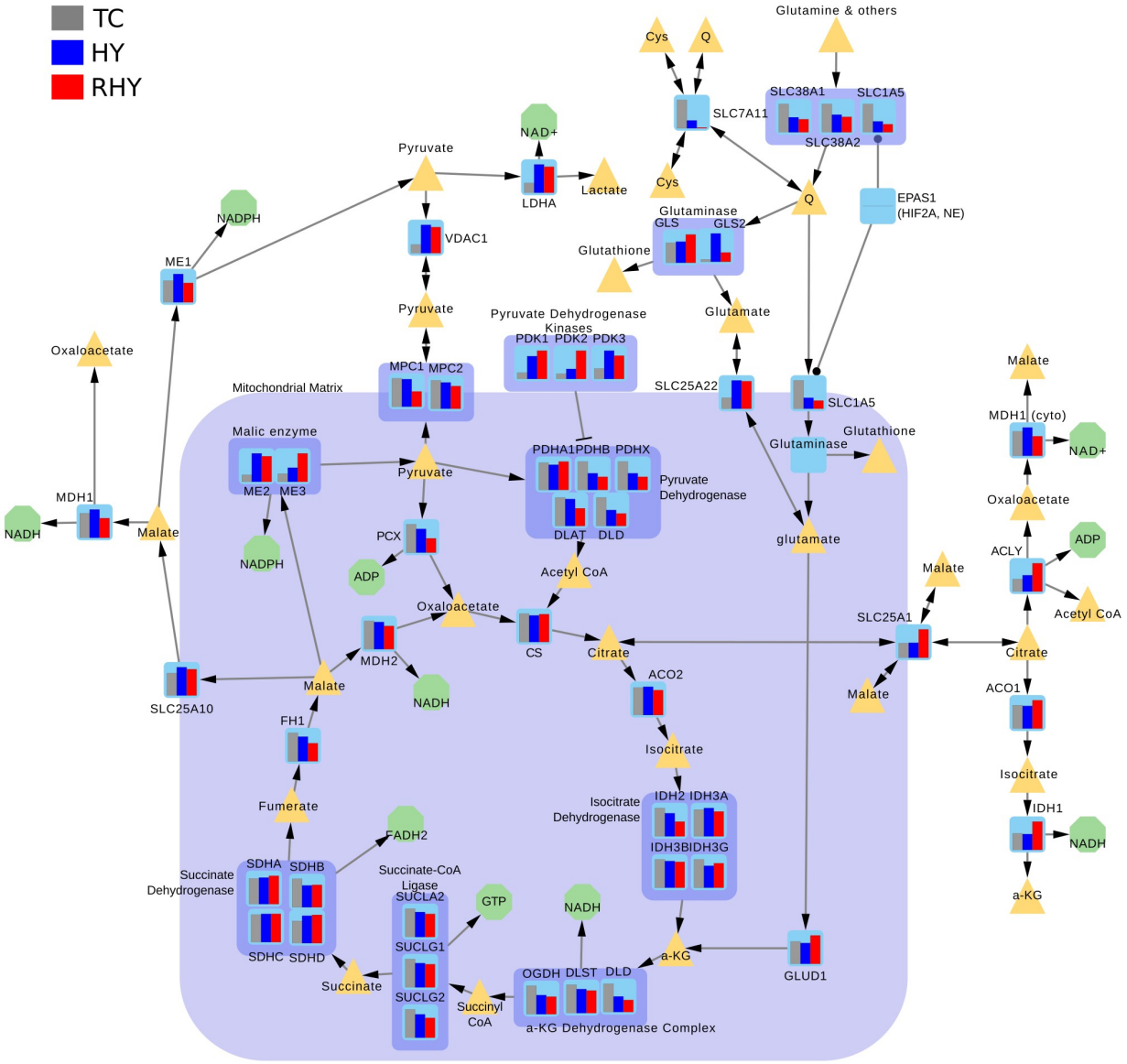

Figure 19: MHY TCA & glutamine metabolism network

Genes are blue squares, with relative MHY RNAseq expression values mapped to grey (TC), blue (HY), and red (RHY). Expression is averaged DESeq2 normalized reads. Genes with similar function, or generate components of a complex, are grouped together. Intermediate metabolites are symbolized as yellow triangles. Energy carrying molecules are symbolized as green octagons. Large blue square denotes mitochondrial matrix localization.

Regarding mitochondrial genes associated with the TCA cycle, very few genes are upregulated. The majority of these genes are either stably expressed across all conditions, or display a drop in expression under hypoxic conditions (**Figure 19**). The α-ketoglutarate Dehydrogenase Complex (OGDH, DLST, and DLD) and Fumarate hydratase (FH1) are both down-regulated under hypoxic conditions (**Figure 19**). With the loss of FH1 activity, there may be a build-up of fumarate, which can act as a competitive inhibitor for oxoglutarate-dependant hydroxylases, including the prolyl-hydroxylases responsible for HIF1α regulation under normoxia[587].

The notable exception to this pattern of TCA cycle gene expression stability or down-regulation is the RHY specific upregulation of isocitrate dehydrogenase 1 (IDH1) (**Figure 19**). This isoform of IDH1 is localized to the cytoplasm, and severs the same enzymatic function as its mitochondrialy localized isoforms[588]. In addition SLC25A1, a mitochondrial citrate/malate antiporter, is also specifically upregulated under RHY, presumably increasing cytosolic citrate levels (**Figure 19**). This citrate can then be converted to isocitrate by ACO1, which is stably expressed across all conditions, and can be converted to α-ketoglutarate and NADH by IDH1. Alternatively, cytosolic citrate can be converted to oxaloacetate and acytl-CoA by ATP Citrate Lyase (ACLY) at the cost of an ATP hydrolysis to ADP. Interestingly, ACLY is significantly upregulated under RHY conditions. The oxaloacetate can then be converted to malate by MDH1, generating a NAD+, which can be used for further IDH1 generation of α-ketoglutarate, or used to increase glycolytic capacity. The generated cytosolic malate can then be imported into the mitochondrial matrix via SLC25A1, and used by malic enzymes to generate NADPH and pyruvate (**Figure 19**). This non-canonical TCA cycle has been recently reported in murine embryonic stem cells[589]. It allows for the generation of glucose-independent acetyl-CoA generation, which can be used for de novo lipogenesis[588].

3.3.2.8 RHY induces cholesterol synthesis & regulation

Based on GO term analyses showing enrichment for cholesterol biosynthesis under RHY conditions, a deeper investigation of the pathway was performed. Specifically in the RHY condition there is upregulation of all members of cholesterol synthesis: HMG-CoA synthase (HMGCS), HMG-CoA reductase (HMGCR), squalene synthase (FDFT1), and 7-Dehydrocholesterol

reductase (DHCR7) (**Figure 20A**). Additionally, there is RHY specific upregulation for several of the genes involved in cholesterol biosynthesis regulation, namely SREBF2 the canonical transcription factor responsible for regulating cholesterol biosynthesis, and the various accessory proteins required for its function (SCAP, MBTPS1) (**Figure 20B**)[590,591]. An exception to this is MBTPS2, a protease involved in the final step before SREBF2 is released into the nucleus. MBTPS2 shows moderate down-regulation under RHY, although it is still expressed (**Figure 20B**).

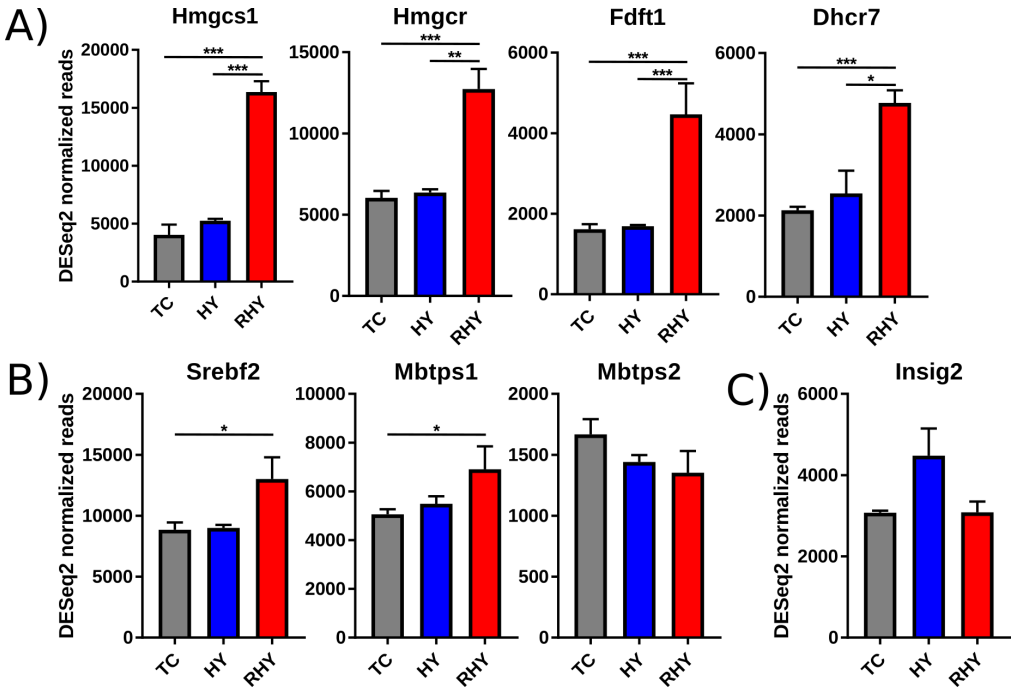

Figure 20: MHY cholesterol biosynthesis & regulation signature

A-C) MHY gene expression histograms. Expression is normalized DEseq2 reads, error bars are SD. * denotes relative significance as calculated by DESeq2 Benjamini-Hochberg adjusted p-value (padj). *: padj < 0.05, **: padj < 0.01, ***: padj < 0.001. **A**) Expression of cholesterol synthesizing genes. **B**) Expression of cholesterol synthesis activation regulators. **C**) Expression of cholesterol synthesis negative regulator.

Cholesterol biosynthesis under hypoxic conditions is unexpected, as cholesterol synthesis is a O_2-dependant pathway, and has shown to be inhibited under hypoxia. This was shown by Hwang et al, where they demonstrated that HIF1 inhibited HMG-CoA reductase (HMGCR) by upregulating Insulin-induced gene 2 (INSIG2)[592,593]. Our results do not contradict

this finding, as we also observe INSIG2 upregulation under HY conditions, however this upregulation disappears under RHY conditions (**Figure 20C**). Tissues-wide cholesterol synthesis upregulation has been reported under fasting conditions[594]. SREBF2 upregulation has been shown to occur under conditions of acidic extracellular pH and nutrient-starvation, in-line with our culture conditions[595].

3.3.2.9 RHY reprograms Zinc homeostasis to promote elevated intracellular levels

Across the MHY conditions, there are several zinc homeostasis genes that are differentially expressed (**Figure 21**). Broadly, zinc homeostasis is regulated through mono-directional channels and zinc chelators. Zinc transporters can be classified based on their orientation relative to the cytosol: ZIPs, encoded by the *SLC39A* gene family, increase cytosolic zinc levels; while ZNTs, encoded by the *SLC30A* family lower it. Both ZIPs and ZNTs have specific subcellular localizations. Cytosolic zinc levels are chelated and buffered by metallothionein proteins, encoded by the *MT1/2* genes[596]. Among the metallothioneins, there is a moderate upregulation of MT1 under HY, which becomes flipped to a moderate down-regulation under RHY conditions. MT2 becomes specifically down-regulated under RHY. In combination, this suggests that RHY culture conditions promote a reduction of Zn ion intracellular buffering.

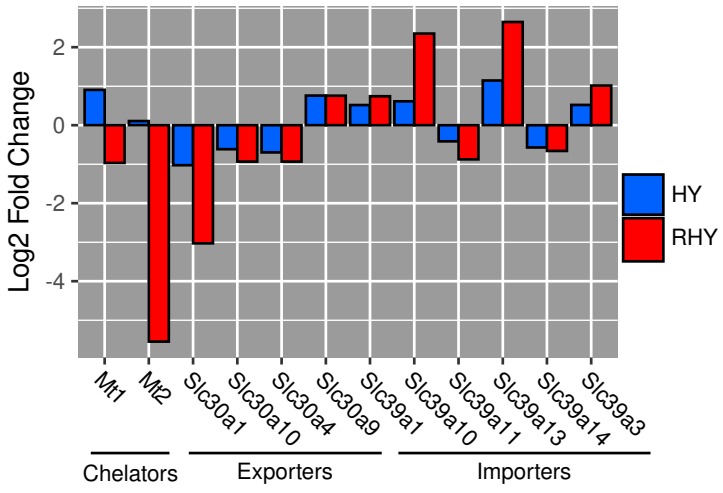

Figure 21: MHY Zinc homeostasis signature

Bar graph of zinc homeostasis genes relative fold-change compared to TC conditions. Fold-changes were calculated using average DESeq2 expression values for genes.

Among the cellular Zinc transporters, RHY tended to induce stronger differential expression compare to HY. ZNT1 (SLC30A1) is down-regulated two-fold under HY, with this down-regulation extending to six-fold down-regulation under RHY. ZNT1 is a zinc transporter which lowers intracellular zinc levels through a Zn^{+}/H^{+} anti-porting mechanism. While ZNT1 is believed to function primarily at the plasma membrane, there are studies suggesting alternative subcellular localizations[597,598]. In contrast to this down-regulation, ZIP10/13 (SLC39A10/13) are both significantly upregulated under RHY conditions. ZIP10 imports Zinc from the extracellular space into the cytosol, and has been linked to invasive behaviour in patient derived breast cancer and breast cancer cell lines[599]. ZIP13 liberates intracellular zinc stores from vesicular stores or the ER[600](Panes et al, manuscript in preparation). ZIP13 expression has also been linked to increasing the metastatic potential of human ovarian cancer cells[601].

3.3.3 MHY DEmiR analyses show RHY induces an aggressive miR signature

Given the known role of miRNA regulation in hypoxic adaptation[602], smallRNAseq was used to determine the differential regulation of miRNAs caused by MHY treatments. Due to the significant amount of DEGs in this dataset, an investigation into the microRNA biogenesis pathway was performed to verify that microRNAs would be functional under our conditions (**Figure 22A**). All genes involved in miR biogenesis were expressed, with no drops in expression across the hypoxic conditions. Of these genes *DGCR8*, a component of the pri-miRNA processing complex[603], as significantly upregulated under RHY conditions. Of the AGO proteins, AGO1 was upregulated two-fold under RHY conditions, while AGO2 and AGO3 maintained stable expression. AGO4 is not expressed in this cell line. Unlike AGO2, AGO1 does not contain slicer activity[384].

This increase in microRNA biogenesis genes in RHY conditions matches the overall increase in DEmiR patterns in comparison to TC conditions (**Figure 22B**). Relative to TC , HY only induced the expression of 14 miRs (**Table 16**). Curiously, HY did not down-regulate any miRs. RHY doubled the amount of up-regulated miRs, and down-regulated 18. These DEmiR patterns are more distinct than the DEGs, as there is less overlap between HY and RHY TC relative DEmiRs than DEGs (**Figure 22B**). The classical hypoxia-induced miR, miR-210, is significantly and relatively equally upregulated under either hypoxic condition (**Figure 22C**). Despite this

relatively equal upregulation, there are still large miR expression pattern differences between the hypoxic conditions (**Figure 22D**). To determine overall patterns in miR regulation, DEmiRs were clustered hierarchically (**Figure 22D**). The original heatmap was broken into ten clusters, as we found that number to generate consistent patterns of regulation (**Figure 60A, supplemental material**). An additional heatmap was generated using DESeq2 normalized readcounts of experimental replicates. It was decided to proceed with averaged DESeq2 normalized readcounts for visual clarity. The replicate-based heatmap can be found in the supplemental material section of this thesis (**Figure 60B, supplemental material**).

Table 16: Overview of MHY differentially expressed miRs

Comparison	Total	Up-regulated	Down-regulated
HY vs TC	14	14	0
RHY vs TC	49	31	18
RHY vs HY	22	5	17

In order to determine how common or unique our DEmiR signature was, we compared it to literature established hypoxia-induced miR signatures. There are two main signatures in the literature: One from Kulshreshtha et al in 2008 with 54 DEmiRs, and another from Shen et al in 2013 with 64 DEmiRs[451,452]. Of these two signatures, There is minimal overlap with our DEmiR signature. Only miR-210 is common among all three sets of miRs, and a total of nine of miRs in the MHY DEmiR list are found in either of these signatures.

Four miRs in the Kulshreshtha et al list are found in the MHY DEmiR list: miR-7, miR-181c, miR-192, and let-7c (**Figure 22D, blue miR label**). The miR-7, despite originally being described as a hypoxia-induced miR, has more recently been shown to be down-regulated by hypoxia in multiple cancer settings. In breast cancer, miR-7 has been shown to suppress metastatsis[451,604,605]. In a human gastric cancer model, miR-7 has been correlated with suppressing glycolysis through repressing LHDA protein production, and sensitizing tumour cells to cisplatin[606]. This promotion of glycolysis and increase in chemoresistance are hallmarks of the hypoxia response, and matches the MHY data precisely. Another study by Seong et al showed

that miR-7 is significantly down-regulated in HeLa cells after 24hrs incubation at 1%O_2[607]. They found that miR-7 was targeting REDD1 (aka DDIT4), a mTOR inhibitor. Therefore in this context, miR-7 was acting as a hypoxia-specific throttle to the mTOR pathway through DDIT4. In the MHY dataset, DDIT4 is also upregulated under hypoxic conditions, in line with this study (data not shown). Based on these and other studies, miR-7 supplementation therapy is being considered for a variety of cancer types[493,608].

miR-181c expression has been associated with chemoresistance and poor-prognosis in high-grade serious ovarian cancer, as well as ROS resistance in cardiac and pulmonary disease[609,610]. miR-192 has been shown to be upregulated in hypoxic melanoma cells, and that this miR is transferred to dendritic cells and cytotoxic T cells via connexin-43 gap junction channels, which are expressed in the MHY dataset (**Figure 59D, supplemental material**)[611]. This transferred miR-192 resulted in decreased cytotoxic ability in the T cells, providing a new mechanism of action for hypoxia-induced immunoresistance. In the MHY data, miR-192 is specifically upregulated RHY conditions. let-7c is established as a down-regulated hypoxia-associated miR by Kulshreshtha et al. Later studies have shown it to function as a tumour suppressive miR, inhibiting invasion, maintaining chemosensitivity, and suppressing cancer growth[612–614]. In the MHY dataset, let-7c is specifically down-regulated under RHY conditions.

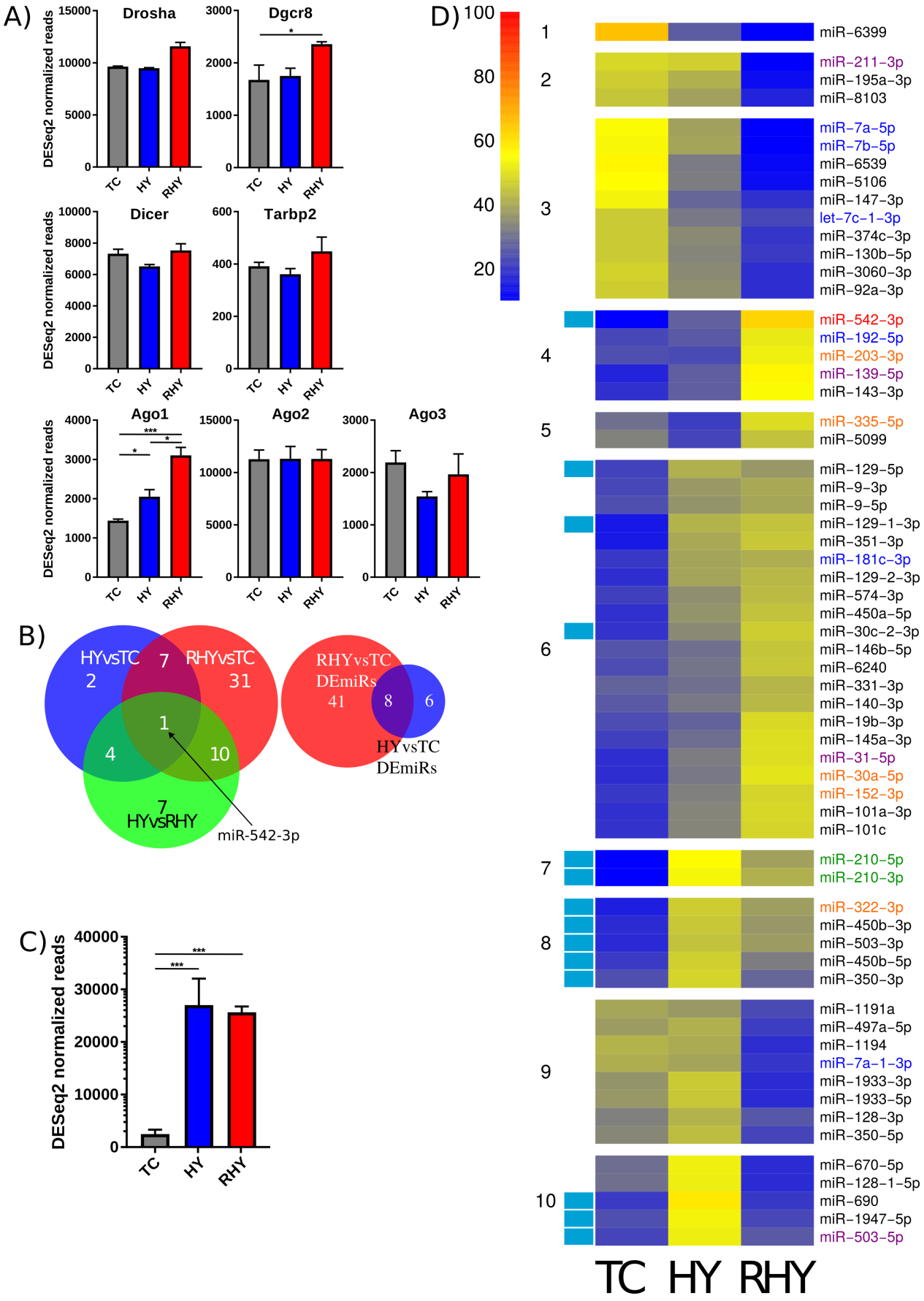
A)
Drosha
Dgcr8
Dicer
Tarbp2
Ago1
Ago2
Ago3
DESeq2 normalized reads
TC
HY
RHY
B)
HYvsTC
2
RHYvsTC
31
7
1
4
10
7
HYvsRHY
miR-542-3p
RHYvsTC
DEmiRs
41
8
6
HYvsTC
DEmiRs
C)
DESeq2 normalized reads
D)
100
80
60
40
20
1
miR-6399
2
miR-211-3p
miR-195a-3p
miR-8103
3
miR-7a-5p
miR-7b-5p
miR-6539
miR-5106
miR-147-3p
let-7c-1-3p
miR-374c-3p
miR-130b-5p
miR-3060-3p
miR-92a-3p
4
miR-542-3p
miR-192-5p
miR-203-3p
miR-139-5p
miR-143-3p
5
miR-335-5p
miR-5099
6
miR-129-5p
miR-9-3p
miR-9-5p
miR-129-1-3p
miR-351-3p
miR-181c-3p
miR-129-2-3p
miR-574-3p
miR-450a-5p
miR-30c-2-3p
miR-146b-5p
miR-6240
miR-331-3p
miR-140-3p
miR-19b-3p
miR-145a-3p
miR-31-5p
miR-30a-5p
miR-152-3p
miR-101a-3p
miR-101c
7
miR-210-5p
miR-210-3p
8
miR-322-3p
miR-450b-3p
miR-503-3p
miR-450b-5p
miR-350-3p
9
miR-1191a
miR-497a-5p
miR-1194
miR-7a-1-3p
miR-1933-3p
miR-1933-5p
miR-128-3p
miR-350-5p
10
miR-670-5p
miR-128-1-5p
miR-690
miR-1947-5p
miR-503-5p
TC HY RHY

Figure 22: MHY DEmiR signature

A) Expression of miR biogenesis & effector genes. Expression is normalized DEseq2 reads, error bars are SD. * denotes relative significance as calculated by DESeq2 Benjamini-Hochberg adjusted p-value (padj). *: padj < 0.05, **: padj < 0.01, ***: padj < 0.001. **B**) Venn diagrams of DEmiR comparisons. **C**) Expression of canonical hypoxia-induced miR, miR-210-3p. Expression values and significance are the same as in A. **D**) Hierarchically clustered heatmap of DEmiRs (padj < 0.05 and minimum expression >= 100 in any comparison). Blue bar is HYvsTC DEmiR list. Normalization method is percent contribution to total expression. Green: miR-210. Red: miR considered DEmiR in all conditions. Orange: Shen 2013 list of hypoxia-miRs. Blue: Kulshreshtha 2008 hypoxia-miRs. Purple: other miRs of interest. Cluster numbers are listed on the left.

Five miRs in the Shen et al list are found in the MHY DEmiR list: miR-152, miR-203, miR-30a, miR-322, and miR-335 (**Figure 22D, orange labelled miRs**). In human gastric cancer cells, miR-152 has been shown to suppress HLA-G expression, aiding in tumour cell escape[615,616]. In glioblastoma cells, miR-152 has been shown to be tumour suppressive by suppressing the expression of KLF4[617]. We contradict this data as we have positive correlation between miR-152 expression and KLF4 (**Figure 59E, supplemental material**). In human melanoma cell lines, miR-203 has been associated with increased proliferation, migration and tumorigenesis[618]. As miR-203 is specifically expressed under RHY conditions, which is when cells express the most EMT markers, we reinforce the claim that miR-203 can be used as a biomarker for early detection of metastasis. While often regarded as tumour suppressive, miR-30a-5p has been shown to confer cisplatin resistance in melanoma cells[619,620]. MiR-322 as well as miR-503, another RHY upregulated miR, are known to be upregulated under serum starvation conditions, explaining their exclusive upregulation under RHY[621]. MiR-322 has been shown to induce angiogenesis by down-regulating Cullin-2, which matches our RNAseq data (**Figure 59F, supplemental material**)[622]. MiR-322 and miR-503 have been suggested to promote EMT through multiple TGFβ-dependant mechanisms, including repression of TGFβR3 which is specifically down-regulated relative to hypoxia under RHY conditions (**Figure 59G, supplemental material**)[623–625]. MiR-335 has been shown to promote cell migration in glioblastoma cells, in a MMP14

dependant fashion[626]. MMP14 is significantly upregulated under RHY conditions (**Figure 59H, supplemental material**).

Besides previously published hypoxia-induced miR signatures, individual miRs have also been linked to hypoxia adaptation. MiR-211 has been reported as a negative regulator of metastasis in melanoma, as well as a negative regulator of HIF1A and PDK4[627,459]. It's loss of expression under RHY specifically aligns with the DEG analysis suggesting higher EMT potential in RHY exposed cells. Despite miR-211 loss of expression, HIF1α mRNA levels do drop under specifically RHY conditions, suggesting alternative HIF1α mRNA regulatory mechanisms (**Figure 59I, supplemental material**). MiR-31 is significantly upregulated under RHY conditions specifically. MiR-31 has been described as tumour suppressive in several cancer types for multiple cancer phenotypes[628]. However in colorectal cancers, miR-31 over-expression contributes to chemoresistance and promotes migration and invasion by inhibiting NUMB, which is down-regulated in RHY relative to HY (**Figure 59J, supplemental material**)[629,630].

3.4 Chapter summary

In this chapter, we have established a melanoma hypoxia-reporter cell line using B16 cells and a lentiviral HIF1α-eGFP fusion protein construct. Using this reporter cell line, we have shown that metabolic stress in conjunction with hypoxia results in elevated levels of HIF1α stabilization. Using deep transcriptome profiling, we have shown that this incubation protocol results in significantly more differentially expressed genes, and specifically more phenotypes involved in secondary *in vivo* hypoxic adaptations like angiogenesis and cellular motility.

Using this transcriptomic data, we have shown that each incubation protocol results in distinct metabolic changes. In glycolysis, metabolic stress results in the upregulation of the Rapoport-Luebering shunt, counter-intuitively reducing the amount of ATP generated by the glycolytic pathway. In branched chain amino acid metabolism, metabolic stress upregulates amino acid import, but throttles metabolism by down-regulating BCAT1 and mitochondrial import. The citric acid cycle was also significantly regulated, with metabolic stress resulting the specific upregulation of an alternative TCA cycle involving IDH1 and mitochondrial malic enzymes. Also unique to the metabolic stress condition was the upregulation of cholesterol

biosynthesis, perhaps due to the complete lack of extracellular cholesterol present. Zinc homeostasis was also differentially regulated under metabolic stress, suggesting cells were increasing zinc import, yet reducing zinc buffering.

MiRNAs were also significantly differentially regulated. While the classical hypoxia-induced miR-210 was equally upregulated across conditions, there were several other miRs differentially regulated that have not been reported in the literature. Mirroring the rRNA-depleted mRNAseq data, the metabolic stress condition regulated significantly more miRs than hypoxic incubation alone. Additionally, the metabolic stress condition regulated miRs in a fashion that has been shown to increase tumour agressivity in other contexts. This suggests that metabolic stress in conjunction with hypoxic incubation is more representative of *in vivo* hypoxia, and should be further used to simulate it.

Chapter 4 – Long term hypoxic culture induces EMT and a novel isoform of WT1

Jordan Quenneville[1,2], Albert Fengalhy[1], Margaux Taux[1,3], Kiersten Thomas[1,2], François Béliveau[4,5], Josée Hébert[4,5], Francois Major[1,6], Etienne Gagnon[1,3]

1. Institute for Research in Immunology and Cancer
2. Department of Molecular Biology, Université de Montréal
3. Department of Microbiology, Infectology, and Immunology, Faculty of Medicine, Université de Montréal
4. Banque de cellules leucémiques du Québec (BCLQ)
5. CIUSSS de l'Est-de-l'Île-de-Montréal, Hôpital Maisonneuve-Rosemont, Montréal QC
6. Department of Computer Science and Operations Research, Faculty of Arts and Sciences, Université de Montréal

4.1 Authors & contributions

Jordan Quenneville: Conception, planning, execution, and analysis of experiments and data including FACS, qPCR, Western Blots, ChIP-qPCR and ChIPseq sample preparation. Development and execution of LTHY incubation protocol. Maintenance of cell lines. Bioinformatic analyses including: generation of DEmiR and DEG analyses (heatmaps, histograms, GSEA), Kozak reliability prediction, transcription factor binding site analysis, GO term annotation bubbleplot generation, and analysis of Leucegene kmer files. Discovery of human and murine truncated WT1 isoforms and generation of splicing maps. Cloning of truncated WT1 expression plasmids and WT1 targeted shRNA plasmids.

Albert Feghaly: Processing of raw RNAseq data to DESeq2 tables of differentially expressed genes. Granted access to Leucegene kmer tables. Troubleshooting for kmer analysis code and python *km* package.

Margaux Taux: Acquired microscopy images to determine truncated WT1 subcellular localization.

Kiersten Thomas: Maintained ZR75 breast cancer cell line. Performed ZR75 long-term hypoxia incubation experiment. Performed and analyzed ZR75 qPCR experiment.

François Béliveau: Generated Leucegene overall survival curves.

Josée Hébert: Supervision of generated Leucegene overall survival curves.

Francois Major: Conception of project, analysis and interpretation of results, figure preparation.

Etienne Gagnon: Conception of project, analysis and interpretation of results, final figure preparation. Preparation, submission, and analysis of Mass Spectrometry data.

4.2 Executive Summary

During tumour progression *in vivo*, the natural development of the hypoxic microenvironment is much more gradual, and reaches anoxic levels over time. To simulate this, we developed a novel long-term hypoxic incubation protocol, which gradually lowers the oxygen level from 5% to 0.1% over the course of 18 days, which we termed the LTHY incubation protocol. While undergoing LTHY, we observed a morphological shift in the cells to a more mesenchymal state, suggesting we had triggered an epithelial to mesenchymal transition (EMT) through hypoxic incubation. To investigate this further, we performed deep transcriptome profiling at 5%, 1%, 0.5%, and 0.1% O_2.

Through analyses of both DEmiRs and DEGs, there were clear indications that LTHY incubation induced a non-canonical form of EMT. Analyses of potential EMT driving genes lead us to investigate WT1 expression under LTHY. To our surprise, our cells were expressing a novel isoform of WT1 derived from a unestablished transcription start site (TSS) close to the 3' end of the canonical *WT1* intron 5. We confirmed the existence of this isoform and characterized it. Despite missing the first half of the canonical gene, this isoform retained nuclear localization and DNA binding ability. By performing ChIPseq analyses on this WT1 isoform, we determined that it is binding to genes associated with transcription and cell adhesion, strengthening its candidacy as the driver of our observed EMT.

We next sought to identify expression of the human ortholog of this truncated WT1 in human cancer samples. Towards this end, we analyzed WT1 isoform expression in the Leucegene leukemia patient database using a kmer approach. This approach not only allowed us to identify leukemic samples expressing the ortholog in question, but also identified expression of a human isoform which may be more functional than the currently established human truncated WT1 isoform. Overall survival of patients expressing these truncated WT1 isoforms is inconclusive, likely due to their low frequency in the Leucegene dataset.

4.3 Results

4.3.1 Establishing the LTHY incubation protocol

When tumour hypoxia occurs *in vivo*, it does not occur as abruptly as our previous *in vitro* experiments. As the tumour grows, the microenvironment becomes hypoxic more gradually, as hypoxic onset is driven by cell division and local oxygen supply and demand[631]. Ineffective angiogenesis as a result of aberrant angiogenic signalling by tumour cells often leads to the generation of vasculature that is inefficient at supplying nutrients and oxygen to the surrounding cells[231,631]. This serves to lengthen the amount of time over which the hypoxic microenvironment develops, as these ineffective blood vessels cannot properly serve their function[632]. In addition to this gradual onset, tumour hypoxia can also reach much lower levels of oxygen than 1%, with tumour cores often reaching anoxia (**Figure 23A**). These cores are often necrotic, with a shell of viable tumour cells surviving in the extremely hypoxic environment[36].

We developed a novel *in vitro* hypoxic incubation protocol called long term hypoxia incubation or LTHY to simulate the gradual onset and deeper hypoxic aspects of *in vivo* hypoxia (**Figure 23B**). In our protocol, cells are brought from normoxia to 5% O_2. At 5% O_2, this is considered average for physoxia across tissues, or on the edge of what is considered hypoxic[15]. Cell are allowed to adapt to this environment for 24hrs, after which the oxygen is lowered to 1%. After 48 hours of incubation at 1%, oxygen is lowered again to 0.5% O_2 for 72 hours. After this incubation period, oxygen is lowered by 0.1% every 72 hours. Cells are passaged at every timepoint, and leftover cells are snap-frozen in QaiZol for RNAseq. Oxygen concentrations were automatically maintained by a computer controller, ensuring the cells were continually exposed to the correct oxygen concentration. Aliquots of cell media were allowed to equilibrate to the hypoxic atmosphere for at least 24 hours prior to use, ensuring that the cells were not exposed to higher oxygen concentrations through their media.

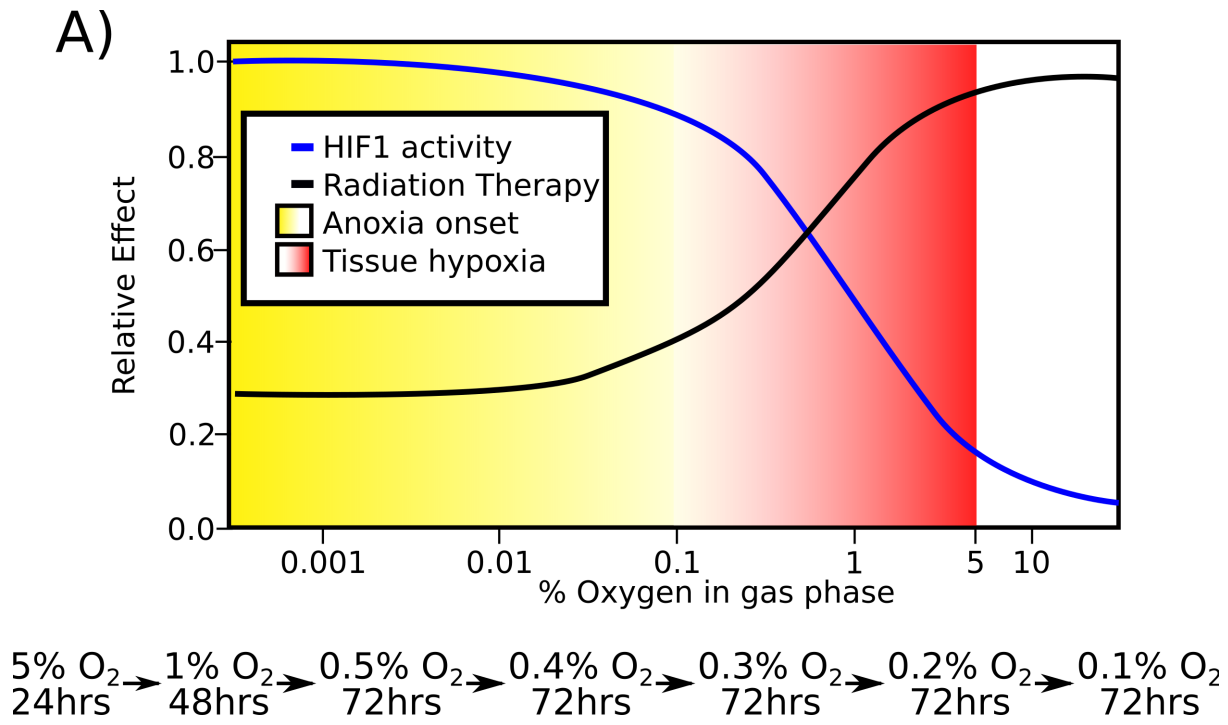

Figure 23: The range of hypoxia and the Long Term Hypoxic (LTHY) Incubation protocol

A) Adapted from REF[759]. The definitions of hypoxia and anoxia commonly used in the literature, their effect on HIF activity, and radiation therapy resistance. **B**) Definition of the Long-Term Hypoxia (LTHY) timecourse.

4.3.2 LTHY induces non-canonical EMT transcriptional program

4.3.2.1 LTHY induces morphological changes after exposure to specific hypoxic conditions

B16-HG cells underwent the LTHY protocol. Cells were maintained in screw-cap T-25 flasks, which allowed for the removal of the cells from the hypoxic incubator while maintaining hypoxia exposure. This also allowed for morphological observation through microscopy while preserving hypoxia exposure. Over the length of the timecourse, a morphological shift was observed in the B16-HG cells. Under normoxia as well as hypoxic incubation above 0.5% O_2, B16-HG cells had a epithelial-like cell morphology, with cells having a more cobblestone-like shape (**Figure 24A**). Once B16-HG cells were exposed to oxygen levels below 0.5%, their morphology shifted to a more mesenchymal shape (**Figure 24B**). Individual cells became more

elongated, with clear membrane ruffling opposite of a lamellipodium. Additionally, the cells were also growing in semi-attached, or completely detached clusters of cells (**Figure 24B**). This stark morphological shift suggested that the cells had undergone some kind of EMT over the course of their hypoxic incubation. To determine if these clusters of cells were dead, samples in suspension were taken during passaging and a Trypan Blue stain was performed. The staining, as well as the fact cells reattached to the flask after all passages confirmed cells were surviving in this environment and there was no notable cell death occurring.

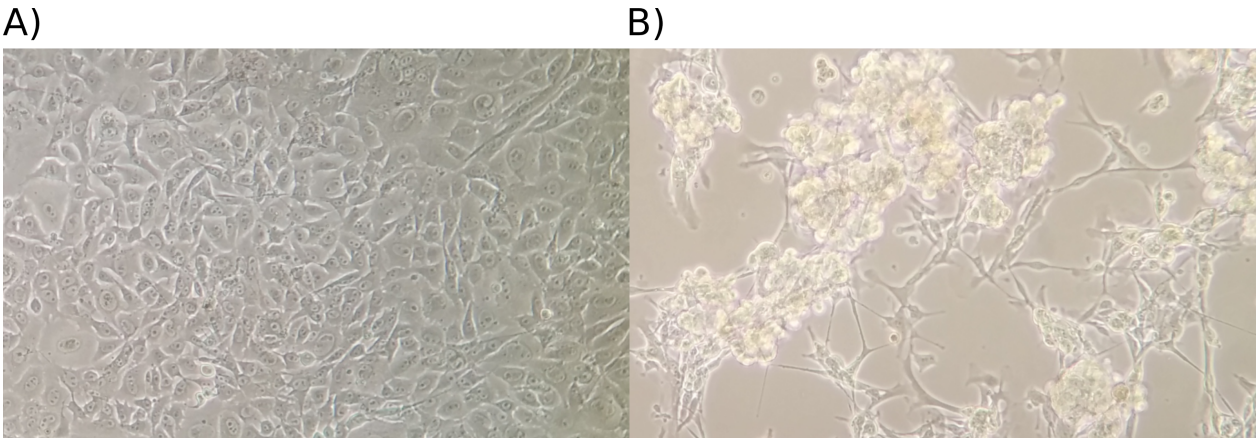

Figure 24: B16 morphology during LTHY incubation

A) B16-HG cell morphology under normal tissue culture conditions. **B**) B16-HG cell morphology after reaching the 0.3% O_2 timepoint of the LTHY protocol.

4.3.2.2 LTHY RNAseq clustering

To determine what transcriptional changes were occurring to induce this morphological shift, rRNA-depleted mRNAseq and smallRNAseq was performed after incubation at 5%, 1%, 0.5%, and 0.1% O_2 over the course of the LTHY protocol using serial biological duplicates. Sample clustering for both datasets was not ideal. For the rRNA-depleted mRNAseq data, the replicates at the early timepoints (5% and 1% O2) cluster well (**Figure 25A**). The proximity between the timepoints on the PCA plot suggests subtle changes have occurred in the transcriptome. It is interesting to note that the oxygen levels of the samples matches the directionality of PC1, with higher oxygen samples on the left, and lower oxygen samples on the

right. At the later timepoints (0.5% and 0.1% O2), replicates become more distant. As a result, unbiased clustering of replicates results in only the two endpoints of the experiment clustering properly (data not shown). Similar replicate issues are present in the smallRNAseq data (**Figure 25B**). Unbiased clustering of replicates only properly clustered the 0.1% O2 samples (data not shown). It is interesting that replicates follow a "V" shaped path across the timecourse. This pattern is present for both replicates, with n2 shifted to the right and down slightly. Due to these clustering issues, we opted to only consider the DEGs and DEmiRs comparing the 5% and 0.1% O_2 conditions, and examined the expression at the other timepoints post-hoc.

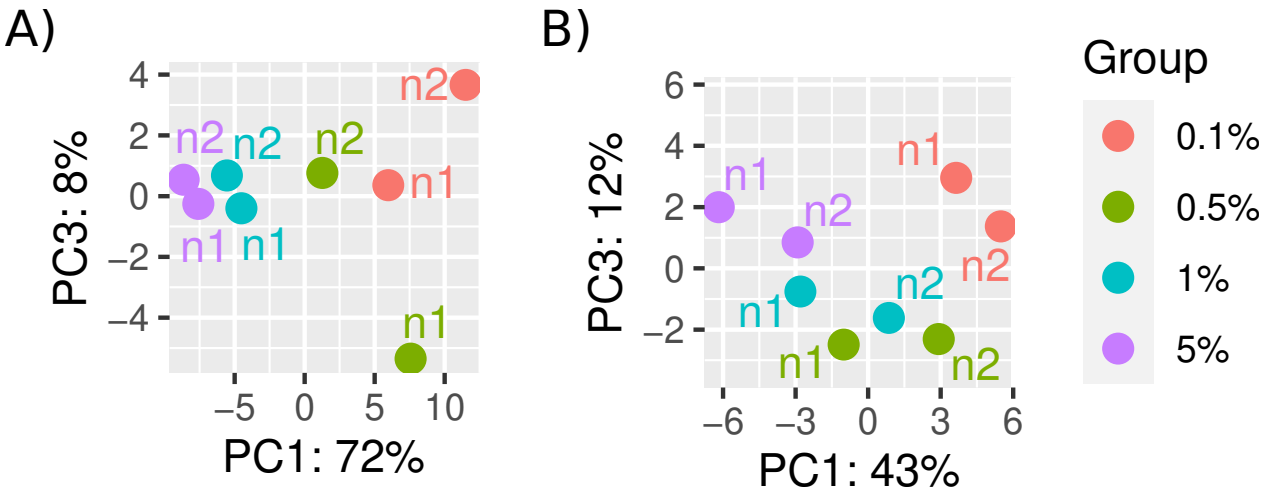

Figure 25: LTHY RNAseq PCA analysis

A&B) PCA analyses of LTHY rRNA-depleted mRNAseq data (A) and smallRNAseq data (B)

4.3.2.3 LTHY smallRNAseq confirms upregulation of hypoxia/EMT associated miRNAs

Given the role of miRNAs in EMT, we investigated miR biogenesis gene expression and the groups of differentially expressed miRNAs across the four hypoxic conditions analyzed. Genes involved in miR biogenesis were stably expressed from 5-0.1% O_2, with minor upregulation of Tarbp2 at 0.1% O_2, suggesting that miR functionality is maintained under these conditions (**Figure 26A**). We next looked at the classical EMT-inhibiting miR cluster, the miR-200 family[633]. Of the family members (miR-200a,b,c, miR-141, miR-429), only miR-200b-3p was expressed above our expression threshold of 100 DESeq2 normalized reads. However it was not significantly differentially regulated at 0.5% O_2, when our EMT phenotype first appeared (**Figure**

26C). These results suggests that our EMT is not being driven by a loss of miR-200 family expression, due to the low or lack of expression of the family members.

Next, we examined the classic hypoxia-induced microRNA, miR-210, to determine if our EMT was the result of differentially expressed hypoxia miRs. To our surprise, the fold-induction was between the 5% and 0.1% O_2 samples was not as strong as in the MHY dataset. By re-running DESeq2 using both the MHY and LTHY smallRNAseq read count tables, we were able to evaluate miR-210 expression levels across both experiments (**Figure 26B**). This revealed that at 5% O_2, there was already a six-fold upregulation of miR-210, explaining the reduced dynamics seen in the LTHY dataset. When compared to the miR-210 expression values in the MHY dataset it is clear that miR-210 begins upregulation at 5% O_2 and once cells are exposed to 1% O_2, miR-210 levels stabilize. Given there was no significant upregulation at 0.5% O_2, it is unlikely that miR-210 is driving the observed EMT.

To obtain a more global view of miR regulation across the LTHY dataset, differentially expressed miRNAs were hierarchically clustered as previously described, using the list of DESeq2 calculated DEmiRs between 5% and 0.1% O_2 as input (**Figure 26D**). Due to replicate variability, we chose to cluster averaged expression values between replicates, and removed the clustering dendrogram for visual clarity. A copy of the original heatmap with dendrogram, as well as a replicate-based clustered heatmap, can be found in the supplemental material of this thesis (**Figure 61, supplemental material**). To investigate which miRs could be involved in the morphological changes we observed while performing the timecourse, we focused on miRs which dramatically change expression at the 0.5% O_2 condition. Our list of DEmiRs were hierarchically grouped and broken into five clusters, with clusters 3, 4, and 5 containing miRs upregulated as oxygen decreases. These clusters became our initial focus for investigation.

Cluster 3 represents miRs that are significantly upregulated at 0.1% O_2. The top upregulated miR in this cluster is miR-125b-1-3p, which has been linked to increased metastatic potential in colorectal cells[634]. Other miRs in this cluster have been shown to inhibit EMT-like phenotypes, such as miR-34b and miR-100[635–638]. The remaining miRs, miR-34c and miR-199a, have conflicting evidence in support or against their role in EMT. There is evidence of a negative

correlation between miR-34c expression and invasion in colorectal, oesteosarcoma, and hepatocellular cancers[639–641]. However there is correlation of miR-34c expression with proliferation and migration in bladder cancer, as well as evidence that the anti-invasive action of miR-34c occurs through the Notch signalling pathway[642,643]. High miR-199a expression is associated with increased tumour invasion and metastasis in gastric and colorectal cancers[644–646]. However there is a body of literature suggesting miR-199a as having anti-tumour effects, through the targeting of HIF1α, ZEB1, and SOX4[647–650].

Cluster 4 contains miRNAs whose expression dramatically increased once cells were exposed to 0.5% O_2, and were therefore of highest interest. This cluster contained some of the classical miRNAs induced during an EMT: miR-221 and miR-222, suggesting the cells had undergone an EMT at the 0.5% O_2 timepoint[478,651]. The expression of miR-221/22 has been shown to be regulated by the transcription factor ZBTB16 (also known as PLZF) in melanoma cells. However the B16-HG cell line lacks expression of this gene, suggesting miR-221/222 expression is regulated by an alternative mechanism[652]. In addition, miR-21a, another EMT driving miR, was found in the same cluster[653,654]. MiR-155, also found in the same cluster, has been linked to hypoxia-induced EMT in multiple studies[655,656]. MiR-3695 is a murine specific microRNA, with no links to hypoxia adaptation or EMT.

Cluster 5 contains miRs which are gradually upregulated as hypoxic conditions deepen. Several of these miRs have been linked to EMT in different cancer contexts. In a murine model of breast cancer, miRs-143/145 were found to enhance migration and invasion[657]. In Wilms tumour cells and colorectal cancer cells, miR-190b upregulation promoted migration and invasion[658,659]. The miR-27 family (miR-27a/b) is also found in this cluster, and has been associated with EMT in multiple cancer types as well as chemoresistance[457,660–663].

Together, the LTHY smallRNAseq analyses reaffirms the possible onset of LTHY-induced EMT. Several of the miRs highly expressed at 0.5% and/or 0.1% O_2 are known EMT-driving miRs. Several of the miRs downregulated at 0.5% or 0.1% O_2 are known EMT-inhibiting miRs. Overall, the majority of the DEmiR signatures are associated with the promotion of an EMT-like state.

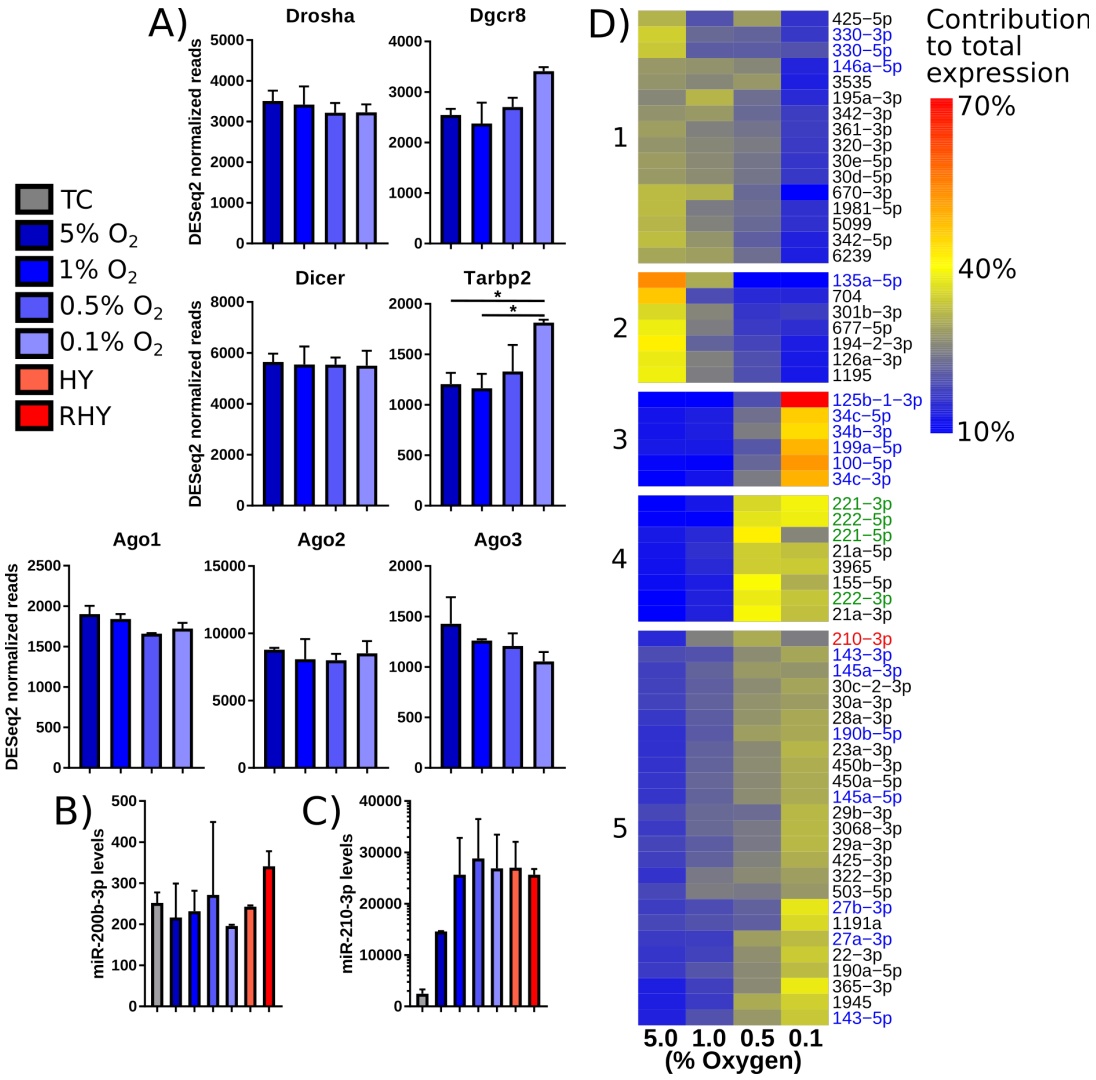

Figure 26: LTHY DEmiR signature

A) LTHY expression of miR biogenesis genes. Values are DESeq2 normalized reads, error bars are SD. * denotes relative significance as calculated by DESeq2 Benjamini-Hochberg adjusted p-value (padj). *: padj < 0.05, **: padj < 0.01, ***: padj < 0.001. **B**) LTHY expression of miR-200b. Values are DESeq2 normalized reads. **C**) LTHY expression of miR-210-3p. Values are DESeq2 normalized reads. **D**) Hierarchically clustered heatmap of LTHY DEmiRs. Cluster numbers are annotated to the left of clusters. DEmiR defined as having a DESeq2 padj < 0.05 in the 5% vs 0.1% O_2 comparison and a minimum expression of 100 normalized DEseq2 reads in any condition. miR expression normalized to contribution of total expression. Red highlights the classical hypoxia miR-210. Green highlights EMT associated miRs miR-221/222. Blue highlights other miRs of interest.

Another potential mechanism of action for miR-induced EMT is through the loss of expression of a EMT-inhibiting miR. Such miRs would be in clusters 1 or 2 of the heatmap. Cluster 2 contains miRs with a more abrupt change in expression at 0.5% O_2. In cluster 2, we can see that miR-135a-5p is downregulated at 0.5% O_2 and below. There is a body of work suggesting that miR-135a-5p is an invasion and proliferation inhibiting miR[664–666]. Its down-regulation at 0.5% oxygen correlates well with our increase in EMT-like properties observed in the cells. Similar to miR-135a-5p, miRs-330-3p/5p and miR-146a-5p in cluster 1 have been found to suppress EMT in lung, ovarian, liver, and colorectal cancers[667–672]. Similarly, miR-30e&d which are moderately down-regulated as the environment becomes more hypoxic, have been shown to inhibit EMT[673–677]. Their loss of expression may contribute to our observed EMT phenotype.

4.3.2.4 LTHY rRNA-depleted mRNAseq confirms upregulation of EMT effector program

To further examine the EMT induced by the LTHY protocol, DEG analyses were performed. DEGs were defined as genes with significant differential expression between the two endpoints of the experiment, 5% and 0.1% O_2, a list of 1448 genes. This list of DEGs was then used to generate a non-hierarchically clustered heatmap using the expression data from all four timepoints collected from the LTHY experiment. Gene expression was normalized using row Z-score, clustering was performed using the kmeans algorithm split into nine clusters using the elbow method[678,679]. Intra-cluster variability test and replicate-based LTHY DEG heatmap can be found in the supplemental material of this thesis (**Figure 62, supplemental material**). Due to replicate variability, we proceeded with averaged replicate expression values for DEG heatmap and GSEA analyses. GO term enrichment using DAVID was performed on each cluster to determine enriched phenotypes (**Figure 27A**). In stark contrast to the MHY clustered heatmap, there were much less GO terms associated with metabolic adaptation. The only metabolic adaptation phenotypes were found in clusters 1&2, which showed moderate downregulation across conditions. Given the observed EMT as well as the high amount of EMT-associated GO terms in the DEG signature, we focused on the EMT aspect of this dataset rather than the modest metabolic changes.

Notably, the genes in cluster six become markedly upregulated at 0.5% O_2 and 0.1% O_2. This cluster is enriched for genes associated with the response to hypoxia, angiogenesis, and

positive regulation of cell migration. In addition, one of the phenotypes in cluster three is negative regulation of cell adhesion. As these genes lose expression at 0.1% O_2, it may be that cell adhesion is increased under these conditions. In cluster seven, there is significant enrichment for genes associated with negative regulation of the TGFβ, SMAD, and BMP signalling pathways. This suggests that the EMT observed in this experiment is not being driven by these pathways, as they appear to be shutdown at oxygen levels below 0.5%. Nevertheless, clusters three and six both point towards the cells engaging in an EMT-like response.

In line with this upregulation of cell motility phenotypes, GSEA analyses comparing 5% and 0.1% O_2 showed that EMT was the top enriched hallmark associated geneset (**Figure 27B**). This enrichment was higher than the Hypoxia Hallmark geneset, suggesting that while cells were engaged in hypoxic adaptation as their microenvironment became more hypoxic, there was a stronger drive to become more motile and metastatic. This is in line with work done by Godet et al, whose group demonstrated that hypoxic adaption induced an increase in motility *in vivo*, using a reporter system which triggers at 0.5% O_2, the same oxygen level that we observe our EMT-like shift[174]. After sufficient time in a hypoxic environment, hypoxic tumor cells become more motile, as if to escape the hypoxic environment[174].

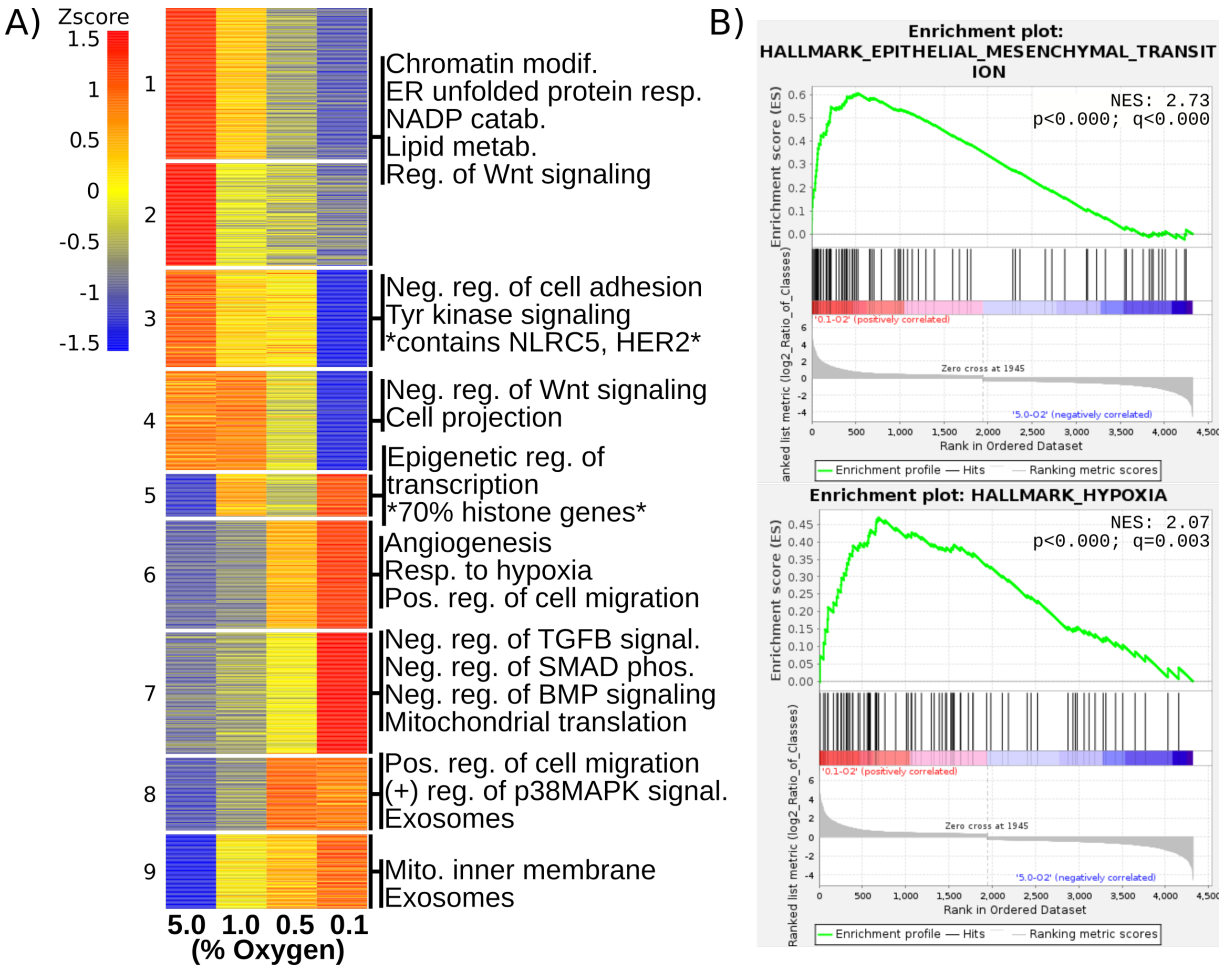

Figure 27: LTHY mRNA signatures

A) LTHY DEG k-means clustered heatmap. Gene expression normalized using row Z-score. Cluster number determined using elbow method. GO term enrichment was done using DAVID and cluster gene lists as input. Displayed GO terms are all significantly enriched ($p<0.05$). Cluster numbers are annotated to the left of clusters. **B**) GSEA of LTHY DEGS. Top: Enrichment plot for Hallmark of Epithelial -Mesenchymal Transition. Bottom: Enrichment plot for Hallmark of Hypoxia. Normalized Enrichment Scores (NES) and statistical significance is found in each plot, as calculated by GSEA.

Examining the genes associated with the EMT phenotypes in both the GO term and GSEA enrichments, there is a clear enrichment for EMT effector genes. However, most canonical EMT drivers such as SNAI1/2, ZEB1/2, GRHL2, EZH2, TWIST1/2, FOXC2 are either not expressed during the LTHY experiment or are not significantly differentially expressed at 0.5% O_2 or 0.1% O_2 (**Figure 28A**). This suggests that these classical EMT transcription factors are not driving the EMT observed during the LTHY experiment.

Nevertheless, the GSEA enrichment for EMT contained several genes which enact an EMT. Genes such as *S100A4*, *FN1*, *COL4A1*, *COL4A2*, *COL8A1* and *COL1A1* are all significantly upregulated at 0.5% O_2 and 0.1% O_2 (**Figure 28B**). Vimentin (VIM) was upregulated at 0.5% O_2 and 0.1% O_2, however this upregulation was not significant (lowest padj = 0.13). It may be that like miR-210, VIM is also already moderately upregulated. To test this, DESeq2 was ran again using both MHY and LTHY gene count tables. Under this analysis VIM is significantly upregulated (padj < 0.05) in all LTHY conditions in comparison to MHY TC samples. For this reason, we still consider VIM as part of our EMT signature.

The classical EMT effector marker N-cadherin (CDH2) was only moderately upregulated at 0.5% O_2 and 0.1% O_2. Interestingly, CDH2 was significantly upregulated under RHY conditions in the previous experiment by about 2.5 fold. This demonstrates that the B16-HG cell line can modulate CDH2 and suggests that CDH2 expression under hypoxia is driven more by nutrient starvation than actual oxygen levels. Loss of expression of E-cadherin (CDH1) could not be determined as its expression is not significant in the B16-HG cell line.

Despite having a EMT-like effector signature, we were still lacking a potential driver for this signature. Therefore, an investigation into non-canonical EMT drivers was launched. We started by collecting the list of genes associated with the GSEA enrichment for EMT as well as the list of genes associated with the GO term “positive regulation of cell migration”. From this list, several possible EMT drivers emerged: HRAS, CPEB1, SPHK1, HBEGF, THBS, CYR61, and WT1 (**Figure 28C**).

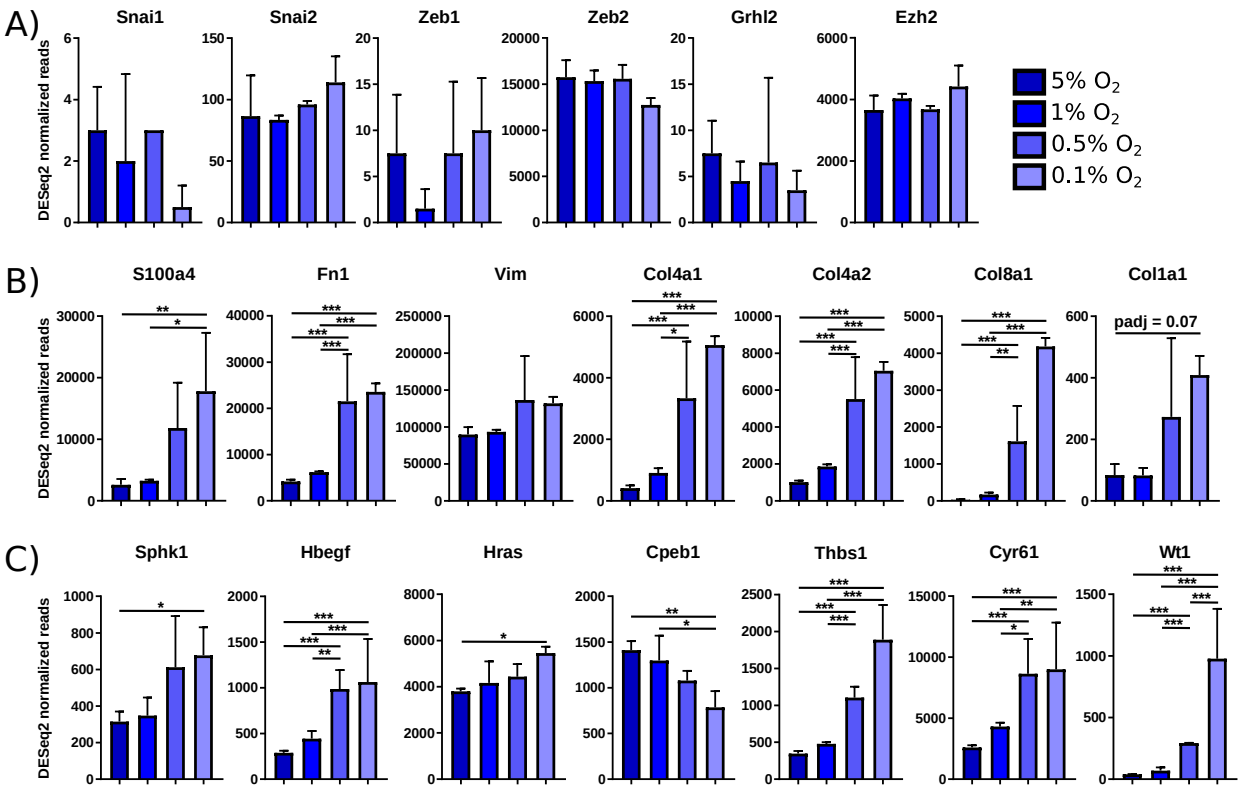

Figure 28: LTHY EMT driver and effector expression signatures

A) Expressions of EMT driving genes. Data not shown for genes with zero expression **B**) Expressions of EMT effector genes. **C**) Expressions of potential EMT driving genes. **A-C**) Values are DESeq2 normalized reads, error bars are SD. * denotes relative significance as calculated by DESeq2 Benjamini-Hochberg adjusted p-value (padj). *: padj < 0.05, **: padj < 0.01, ***: padj < 0.001.

HRAS has been shown as a driver of EMT in tumour settings, and this EMT promoting activity can be induced by an increase in HRAS expression[680,681]. However based on our RNAseq data, we did not see a distinct upregulation of HRAS at 0.5% O_2, when the other EMT markers occur. This suggests that HRAS it is not the driver of this EMT program. CPEB1 down-regulation has been shown to induce EMT in breast cancer[682]. However the down-regulation of CPEB1 is only significant at 0.1% O_2, well after the EMT-like morphological changes had occurred, suggesting that it is not the primary driver in this setting. In breast cancer, SPHK1 has been shown to drive the expression of EMT effector genes as well as stem cell markers[683]. SPHK1

expression did significantly increase and stabilize at 0.5% O_2, although the fold-induction is relatively low at about 3 fold. For this reason, other genes were prioritized for investigation. For similar reasons investigation into Hbegf was put on hold, despite having been described as a autocrinic EMT driver in oral squamous cell carcinoma [684]. THBS1 has been shown to be able to drive an EMT program in melanomas and has been correlated with increased aggressiveness in patient samples[685]. However it was found that THBS1 was not the primary driver of this EMT program, and was therefore de-prioritized for investigation. CYR61 has been shown to be a EMT driver in osteosarcoma through a TWIST-1 dependant pathway[686]. However TWIST-1 is not expressed in the B16-HG cell line. WT1, while originally described as a tumour suppressor, has been shown to be oncogenic in leukemia, as well as several solid tumours[687]. During embryogenesis, WT1 has a role in inducing EMT or MET, depending on the tissue context[688]. We believed that WT1 could therefore be the driver of our observed hypoxia-induced EMT. Additionally WT1 showed the highest fold-change relative to 5% O_2, and was therefore made the primary focus for investigation.

4.3.3 Characterization of novel hypoxia induced WT1 isoform and its regulation

4.3.3.1 qPCR confirms upregulation of WT1 under hypoxic conditions in multiple cell lines

From the LTHY RNAseq data, WT1 expression only passes a read cut-off of 100 normalized DESeq2 reads once the B16-HG cells are exposed to 0.5% O_2 or less. To more easily test WT1 expression levels in B16-HG cells, two sets of murine WT1 qPCR probes were designed and validated using the original LTHY RNA used for RNAseq. One set was specific to the 5' end of the WT1 mRNA, while the other bound to the 3'UTR. Only the 3'UTR specific probe set generated signal using the 0.1% O_2 B16-HG cDNA, and matched the RNAseq data well. To verify that hypoxic WT1 upregulation was not a quirk of the B16-HG cell line, WT1 regulation was monitored across the LTHY protocol in B16 wild-type cells (**Figure 29A**). The qPCR upregulation of WT1 in wild-type B16s correlated well with the B16-HG LTHY RNAseq data. This confirmed that WT1 upregulation is not an artifact of the generation of the B16-HG cell line.

To verify that the WT1 upregulation was not a murine specific phenomena, human melanoma cell lines (MEL1300 and SK23) and the human breast cancer cell line ZR75 were subjected to the LTHY protocol as well (**Figure 29B&C**). WT1 regulation was then probed for at different oxygen levels by qPCR. Two different qPCR probes sets were used for the human melanoma cell lines. The first bound within the first exon (WT1 5' probe), and the second spanning exons 7-9 (WT1 3' probe) (**Figure 29B**). In both cell lines, there was significant upregulation of WT1 at 0.5% and 0.1% O_2 relative to expression levels at 5% O_2. Both cell lines reduced WT1 expression at oxygen levels lower than 0.5%, in contrast to the B16-HG cell line. Curiously, MEL1300 cells maintained relatively high expression of the 3' end of the gene while expression of the 5' probe dropped, suggesting expression of alterative isoforms.

For the ZR75 LTHY incubation, cells were cultured either in the presence or absence of estrogen. This was in an attempt to better simulate the *in vivo* microenvironment, as growth factors such as estrogen may be depleted due to consumption by surrounding cells or due to cancer treatment. This estrogen modulation revealed two different WT1 regulatory patterns (**Figure 29C**). In the presence of estrogen, WT1 expression climbs dramatically as oxygen continues to decrease, reaching the highest fold change seen in any of our tested cells (60 fold induction). In the absence of estrogen, WT1 expression increases to approximately 20 fold relative to 20% O_2 at 0.5% O_2, then stabilizes afterwards. In summary, these qPCR experiments show that WT1 is consistently upregulated under hypoxic conditions in multiple cancer settings. (**Figure 29D**) The degree of upregulation appears to be cell line specific, and probe-specific regulation patterns suggest differential regulation of specific WT1 isoforms.

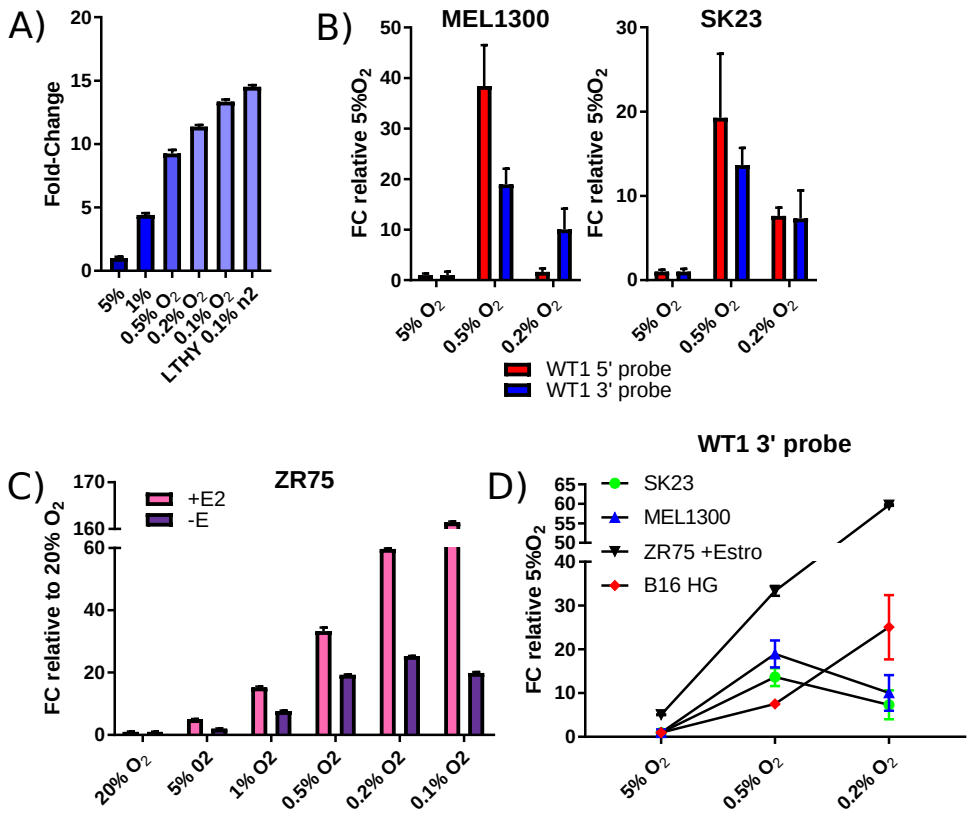

Figure 29: Validation of LTHY-induced WT1 upregulation

A) qPCR re-validation of WT1 upregulation. B16-WT cells underwent LTHY protocol as previously described. Used murine WT1 3' probe set for detection. Fold-change calculated using HPRT as house-keeper. **B**) WT1 upregulation in human melanoma cell lines MEL1300 and SK23. Hsa WT1 probset #1 used for WT1 5' probe, and Hsa probset #2 used for WT1 3' probe. Fold-changes calculated using 18s as housekeeper. **C**) WT1 regulation in ZR75 cells. Oxygen values are timepoints of LTHY incubation. Hsa WT1 probeset #3 used. Fold-changes calculated using RPLP0 as housekeeper. **D**) Summary of WT1 regulation under LTHY conditions **A,B,C**) Error bars are technical triplicate SD propagated using error propagation.

4.3.3.2 The LTHY protocol is necessary for optimal WT1 upregulation

Given that the original WT1 upregulation observed in the B16-HG RNAseq data occurred after six days of culture, we addressed if this upregulation was an effect of long-term hypoxic culture, and not the degree of hypoxia to which the cells were exposed to. This was tested by culturing B16-wt cells at 1% O_2 for two weeks, collecting cells at 2, 5, 8, 11 and 14 days of

culture, and examining WT1 regulation by qPCR (**Figure 30A**). At no point during the 2 week incubation did WT1 become significantly upregulated as previously shown. Therefore, the upregulation of WT1 seen in the LTHY RNAseq data is not due to total hypoxia exposure time. But rather, the degree of hypoxia serves to induce WT1.

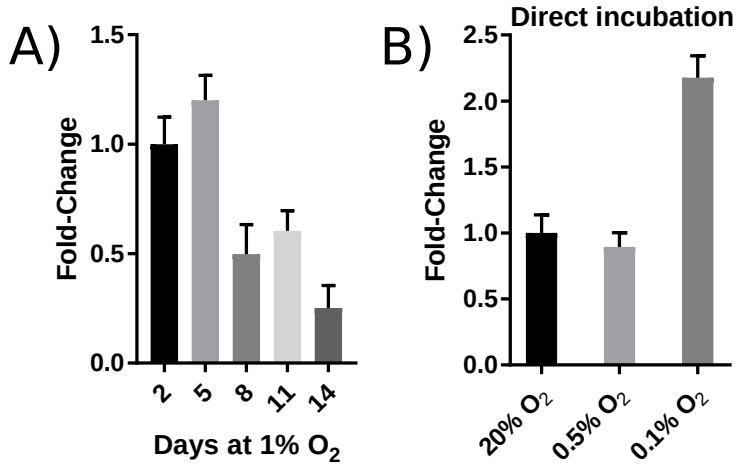

Figure 30: WT1 induction optimization

A) Relative WT1 levels following prolonged incubation at 1% O_2 in B16-WT cells. **B**) qPCR WT1 levels following 72hrs of direct incubation at indicated O_2 levels. **A,B**) Error bars are technical triplicate SD propagated using error propagation. Housekeeper used was 18s.

To determine if hypoxic upregulation of WT1 observed was strictly regulated by oxygen levels, B16-WT cells were transferred from 20% O_2 to 0.5% or 0.1% O_2 for 72 hours, and WT1 levels were measured by qPCR (**Figure 30B**). While WT1 was upregulated after direct incubation at 0.5% or 0.1% O_2, the degree of upregulation was not as high when cells underwent the LTHY protocol. This indicates that WT1 upregulation is optimal when cells have been habituated to a hypoxic environment. Therefore, optimal WT1 upregulation requires preincubation at milder hypoxic conditions, as well as the progression to deeper hypoxic conditions, as in the LTHY protocol. This combination of factors is more representative of the onset of hypoxia in an *in vivo* setting, suggesting that WT1 upregulation would occur to similar or greater degrees in the hypoxic tumor microenvironment.

4.3.3.3 RNAseq reads confirms four novel WT1 isoforms arising from a novel TSS

Due to the large amount of potential WT1 isoforms that could be expressed, as well as the hints of different isoform expression provided by the human melanoma cell line qPCR results, an isoform expression analysis was done. Using IGV, the B16-HG LTHY RNAseq reads were visualized across the *WT1* locus (**Figure 31A**). To our surprise, there was a complete lack of reads coverage across the first five exons of the gene, with coverage beginning at the 3' end of intron 5. Given the increase in read coverage at this site in combination with the lack of reads from the upstream exons, it is highly likely that this is the transcription start site (TSS) for this novel transcript. Due to the uncommon length of *Wt1* intron 5 at ~20kb, it is possible that this intron contains transcription factor binding sites (TFBSs) that drive this novel TSS. From the end of intron 5, read coverage continues across the locus to the end of the 3'UTR.

The exonic spikes in coverage suggest that this RNA does not remain as a primary transcript and is being spliced properly (**Figure 31A**). Potential splicing paths were investigated by generating a sashimi plot of the local genomic area (**Figure 31B**). From this, we could see there were two splicing events: one at the end of intron 5 or exon 6, joining at the beginning of exon 7, and another between exons 9 and 10. The exon 9-10 splicing event represented the KTS motif inclusion or exclusion[338,348]. The exon 6 splicing event was the expected splicing of exon 6 to exon 7. The intron 5 splicing event is completely novel in mice and occurs 58nt upstream of the 5' end of exon 6, removes exon 6 from the spliced transcript, and joins exon 7 at the correct location (**Figure 31C**).

From these two different splicing events four different mRNA isoforms arose (**Figure 31D**). All transcripts began transcription from within the end of intron 5. Primary transcripts could then be spliced to remove exon 6 and splice directly over to exon 6. These transcripts will be referred to as exon-6-skip in this thesis. Transcripts that do not undergo the intron 5 splice and instead include exon 6 and splice over to exon 7 will be referred to as exon-6 in this thesis. From these two 5' splicing events, either transcript could include or exclude the KTS motif. Collectively, these mRNAs resulting from these splicing events will be referred to as truncated WT1 isoforms (tWT1) for this thesis.

Based on known WT1 mRNA isoforms in both murine and human genomes, the particular isoform generated by the exon-6-skip splicing path has not been described (**Figure 31E&F**). Regarding known murine WT1 mRNA isoforms, there is only one known truncated isoform (**Figure 31E, bottom isoform**), referred to as murine Wt1-208 in the Ensembl database[347]. Wt1-208 is most semblant to the exon-6 tWT1 isoforms identified in the LTHY RNAseq data, as Wt1-208 retains exon 6 in the transcript. The major difference between the two is the lack of exon 10 (which contains the canonical stop codon) in Wt1-208. Wt1-208 is also listed as a retained intron and does not produce protein[347]. As for the exon-6-skip splicing path, there are no listed murine WT1 mRNA isoforms containing this particular splicing path. We therefore regard the exon-6-skip splicing path as novel in mice.

Regarding the known WT1 mRNA isoform in humans, there is much more variety of transcripts and splicing paths (**Figure 31F**). Focusing on the isoforms similar to our identified tWT1 transcripts (**Figure 31F**, bottom four transcripts), there is a transcript similar to the exon-6 tWT1 isoform, however there is no human truncated WT1 mRNA isoform which matches the exon-6-skip splicing path identified in the LTHY RNAseq data (**Figure 31D**, bottom two isoforms). Interestingly, many of these truncated WT1 mRNA isoforms identified in humans are labelled as protein coding, but with no CDS defined, indicating that very little is known about this class of truncated WT1 isoforms[347].

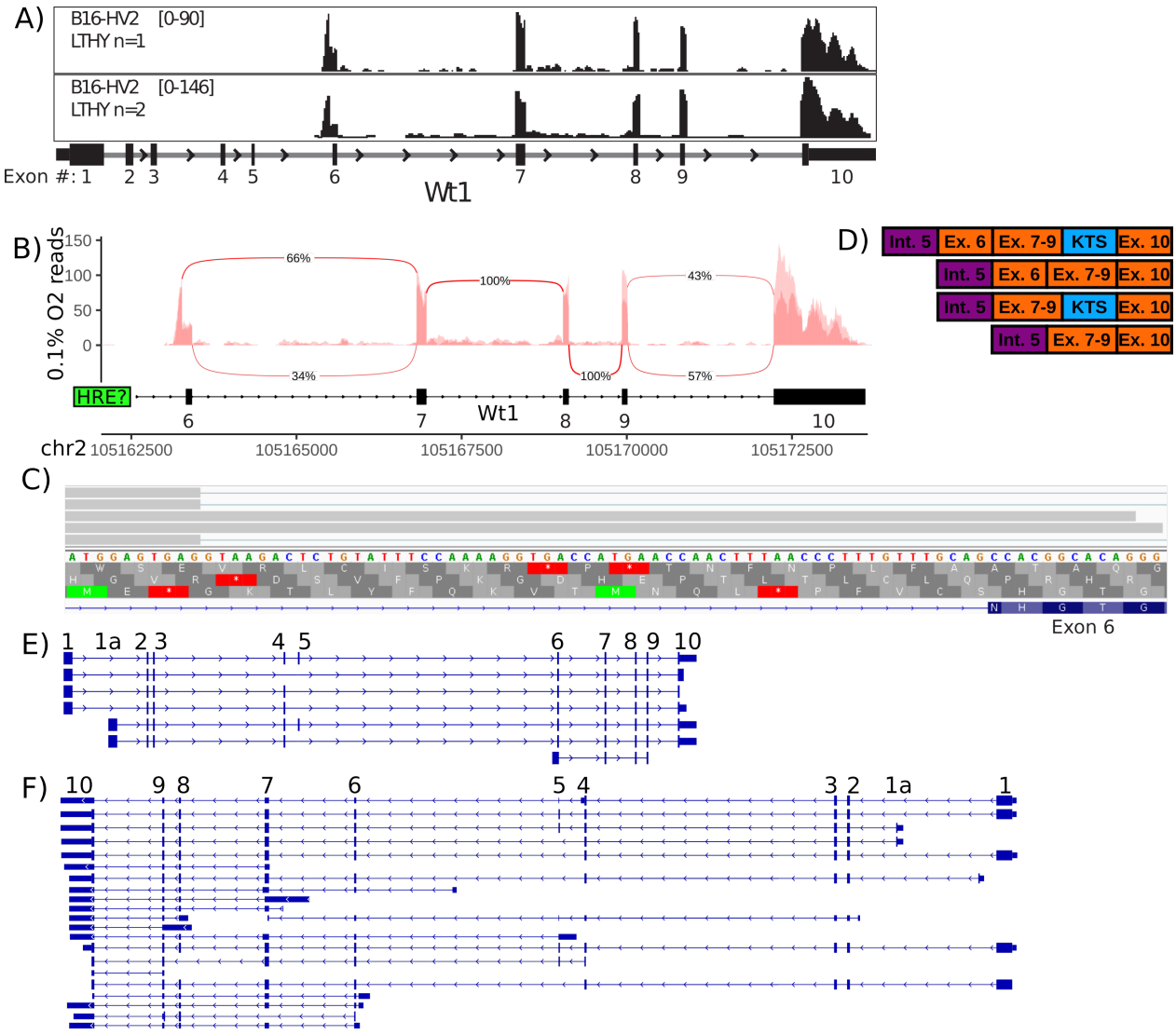

Figure 31: WT1 RNAseq read coverage & potential isoforms

A) LTHY read coverage of the Wt1 locus, generated in IGV. Number ranges are coverage depths at the individual nucleotide level. **B**) Splicing events observed in LTHY data. Percentages are the average between replicates, coverage depths are overlaid. **C**) Example of intron 5 to exon 7 splicing and read-through events in the LTHY RNAseq reads. Top: Mapped RNAseq reads. Middle: Genomic sequence and possible peptide sequences. Bottom: WT1 mRNA, beginning of exon 6 on far right. **D**) Possible tWT1 isoforms. Purple: Intronic sequence. Orange: Exonic sequence. Blue: KTS motif. **E&F**) Known WT1 mRNA isoforms in murine (GRCm39, **E**) and human (GRCh38, **F**) genomes.

4.3.3.4 HIF1α ChIP-qPCR identifies functional HRE elements inside of *Wt1* intron five.

Since read coverage of tWT1 began within intron 5, transcription factors were likely binding somewhere within intron 5 to drive this novel transcription. A TFBS analysis was done using R and the TFBStools library across the entire intron 5 sequence (**Figure 32A**). Among several other transcription factor binding sites, several HREs were found across the sequence (**Figure 32A**, bottom of bars). Given that these HREs were all high scoring and significant, they were all tested for HIF1 binding by ChIP-qPCR. B16-wt cells underwent the LTHY incubation protocol, but were processed for ChIP-qPCR after 36 hours of exposure to 0.5% O_2 using a polyclonal anti-HIF1α antibody. This was done to capture the initial mechanism of tWT1 upregulation, which was the most likely to be directly regulated by hypoxia. Of the probed WT1 intron 5 regions, 4/6 showed a similar or better enrichment to *Car9* and *Vegfa*, canonical HIF1 regulated genes (**Figure 32B**). Interestingly the HRE most proximal to the tWT1 TSS showed weak enrichment, suggesting that HIF1 functions more as an enhancer rather than a direct driver of tWT1 transcription. There is previous evidence for HIF1 functioning as an enhancer activating hypoxic transcription of genes such as erythropoietin[689].

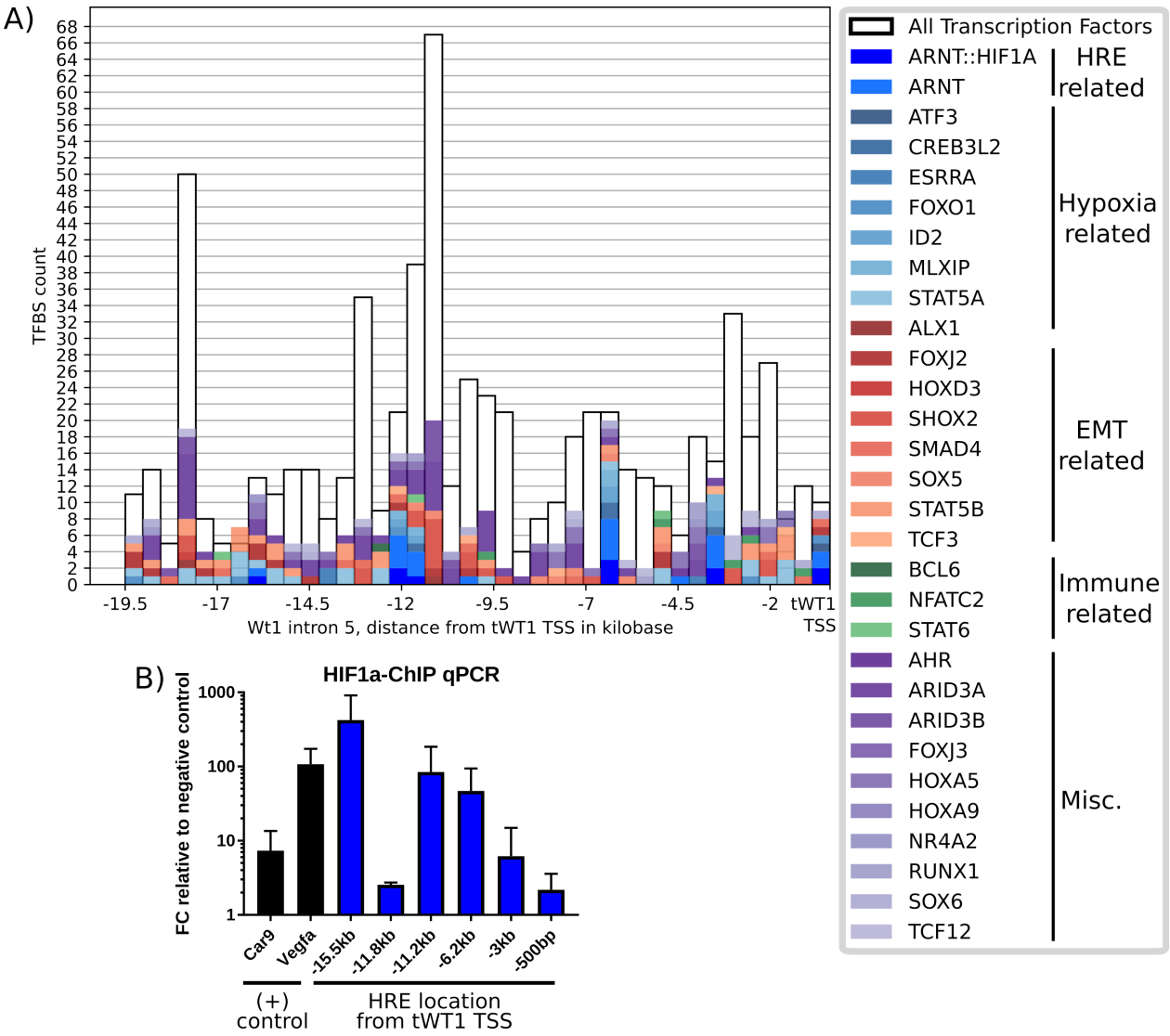

Figure 32: Hypoxic control of tWT1 transcription

A) Transcription Factor Binding Site analysis of murine *Wt1* intron 5 from beginning of intron 5 to tWT1 TSS. Only considered TFBSs with a score ≥ 0.95. Intron 5 sequence is broken into 40 bins, ~500bp/bin. Coloured TFs have an minimal expression of ≥ 100 DESeq2 normalized reads across any LTHY condition. **B**) HIF1a-ChIP-qPCR. B16-HG cells were underwent the LTHY protocol up to 36hrs of exposure at 0.5% O_2, and were then processed for ChIP-qPCR. Negative control used was Fyn_neg probe. Experiment represents biological triplicates, error bars are SD.

4.3.4 Functional characterization of novel tWT1 isoforms

4.3.4.1 Exon6-skip tWT1 generates a polypeptide derived from a intronic start codon

Because these truncated WT1 mRNAs were being highly expressed under deeply hypoxic conditions, we hypothesized that these mRNAs were functional and were capable of generating a polypeptide. Within the exon 6-10 spliced sequence, there are four start codon in the correct frame to generate the remaining WT1 protein sequence, as well as a start codon which is out of frame for the canonical WT1 protein sequence and is found in exon 6 (**Figure 33A**). All of these in-frame start codons are within exons 7-10, and are found in all tWT1 transcripts. Within the sequence derived from intron 5, there are two start codons. The first occurs exceptionally close to the beginning of the transcript, at the third nucleotide. This start codon generates a small peptide derived from the intron 5 sequence 23 amino acids in length. Shortly after this open reading frame, there is a second start codon. This start codon is only in-frame when the exon 6 skip splicing event occurs, if that splice does not occur and the transcript contains the exon 6-7 splice, then this start codon is out of frame. Given all of these start codons and potential protein sequences, we began by first verifying if any of these tWT1 sequences could generate protein.

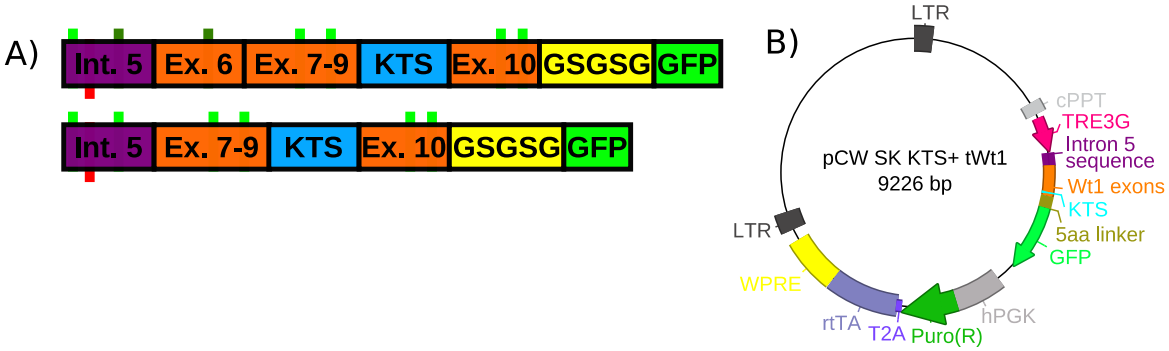

Figure 33: Potential tWT1 start codons and plasmid designs

A) tWT1-GFP isoform coding sequences. Purple: Intron 5 sequence. Orange: Exonic sequence. Yellow: Flexible linker, with peptide sequence. Green: GFP CDS. Green boxes above coding sequence are start codons; Bright-green are in-frame, dark green are out of frame. Red boxes below coding sequence are in-frame stop codons. **B**) plasmid map for lentiviral Doxycycline-inducible expression plasmids.

To test these tWT1 mRNA isoforms for protein production, we generated Doxycycline (DOX) inducible lentiviral plasmids of these different tWT1 transcripts (**Figure 33B**). The tWT1 transcript from the beginning of the 5'UTR to the end of the CDS was inserted into the DOX-inducible plasmid directly upstream of the DOX-inducible promoter TRE3G. To generate a detectable protein synthesis signal, the stop codon of the tWT1 CDS sequence was removed, and replaced with a GSGSG liner followed by a start codon deficient eGFP CDS. Using these plasmids, we expected no GFP signal under DOX deficient conditions. With DOX supplementation to the media, we expected eGFP signal to be generated only if the tWT1 transcripts were capable of generating protein. Given that the KTS motif is irrelevant for determining if these transcripts generate protein, tests began using the exon 6 skip KTS+ and exon 6 KTS+ transcripts.

Initially, HEK cells were transduced with the exon-6 KTS+ construct. After DOX induction, there was a clear increase in GFP signal, indicating protein production (**Figure 34A**). Subsequently, wild-type HEK and B16 cells were transduced with the exon-6-skip KTS+ construct. Following DOX induction under standard tissue culture conditions, there was a clear increase in GFP signal in HEK cells (**Figure 34B**). This proves both of the 5' splicing variants of tWT1 are capable of producing protein. Despite there being clear evidence of GFP induction in HEK cells, there was substantially less induction in B16 cells, and the cells which were GFP+ were not as dynamic as the transduced HEK cells. Given that tWT1 expression occurred under hypoxic conditions, we hypothesized that tWT1 in B16s may be stabilized under hypoxic conditions. Following simultaneous DOX and hypoxia incubation, there was a large increase in the percentage of GFP+ cells in both B16s and HEK cells (**Figure 34B, right column**).

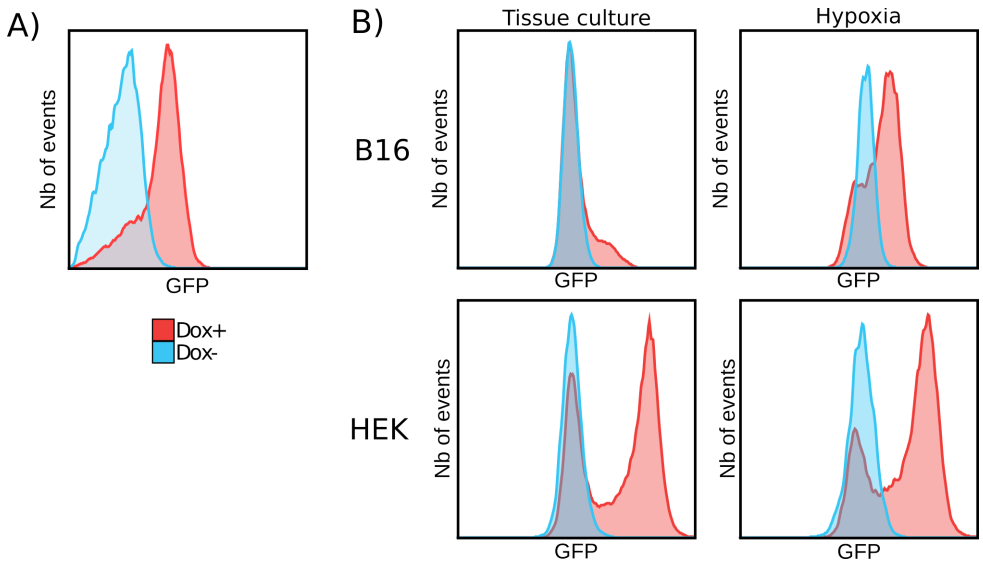

Figure 34: tWT1 peptide confirmation

A) FACS analysis of HEK-pCW-Exon-6_KTS+_GFP cells with (red) or without (Blue) Dox treatment. **B**) FACS analysis of HEK and B16s stably expressing pCW-exon-6-skip_KTS+_GFP under 20% or 1% O_2. **A,B**) Dox treatment was 2ug/mL for 72hrs.

Having established that the tWT1 transcripts could indeed generate protein, we sought to determine their relative functionalities. We began by performing microscopy on cells expressing the various transcripts. We hypothesized since WT1 is a transcription factor, functional isoforms of tWT1 should localize to the nucleus, given the presence of predicted nuclear localization signals (NLS) within exons 8-10[690]. To verify this, cell lines expressing DOX-inducible C-terminally tagged GFP tWT1 isoforms were examined by microscopy (**Figure 35**). As expected, cells expressing only GFP had cytoplasmic GFP signal localization (**Figure 35A**). Unexpectedly, the exon-6 isoform did not display nuclear localization, suggesting dysfunction (**Figure 35B**).

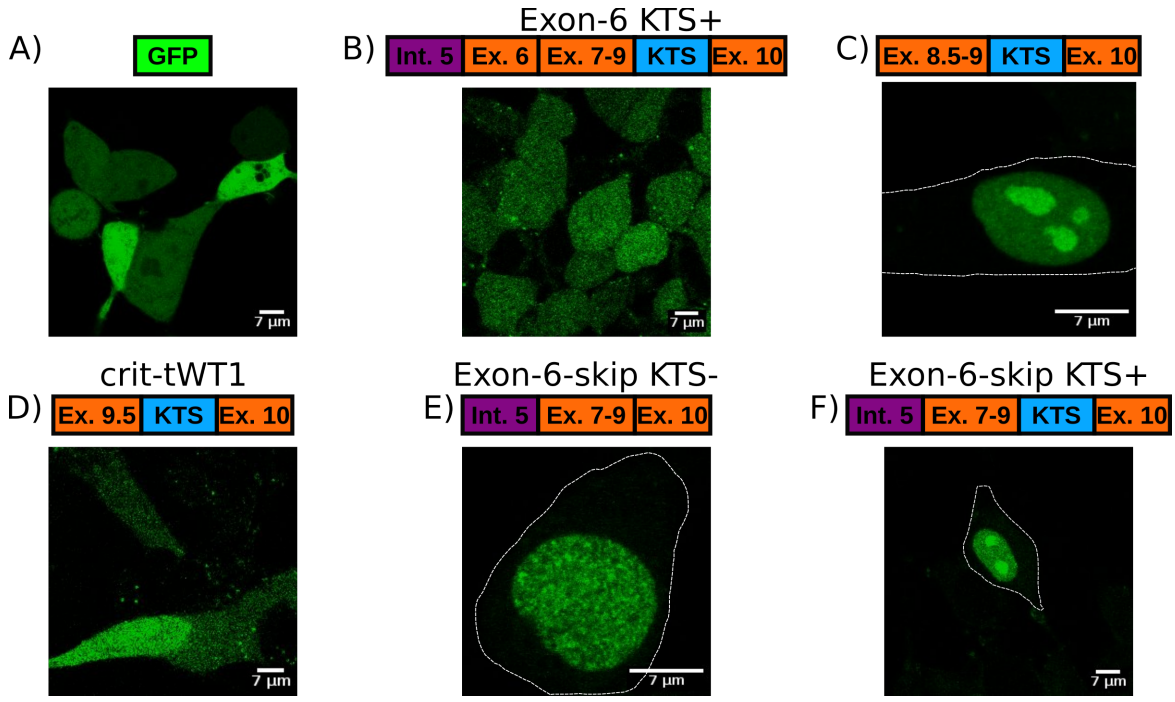

Figure 35: tWT1 subcellular localization

Microscopy images of tWT1-GFP fusion constructs. All GFP fusions were at the C terminus. Construct name (if applicable) is on top of the CDS diagram. **A**) Visualization of cytoplasmic GFP in HEK cells. **B**) Visualization of exon-6 KTS+ GFP fusion protein in HEK cells. **C**) Visualization of further truncated version of tWT1 GFP fusion protein in D4M cells. **D**) Visualization of crit-tWT1 GFP fusion protein in D4M cells **E**) Visualization of exon-6-skip KTS- GFP fusion protein in B16 cells **F**) Visualization of exon-6-skip KTS+ GFP fusion protein in B16 cells.

To identify what degree of WT1 truncation was needed to generate the lack of nuclear localization seen in exon-6 KTS+ GFP construct, further truncated versions of WT1 were generated. The first began translation approximately halfway through exon 8, and begins at the exact start of the first zinc finger (see **Figure 36B**, 85 a.a. into shown protein sequence, SHLQ**...**). This shorter isoform displayed excellent nuclear localization, suggesting that exon-6 KTS+ translation was not being initiated from the two exonic start codons upstream of this smaller tWT1 isoform (see **Figure 36B**, 71 a.a. into shown protein sequence, MCAYP... or at 89 a.a MHSR...) (**Figure 35C**). Therefore, exon-6 spliced tWT1 isoforms must initiate in-frame translation from one of the two most 3' start codon found on line 543 of the shown transcript in

Figure 36B, expressing at most 13 amino acids of WT1. Given this short peptide sequence length, it is unlikely to be functional. Within exon 6 there is a start codon, however it is out of frame. It is possible that translation is being initiated from here, but this was not investigated.

To determine the approximate location of the NLS signal in tWT1, a second even more truncated version of WT1 was generated (**Figure 35D**). This isoform began translation within exon 9, from within the third zinc finger. This extremely truncated isoform did not display nuclear localization, and was termed crit-tWT1 as it was critically truncated to the point of loss of localization. We therefore conclude that the WT1 NLS is between zinc fingers 1-2, including zinc finger 1. This also correlates with the translation initiation site for exon-6 isoforms, as they display similar expression patterns.

Given the apparent lack of functionality of exon-6 isoforms, we focused our attention to exon-6-skip isoforms. When expressed, both KTS+ and KTS- isoforms localize exclusively to the nucleus (**Figure 35E&F**). In addition to nuclear localization, exon-6-skip KTS+ also displayed nucleolar localization, in line with previous studies and the increased RNA binding ability of full length KTS+ WT1. Exon-6-skip KTS- tWT1 isoform did not display this nucleolar localization, and had smaller punctuated points of increased GFP intensity. Given the localization patterns of all tWT1 isoforms, we concluded that the exon-6-skip isoforms were the major functional isoforms derived from this novel TSS.

Having identified the functional tWT1 isoforms, we sought to determine the active start codon in exon-6-skip isoforms of tWT1. Based on the exon-6-skip mRNA sequence there were six in-frame start codons, two from the intron 5 sequence and four within the exonic sequence. The first and last two could be safely ignored for the following reasons. If translation initiated from the first start codon, a ribosomal read-through event of the intronic stop codon would be needed to generate the functional protein, which is highly unlikely given its expression levels. The last two start codons in the WT1 CDS are not sufficient to produce nuclear localization, and therefore can be ignored. This leaves three possibilities: the intronic start codon, and two start codons surrounding the first zinc finger. From these start codons, the expected GFP-fusion protein sizes are 47, 42.5, and 40kDa respectively (in KTS+ isoforms).

Given these protein sizes, it is possible to differentiate between them via a Western blot. Using HEK cells stably expressing either GFP alone or the exon-6-skip KTS+ GFP fusion in the presence or absence of DOX, a Western blot was performed to investigate the relative size of the fusion protein (**Figure 36A**). The blot showed a GFP specific band approximately halfway between 40 and 50kDa, suggesting translation was being initiated from the intronic start codon. To confirm this, an enriched sample of the exon-6-skip KTS+ GFP fusion protein was prepared by immunoprecipitation followed by size separation by SDS-PAGE. Once sufficiently migrated, the 40-50kDa section of the gel was excised and submitted for mass spectrometry (**Figure 36B**). Using the theoretical tWT1 protein sequence as reference, several isolated peptides specific to the tWT1 protein were found (**Figure 36B**, protein sequence, coverage highlighted in green). It should be noted, there was peptide coverage upstream of the exonic start codons, confirming that they are not the translation initiation start site. Focusing on the beginning of the mass spectrometry coverage, we could see coverage stopped just before the second intronic in-frame start codon. Given that the preceding protein sequence did not contain a trypsin cleavage site, we can conclude that translation was initiated from the second intronic start codon. In line with this, there was no peptide coverage of the methionine encoded by the second intronic start codon, which is to be expected if it was the initial methionine[691]. Additionally, the Kozak sequence context surrounding the second intronic start codon was strongest in the transcript, further suggesting its functionality (**Figure 36C**).

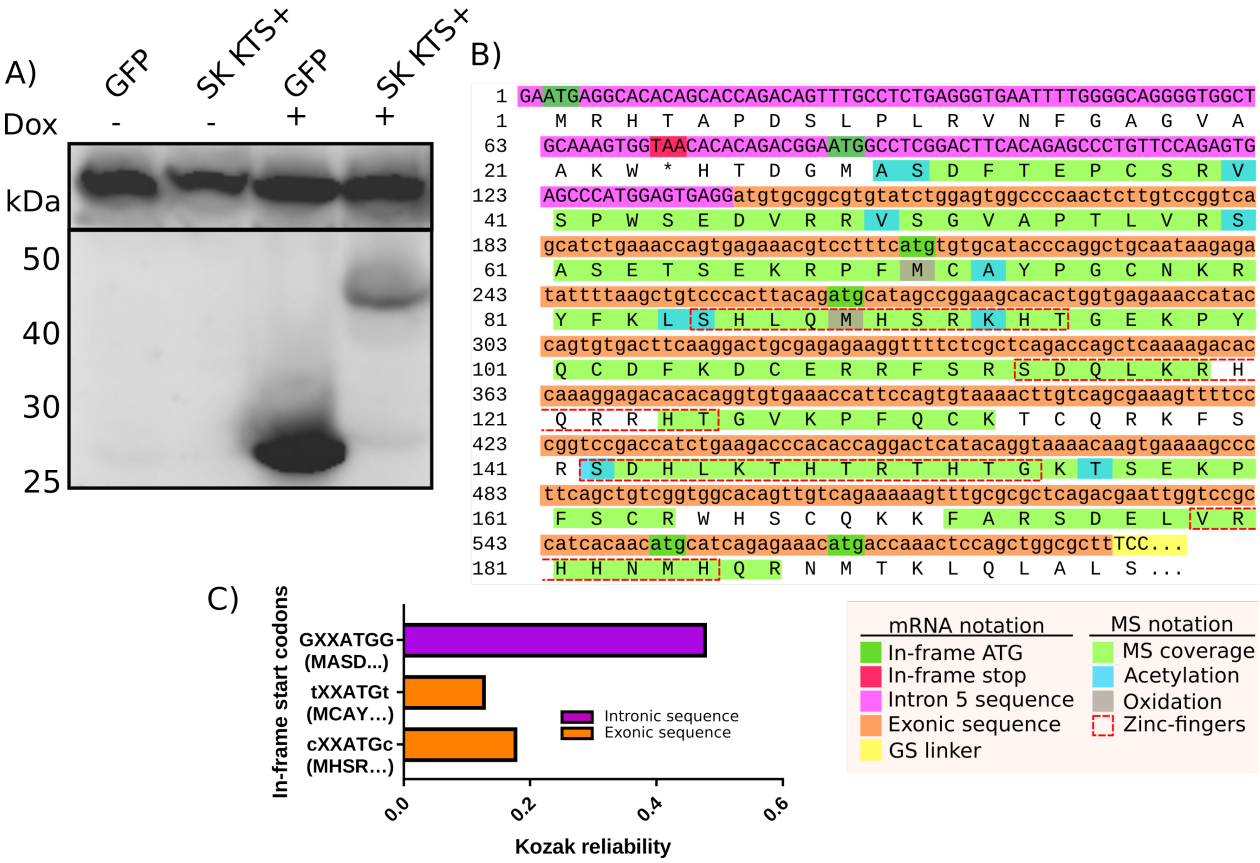

Figure 36: Confirmation of tWT1 translation initiation site

A) Western Blot of HEK cells expressing DOX inducible GFP or tWT1_exon-6-skip_KTS+_GFP. Top: anti-Calnexin. Bottom: anti-GFP. **B**) Mass Spectrometry (MS) coverage of tWT1_exon-6-skip_KTS+ purified from HEK cells. mRNA sequence is shown on top, expected protein sequence is shown underneath. mRNA and protein lengths are shown on the left. Refer to the legend below for full annotation list. **C**) Kozak reliability scores for in-frame start codons of exon-6-skip KTS+ mRNA.

Using the tWT1 mass spectrometry data, we were able to identify two types of post translation modifications (PTMs). The first and most prevalent was acetylation, which can be found on alanine, serine, valine, leucine, lysine, and tyrosine residues. The acetylation could be an example of N-terminal acetylation, a common PTM in the human proteome[692], and adds further weight to the conclusion that translation is initiated here. There is also evidence of acetylation within zinc fingers 1 & 3. How this may affect their DNA binding ability is unknown.

The second PTM found in our data is methionine oxidation. Of the three methionines of which we had coverage of, the first two are oxidized. The second oxidized methionine is in the middle of the first zinc finger, potentially altering its structure and function. Curiously we did not have oxidation of the second to last methionine in the tWT1 protein sequence, which lies in zinc finger 4. Wether or not these modifications is genuine or an artifact of the mass spectrometry sample processing protocol, remains to be determined[693]. Wether or not these PTMs are normal in WT1, or carry regulatory significance, is unknown.

4.3.4.2 ChIPseq of Exon6-Skip tWT1 reveals binding to a subset of WT1 binding motifs

Once we had confirmed the existence and localization of the exon-6-skip tWT1 isoforms, the next logical question was whether they retained functionality as a transcription factor. As these tWT1 isoforms expressed all four zinc fingers, it was expected they retain their nucleotide binding ability. We focused our attention on the KTS- version of the exon-6-skip tWT1 isoform, as it was the initial isoform studied for DNA binding, and was the better characterized of the two isoforms[694,695]. We generated a stable B16 cell line expressing the DOX-inducible exon-6-skip KTS- GFP fusion protein for ChIP-qPCR experiments. As a negative control, we generated a B16 cell line expressing the crit-tWT1, which only expresses the last zinc finger and lost nuclear localization. To ensure the correct epigenetic context upon DOX induction, these cell lines underwent the LTHY protocol up to 72 hours of incubation at 0.5% O_2. Upon incubation at 0.5% O_2, DOX was added to the cell media to induce tWT1 expression (**Figure 37A**). After 72 hours of induction, the cells were lysed and processed for ChIP-qPCR using an anti-GFP antibody. A sample of cells was collected for FACS analyses to confirm the presence of intracellular DOX and exon-6-skip KTS- GFP protein production (**Figure 37B**). FACS analyses confirmed that hypoxic cells were absorbing the DOX, and that GFP was present.

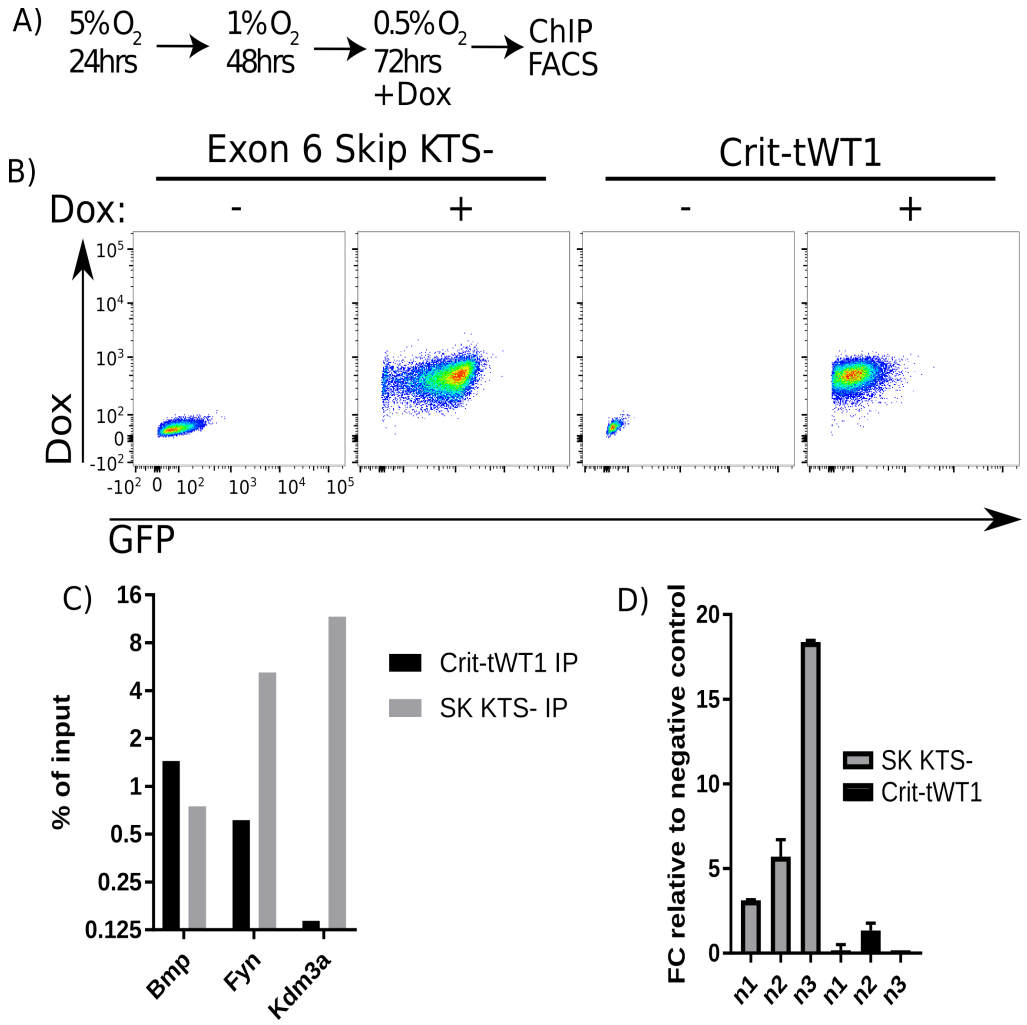

Figure 37: Exon 6 Skip tWT1 DNA binding ability

A) LTHY incubation protocol for ChIP-qPCR and ChIPseq experiments **B**) FACS data showing induction levels of exon-6-skip KTS- GFP and crit-tWT1 GFP in B16 cells after LTHY incubation. **C**) exon-6-skip KTS- GFP and crit-tWT1 ChIP-qPCR. ChIP was performed at 0.5% O_2 and done using an anti-GFP antibody. **D**) ChIPseq quality control prior to sample submission. Fold-change was calculated relative to *Fyn* negative control probe and sample input controls using the 2^{-ddCt} method.

To verify if the exon-6-skip KTS- tWT1 isoform retained DNA binding, we used WT1 ChIP-qPCR probes previously used in the literature targeting the WT1 binding site in the promoters of *Bmp*, *Fyn*, *Kmd3a*, and their negative control[526]. The negative control binds to a WT1 binding

sequence negative region of the *Fyn* gene. Using these probes, we found that the exon-6-skip KTS- GFP fusion protein was highly enriched on the *Fyn* and *Kmd3a* promoter sites relative to both input controls and the crit-tWT1 (**Figure 37C**). This confirmed that exon-6-skip KTS- tWT1 retained DNA binding activity.

To gain a comprehensive view of the DNA binding sites of exon-6-skip KTS- tWT1, The ChIP was repeated for ChIPseq submission. We processed both exon-6-skip KTS- GFP B16 cells, as well as crit-tWT1 expressing cells for ChIPseq. Proper induction of both constructs was confirmed via FACS prior to ChIP, and ChIP-qPCR was performed on *Fyn* to confirm enrichment (**Figure 37D**). While all sample did show enrichment, there is a large spread between replicates. Nevertheless, enrichment was present and considerably greater than the crit-tWT1 ChIP sample, so we proceeded with the ChIPseq. ChIPseq was performed using both ChIP input DNA and crit-tWT1 IP as negative controls.

Following acquisition of the ChIPseq data, we examined TFBS motifs as an internal control. Known and de novo TFBS motif analysis was performed on the ChIPseq data using HOMER[533]. In both analyses, the top enriched TFBS motifs were WT1 DNA binding motifs (**Figure 38A**). This confirms at exon-6-skip tWT1 KTS- is still capable of binding to DNA, and that its truncation has not affected its binding motif.

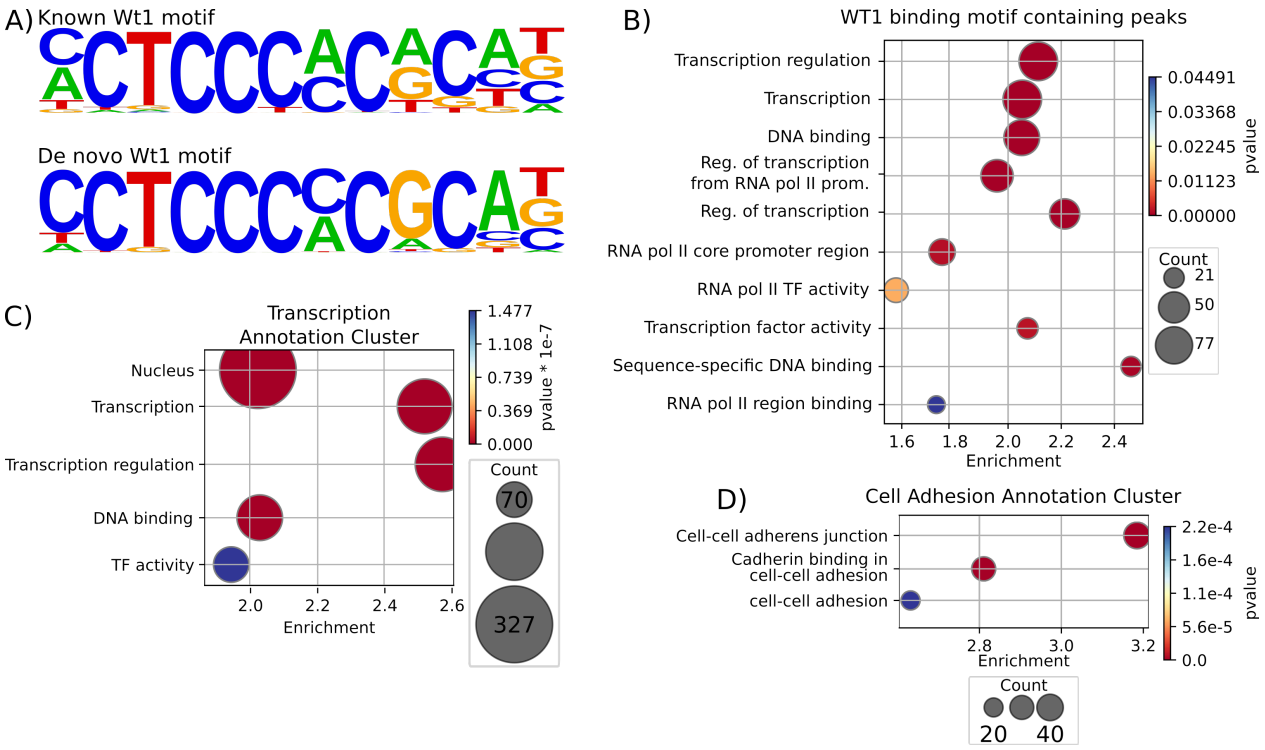

Figure 38: Exon-6-skip tWT1 ChIPseq analysis

A) Known and de novo TF motif analysis of exon-6-skip KTS- tWT1 ChIPseq data. Known motif p-value = 1e-78, is found in 36% of called peaks. De novo motif p-value = 1e-105, motif is found in 30% of ChIPseq peaks **B**) GO term annotation bubbleplot of genes with WT1 motif containing peaks. **C,D**) GO term bubbleplots of ChIPseq called plots.

HOMER analyses determined that only approximately 30% of the called ChIPseq peaks contained the known WT1 DNA binding motif. We therefore investigated if there was functional enrichment in the genes with ChIPseq peaks containing the WT1 DNA binding motif. Genes associated with ChIPseq peaks containing the WT1 DNA binding motif were identified using a in-house Python script specifically coded for this purpose, and can search for any transcription factor binding motif in any genomic range. These genes were then tested for functional enrichment via DAVID GO term functional clustering. From this analysis, we found enrichment for several GO terms associated with transcription and transcriptional regulation (**Figure 38B**).

Given that only 30% of the ChIPseq peaks contained the WT1 binding motif, a secondary GO term enrichment analysis was done on the complete list of enriched ChIPseq peaks (**Figure 38C&D**). While there was a similar enrichment for transcription associated GO terms as in the previous analysis (**Figure 38C**), there was also enrichment for clusters of GO terms associated with cell-cell adhesion (**Figure 38D**). Notably, the functional enrichment analysis done with WT1 motif containing genes had weaker enrichment scores and p-values than the cell adhesion annotation cluster. This is surprising, as it suggests that the cell adhesion genes found in our ChIPSeq data do not contain WT1 binding motifs. If this is the case, it indicates that WT1 is indirectly binding to these genetic loci. TFBS motif analyses for these non-WT1 binding site containing, cell-adhesion associated genes may reveal what transcription factors tWT1 is in complex with. These three functional enrichment analyses suggests that exon-6-skip tWT1 predominately binds to genes encoding other transcription factors as well as specific cell adhesion genes, further supporting the hypothesis that tWT1 is driving our LTHY associated EMT.

To find transcriptional targets of tWT1, we focused on genes with clear ChIPseq enrichment which also became upregulated during the LTHY protocol at 0.5% and 0.1% O_2. With both transcriptomic and ChIPseq data correlating, we believed that these genes would make ideal candidates for further validation. Of the cell adhesion associated genes, *SMAD7* showed both exon-6-skip KTS- ChIPseq enrichment as well as an increase in mRNA levels at 0.5% O_2 and 0.1% O_2 (**Figure 39A&B**, top). Of the transcription associated genes, *GADD45g* had several exon-6-skip KTS- tWT1 ChIPseq peaks around the locus (**Figure 39A**, bottom). That in combination with the upregulation of *GADD45g* under tWT1 inducing levels of hypoxia suggests *GADD45g* is a genuine transcriptional target of exon-6-skip KTS- tWT1 (**Figure 39B**, bottom). To verify these genes are targets of tWT1, we generated stable B16 cell lines expressing shRNAs targeting WT1. We used previously published shRNAs, and ensured that the shRNAs would target our truncated isoforms[696]. However, these shRNAs were not sufficiently effective in our cells, as WT1 was still being upregulated 3-fold at 0.5% O_2 and 11-fold at 0.2% O_2 relative to tissue culture conditions (**Figure 39C**). Without sufficient depletion of WT1, we could not validate our which genes in out ChIPseq data were genuine tWT1 targets.

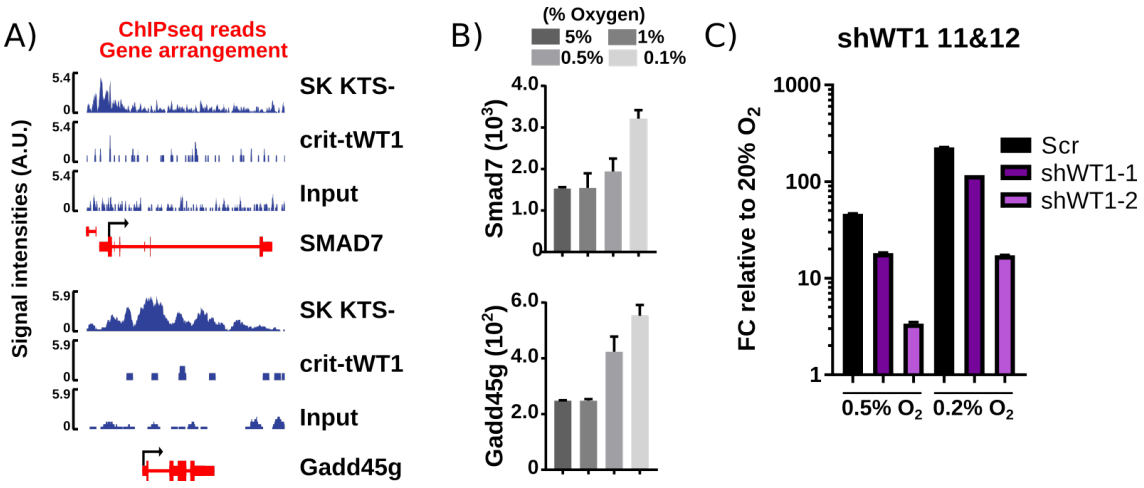

Figure 39: tWT1 ChIPSeq specifics

A) exon-6-skip KTS- tWT1 ChIPseq coverage and LTHY RNAseq expression profiles for genes Smad7 (top) and Gadd45g (bottom) **B**) WT1 qPCR expression under WT1 knock-down conditions. Values are normalized to tissue culture grown samples of each cell line, housekeeper used is 18s.

4.3.5 Identification of tWT1 isoforms in human cancer samples

Having validated the existence and functionality of this novel WT1 isoform in murine cancer, we naturally asked the question if this isoform could also be found in humans, particularly primary patient samples. Overall, there are no major differences in the *WT1* locus between humans and mice (**Figure 40A**). In the literature, there has been sparse publications describing a similar isoform of WT1 in humans called isoform G[350,697]. This isoform begins transcription at the 3' end of intron 5 and initiates a splicing event while still in intron 5, similar to the murine exon-6-skip tWT1 isoforms (**Figure 40B**, top). Unlike the murine tWT1 isoforms (**Figure 40C**, top), human isoform G links this splice to exon 6 and continues splicing the rest of the gene normally. Like the murine ortholog, human WT1 isoform G contains two in-frame start codons from exon 3 of the transcript, and generates identical peptides between the species. Despite this, the original study found that translation initiated from a non-canonical CTG start codon within the intron 5 sequence[350].

To identify specifically isoform G in expression in human cancer samples, we required access to the patient RNAseq sequencing data, as many publicly available databases do not report isoform specific expression of genes. As such, we needed access to read-level data of patient cancer samples, preferably in a cancer type with known WT1 expression. To meet these requirements, we used the Leucegene project, a dataset of over 500 Acute Myeloid Leukemia (AML) primary samples taken from patients across the course of their treatment, and have been processed for RNAseq[698]. In AML, WT1 is often over-expressed, and is currently under investigation as a gene of interest for vaccine therapy[321,699]. As part of another project at IRIC, these RNAseq datasets were converted to kmer tables to which we were granted access for the identification of tWT1 expressing samples.

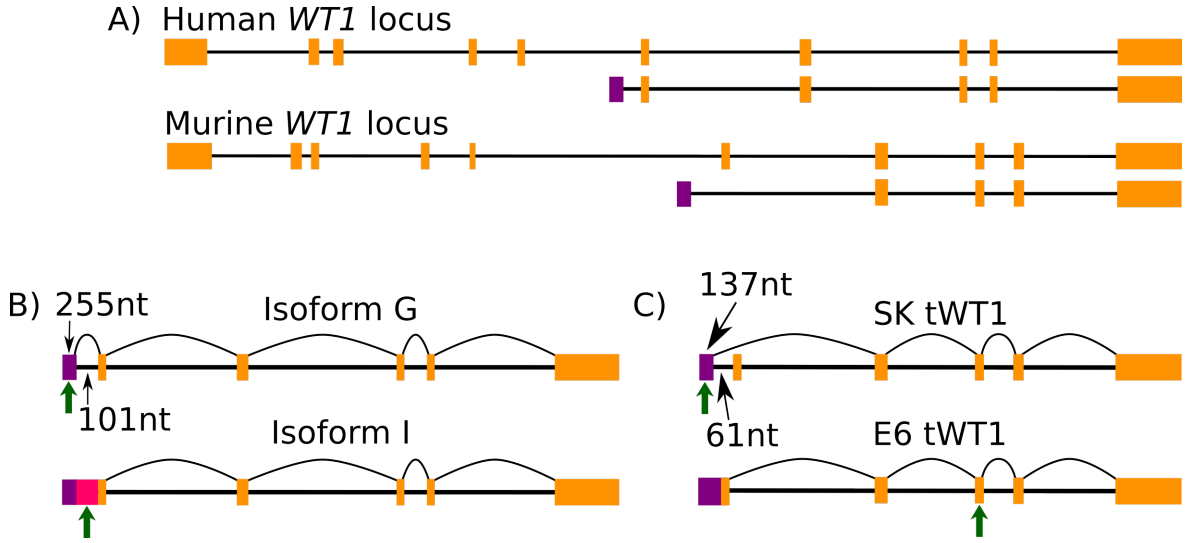

Figure 40: *WT1* locus structure and tWT1 orthologs in humans and mice

A) Genomic loci of human (top) and murine (bottom) *Wt1*. Shortened transcripts beneath loci are exonic structures of human WT1 isoform G and murine exon-6-skip tWT1 respectively. **B**) Human tWT1 isoforms. Top: Splicing diagram of human WT1 isoform G. Bottom: Splicing diagram of km-identified human tWT1 isoform I. Purple: tWT1 exonic sequence derived Wt1 intron 5. Pink: Intronic read-through section. Orange: *Wt1* exonic sequences. **C**) Murine tWT1 isoforms: Top: Splicing diagram of exon-6-skip (abbreviated SK) tWT1 isoform. Bottom: Splicing diagram of exon-6 (abbreviated E6) tWT1 isoform. Purple and orange represent same regions as in B. **B,C**) Green arrows symbolize established or theorized functional start codons.

Using these RNAseq kmer tables, we were not only able to identify samples expressing WT1, but samples expressing isoform G and another truncated WT1 isoform as well, using the *km* kmer analysis software (**Figure 40B**, bottom)[523]. We identified several samples which lacked the isoform G specific splicing event linking the first two exons of isoform G together, and instead has an intronic read-through of intron 1. This intronic read-through isoform will be referred to as Isoform I for the remainder of this thesis. After aligning isoform I with Ensembl annotated mRNA isoforms for WT1, we found that isoform I had been previously identified by the Havana project, and is referred to as WT1-211[347]. It is only found in the Ensembl database, and is classified as unable to generate protein. Despite this, isoform I was further investigated for similarities between it and murine exon-6-skip tWT1.

Interestingly, the intronic read-through in isoform I introduces an in-frame start codon from the intronic sequence. By performing a protein sequence alignment between the human and murine tWT1 isoforms, we can see that while the protein sequence derived from the canonical WT1 exons are highly similar, protein sequence by the N terminus are highly dissimilar (**Figure 41A**). While the N-terminal protein sequences may be dissimilar, the strength of their start codons are very similar (**Figure 41B**). In the human WT1 isoform I, this new start codon has a high Kozak reliability, at levels similar to those seen in the murine exon-6-skip tWT1 transcript (**Figure 41B**). These human splicing patterns are similar but reversed when compared to the murine tWT1 5' splicing events. In mice, a splicing event is used to bring an out of frame start codon into frame (**Figure 40C**). Whereas in humans, a lack of splicing introduces a new in-frame start codon. While mechanistically opposites, both events result in a strong start codon being brought into frame. Based on our experiments done on exon-6-skip and exon-6 tWT1, it is possible human WT1 isoform I is more functional than isoform G.

Through analysis of the Leucegene kmer tables, we found that 71% of samples expressed WT1 above or at 1RPKM. Of these samples, 24% contained isoform G or I expression. Of these samples expressing isoform G, 48% of them did not express any of the full-length WT1 isoforms (**Figure 41C**). When the frequency of these splicing events are rendered as a dot plot, we can see that there are several samples which preferentially or exclusively express this intronic read-through event, suggesting that cells were expressing both isoforms (**Figure 41D**).

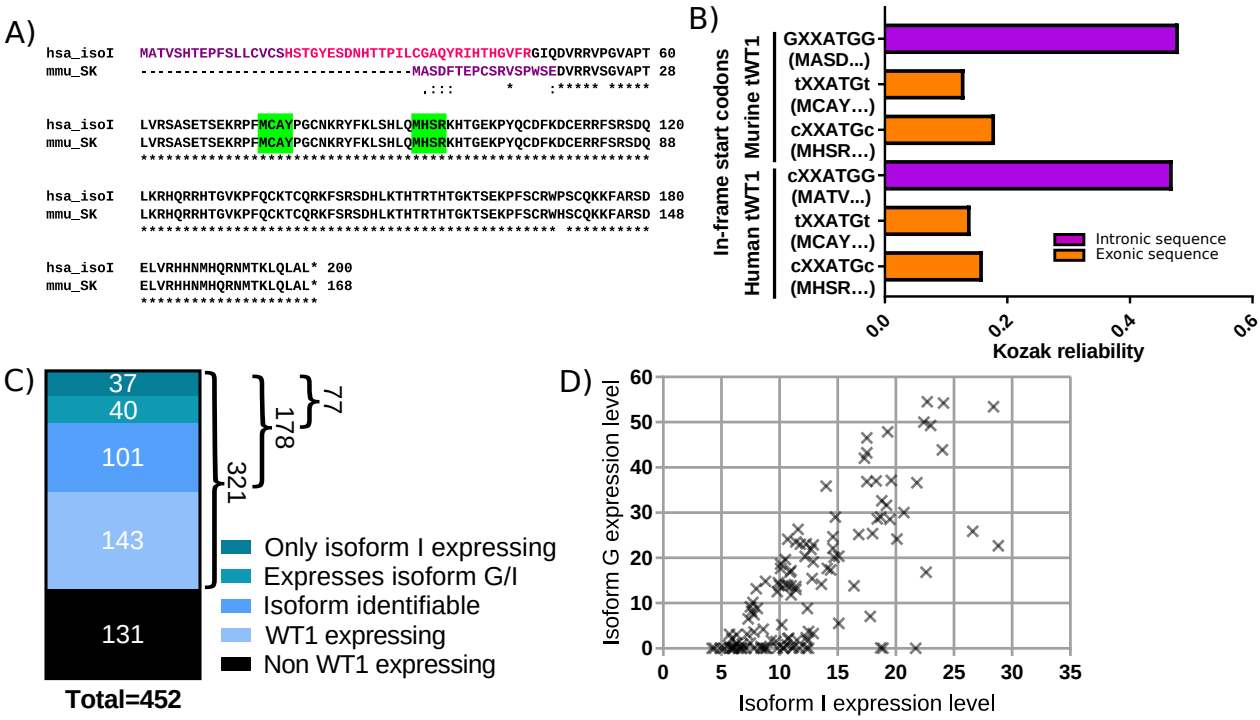

Figure 41: Leucegene WT1 isoform I expression

A) Protein sequence alignment between human WT1 Isoform I (hsa_isoI, top) and murine exon-6-skip tWT1 (mmu_SK, bottom). Purple: Protein sequence is derived from first exon of isoform G mRNA. Pink: Protein sequence is derived from isoform G intron 1 read-through. Black: Protein sequence is derived from *WT1* exonic sequences. Green highlighting: Alternative start codons **B**) Comparison of Kozak reliability scores between human WT1 isoform I and murine exon-6-skip tWT1. **C**) Breakdown of Leucegene samples with regard to WT1 expression **D**) Breakdown of isoform G/I expression in isoform G/I expressing Leucegene samples.

Having established that tWT1 expression can be found in human cancer samples, we next asked ourselves if isoform G/I correlated with patient outcome. To investigate this, survival curve analyses were performed using the maximal number of isoform G/I expressing samples, which corresponded to patient samples taken from all AML patients at any point during their treatment (**Figure 42**). Survival curves were generated from this sample list, and samples were split based on one of the following criteria: 1) Sample expresses full-length WT1 exclusively 2) Sample expresses tWT1 3) Sample expresses tWT1 exclusively. Expression of full-length WT1 provided a significant overall survival advantage (p=0.0013) (**Figure 42**, left). When examining

samples which expressed tWT1, which was a ~1:1 mixture of samples expressing full-length WT1 and tWT1 (**Figure 41C**), this survival advantage is lost (**Figure 42**, middle). When examining samples which exclusively expressed tWT1, these samples have a lower overall survival rate than those which do not express tWT1, however this observation is not significant (p = 0.29) (**Figure 42**, right). Whether or not tWT1 expression is truly insignificant to patient overall survival, or the insignificance is an artifact of a low sample number, remains to be investigated.

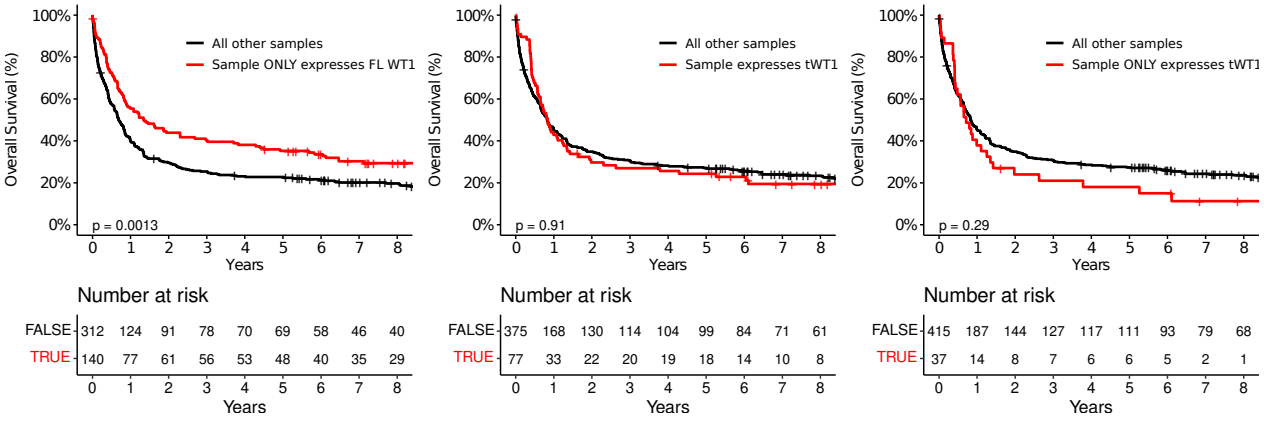

Figure 42: Leucegene sample tWT1 expression survival outcomes

Overall survival outcome of Leucegene patients considering WT1 isoform expression. Number of samples in each category remaining are found below each curve. **Left**: Of the WT1 isoforms, samples only express the canonical full-length (FL) WT1 (red). **Middle**: Of the WT1 isoforms, samples express isoform G or I. (red). **Right**: Of the WT1 isoforms, samples only express isoform G or I.

4.4 Chapter Summary & Discussion

In this chapter, we began by attempting to better simulate the *in vivo* hypoxic microenvironment by establishing a novel *in vitro* long-term hypoxia incubation protocol. The goal of this protocol was to simulate the depth and time *in vivo* hypoxia microenvironment takes to develop. While undergoing this incubation protocol, we observed a drastic change in the morphology of the cells, suggesting that this protocol induced some kind of EMT. To investigate this further, we performed deep transcriptome profiling.

Based on both DEmiR and DEG analyses, it was clear we had induced a non-canonical form of EMT. The only member of the miR-200 family that was sufficiently expressed was miR-200b, which was not significantly expressed or differentially expressed at the point the EMT was observed. However there was significant upregulation of EMT-promoting miRs such as miRs-221/222. From the miR analyses, there was clear indication of an EMT. The DEG analysis further supported the hypothesis that we had induced a non-canonical EMT through the LTHY protocol. Globally, the mRNA signatures were enriched for EMT associated phenotypes such as positive regulation of cell migration. But there was also enrichment for negative regulation of several canonical EMT driving pathways, such as TGFβ, SMAD, and BMP. A closer analysis of EMT driving genes revealed that only ZEB2 and GRHL2 were significantly expressed. Nevertheless, we had a strong EMT effector signature, through the expression of motility related genes like *S100A4* and several extracellular matrix modifying genes like *FN1*, *VIM*, and multiple collagen genes. Since we had a clear hypoxia-linked EMT, we looked at alterative EMT drivers, eventually settling on WT1 for further investigation.

After confirming hypoxic upregulation of WT1 in multiple cancer contexts, we focused on the exact requirements for hypoxic upregulation. Based on our initial findings, it was possible the WT1 upregulation was the result of a sufficiently long incubation under hypoxia, or specific hypoxia levels are required to induce expression. We found that either aspect alone was insufficient to optimally induce WT1 expression, and that optimal hypoxic WT1 upregulation required both a pre-incubation at milder hypoxic levels and sufficient time at these milder levels to adapt. Determining how short these timepoints can become, and what adaptation is occurring to allow for this observed WT1 upregulation will shed light on this understudied aspect of hypoxia adaptation.

Through examination of differential splicing patterns of WT1 in the LTHY RNAseq data, we identified a novel murine isoform of tWT1, likely from a novel TSS just upstream of exon 6. Interestingly, our RNAseq data showed that WT1-AS was not expressed throughout the LTHY experiment (data not shown), further supporting the idea that transcription was initiating from an alternative start site. This novel TSS not only demonstrates the existence of an alterative

WT1 isoform, but also a new mechanism for expressing WT1 isoforms in the absence of WT1-AS expression.

Using ChIP-qPCR, we showed that HIF1 is able to bind to multiple HREs upstream of the novel TSS. However these HRE binding profiles places HIF1 in the role of an enhancer. Further investigation into TFs binding proximally to the TSS is required. Through exogenous expression of these tWT1 transcripts, we determined isoforms made of a novel exon 6 removal splicing event generated functional protein, termed exon-6-skip tWT1. Through ChIPseq analyses, we determined exon-6-skip tWT1 binds to several genes associated with transcription and cell adhesion. However without a sufficiently effective way to reduce endogenous WT1 levels, validation of these targets is difficult. To overcome the failure of the shRNA attempt, knock-down could be re-attempted using a dual shRNA approach. Alternatively, endogenous WT1 protein levels could be completely inhibited through a CRISPR/Cas9 approach. Should both these avenues fail, prospective tWT1 binding sites can be cloned into a fluorescent reporter plasmid to probe for WT1 binding activity.

To determine the relevancy tWT1 expression in a patient setting, we attempted to find tWT1 expression in human patient samples using the Leucegene dataset. We initially probed for the most likely human ortholog to our murine tWT1 transcripts, human WT1 isoform G. Our isoform G detection method allowed for the simultaneous detection of isoform I, which derived from a read-through event of the first intron of isoform G. When this event occurred, it introduces a strong in-frame start codon. Strikingly, this new start codon had a much stronger kozak sequence than the other in-frame start codons. The kozak context for the relevant start codons in the human isoform I and the murine exon-6-skip tWT1 were very similar. This suggested protein production from isoform I would be much more effective, as we saw in our murine experiments. Further studies to verify the existence of this human isoform as well as its translation start site are required. Should this isoform exist its novel peptide sequence could made for an attractive immunotherapy target due to its highly specific and abnormal expression, assuming it can be presented on major histocomaptibility complexes.

Should isoform I function as predicted, it would be a fascinating example of convergent evolution in cancer. Two different cancer types from two different species, both attempting to express a truncated version of WT1 through opposing mechanisms: novel splicing in mice and intronic read-through in humans. The exact reason as to why is still not fully understood. As previously mentioned, further investigation into exon-6-skip tWT1 KTS- gene regulation is needed. Additionally, we have yet to investigate the role and function of exon-6-skip tWT1 KTS+, which also showed nuclear and nucleolar localization. Perhaps the KTS+ version of this isoform is binding to rRNA genes to prevent further generation of ribosomes in an attempt to limit overall translation rates. Alternatively, perhaps the roles of these isoforms lies not in DNA binding, but in RNA binding. Studies have shown that both isoforms are capable of binding to RNA through their zinc finger domains[341]. Perhaps these truncated isoforms are functioning as dominant negative RNA binding proteins, shielding their targets from other interactions such as RNA binding proteins or miRNA mediated degradation. In this case, their effect on the transcriptome may be extremely subtle. Investigations using RNA immunoprecipitation sequencing on these isoforms would shed light on the role these proteins play in influencing the RNA landscape of the cell.

Ultimately, we expected expression of these tWT1 isoforms would correlate with reduced overall survival in patients. Unfortunately, our survival analysis using the Leucegene dataset showed non-significance for isoform G/I expression. However, the fact that the prognostic value of WT1 expression flips from being a positive marker of survival to a non-significant negative marker is striking. I believe that with additional samples, we will show tWT1 expression to be a negative prognostic marker. The difficulty resides in gaining access to these samples. Due to the novelty of these isoforms, access to the unaligned RNAseq reads from the patient samples is required, or kmer tables generated from unaligned reads. This is necessary to identify samples expressing these isoforms, as most databases only report on gene-level expression, not isoform level. Additionally, isoform I is classified as as a non-productive transcript, and many RNAseq quantification pipelines may not detect it. For these reasons, access to RNAseq read level data to properly identify isoform I expression is essential.

Chapter 5 – Designing single-sequence, multi-targeting RNAi molecules

Jordan Quenneville[1,2], Albert Fengalhy[1], Francois Major[1,3], Etienne Gagnon[1,4]

1. Institute for Research in Immunology and Cancer
2. Department of Molecular Biology, University of Montreal
3. Department of Computer Science and Operations Research, Faculty of Arts and Sciences, Université de Montréal
4. Department of Microbiology, Infectology, and Immunology, Faculty of Medicine, Université de Montréal

5.1 Authors & contributions

Jordan Quenneville: Cloning of plasmid constructs. Generation and maintenance of cell lines. FACS acquisition and data analysis. Coding of GFP NT3'UTR seed evaluation script. Creation of ration-design synmiR design algorithm. Coding of pNL-NT design script. Generation of figures.

Albert Fengalhy: Coding of Abox optimization algorithm. Troubleshooting code.

Francois Major: Conception of project, analysis and interpretation of results, figure preparation. Creation of minimal-mutation synmiR design algorithm.

Etienne Gagnon: Conception of project, analysis and interpretation of results, figure preparation.

5.2 Executive Summary

When studying RNAi, the RNA duplex structures generated by the RNAi molecule and the local target site can be broadly classified into two categories. The first being the perfect complementary found in exogenous artificial RNAi such as siRNAs or shRNAs. The second is the seed-driven complementarity found in miRNA:target duplexes, where seed complementarity appears to drive specificity while the tail sequence minimally aligns to the target sequence. Despite not aligning to the target sequence as in siRNA structures, the tail sequence in miRNAs is nonetheless functional.

In an attempt to bridge the gap between these two RNA duplex architectures, as well as illuminate the functionality of miR tail sequences, we explored suboptimal shRNA tail sequences in a exogenous gene knock-down setting. We began by generating RNAi structures where the seed is bound, but the central portion of the duplex was unbound, with the remaining tail sequence forming a duplex at increasingly larger distances from the seed duplex. To our surprise, we found as the tail duplex was allowed to form further from the seed duplex, knock-down efficacy increased accordingly. This led us to propose a model for designing single-sequence, multi-targeting RNAi molecules which we termed synmiRs for synthetic microRNAs.

To design these synmiRs, we developed two distinct design algorithms, which employ fundamentally distinct design philosophies. The first used a minimal-mutation algorithm to mutate shRNA sequence to bind to multiple sites. The other uses a rational-design algorithm, where tail binding sequences were ranked based on sequence conservation and used to design synmiR tails based on user-defined design parameters. To test these synmiRs, we developed a HIV reporter system using mCherry expression as an output. In a dual transfection assay in HEK cells, we show that these designs are genuinely multi-targeting. However additional work is needed to demonstrate their mechanism of action. Should these design algorithms prove to be effective at designing synmiRs in multiple contexts, we will have provided a predictive model for miR-tail functionality, as well as provide a new category of RNA therapeutic.

5.3 Results

To develop effective multi-targeting RNAi molecules, we need a better understanding of the rules and limits governing effective RNAi. It's known that perfect siRNA:target complementarity is effective at reducing target protein output by the RNA-slicing activity of AGO2[362]. It's also known that miRNAs can effectively inhibit target protein production despite a lack of obvious complementarity beyond the seed[361]. To help elucidate the function of the miR tail sequence Yifie Yan, a PhD student in the Major lab, studied the effect of mutating regions of the miR-tail. In his now published work, he showed the Abox region (miR nts 9-11) was the least functionally relevant[419].

Building from that finding, we hypothesized if the Abox was dispensable for RNAi efficacy, that this could liberate the rest of the tail sequence to bind further away from the miR:target seed duplex. If this proved to be effective, it would open up possibility space for multi-targeted RNAi, as synmiR tails could bind to the same sequence at varying lengths from the seed at different target sites. A single RNAi molecule could bind to multiple targets with differing secondary structures so that each structure would be effective on each target site, but each site would have a distinct structure. While not published at the time of this hypothesis, this concept has gained traction recently with *in vitro* reports suggesting that the central portion of the RNAi:target duplex can be mismatched while still retaining efficacy[361,417,418].

5.3.1 miR tail binding does not need to occur with a 0 nt offset

5.3.1.1 Developing an *in vivo* reporter cell line for controlled RNAi

Within the RNAi literature there are several different methods for detecting RNAi, ranging from *in vitro* assays of AGO2:RNA dissociation, to cellular luciferase reporter assays[418,419,421]. For our experiments, we opted to use a cell-based reporter to get an accurate understanding of how our designs would work *in vivo*. We repurposed a previous construct developed in the Major lab, which consisted of a retroviral plasmid constitutively expressing a GFP mRNA with a synthetically designed 3'UTR 188nt in length. This 3'UTR was designed to be non-targeted by the microRNAs expressed in the human prostate cancer cell line DU145, making it an ideal testing sequence for exogenous RNAi in that cell line.

To increase the ease of generating stable cell lines, we cloned this mRNA sequence into the pLKO backbone (**Figure 43A**). To take full advantage of the previous analyses, we transduced DU145 cells with this construct and sorted on narrow GFP expression ranges (**Figure 43B**). From the bulk transduced population, we sorted on low GFP and median GFP expression. This dramatically reduced the spread of GFP expression, allowing for easier detection of GFP expression reduction, and controlling for relative abundance of GFP mRNA copies within the cell. Both of these factors are critical for determining GFP knock-down levels by FACS. After sorting, we could then use the reporter cell line for testing synthetic RNAi designs using the GFP 3'UTR.

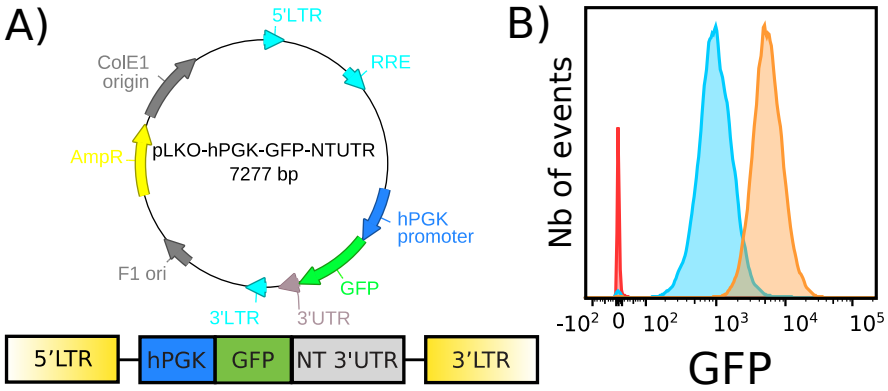

Figure 43: saRNA target reporter construct

A) Top: GFP reporter plasmid map. Bottom: Lentiviral payload of GFP reporter plasmid **B**) GFP expression levels of sort DU145-GFP NT3'UTR cells. Red: WT DU145 cells. Blue: DU145-GFP NT3'UTR sorted for low GFP expression. Orange: DU145-GFP NT3'UTR sorted for median GFP expression.

5.3.1.2 Designing bridged RNAi molecules

Having established the reporter for this project, the next step was to design sets of RNAi sequences to test how sliding the shRNA tail binding section further away from the seed duplex might affect knock-down efficacy. To investigate this, we chose to design a set of RNAi molecules for each site, where the purpose of each member of the set was to generate a different RNAi:target duplex structure (**Figure 44A**). To provide context for the maximal knock-down efficacy for any given site, a perfect shRNA 21nt in length would be designed for each site (**Figure 44A**, shRNA structure). From this shRNA design, three imperfect RNAi molecules were designed, collectively termed small artificial RNAs (saRNAs) due to their imperfect target binding profile (**Figure 44A**, bridge structures). All saRNAs maintained ideal proper seed binding, which for this project was defined as Watson-Crick complementarity from saRNA nts 1-8 (**Figure 44A**, blue/purple sections). All saRNAs were also designed to maintain the same 3' supplemental binding profile, which was defined as Watson-Crick complementarity from saRNA nts 12-21 (**Figure 44A**, green/yellow sections). This meant saRNA set members have different tail sequences, however this was inevitable when testing different saRNA:target duplex structures on the same target site. All saRNA set members also had the same Abox sequence, which was defined as saRNA nts 9-11 (**Figure 44A**, unhighlighted central sections of bottom strands). The

Abox was designed to minimize binding to the target sequence by mirroring the target sequence. If a given Abox sequence would promote binding in a given context, that sequence is modified to reduce binding in that context. Therefore, Abox sequences were set to mirror the target sequence in a 3nt bridge context, and were minimally mutated otherwise to reduce interactions with the extending target bridge sequence. The bridge sequence is defined as the target sequence between the seed duplex and the supplementary binding sequence (**Figure 44A**, unhighlighted central sections of top strands). For our experiments, we decided to test bridge sequence lengths of 3-5nt, as we expected that increased bridge length would result in reduced knock-down efficacy due to the increasing deviation from the ideal shRNA structure.

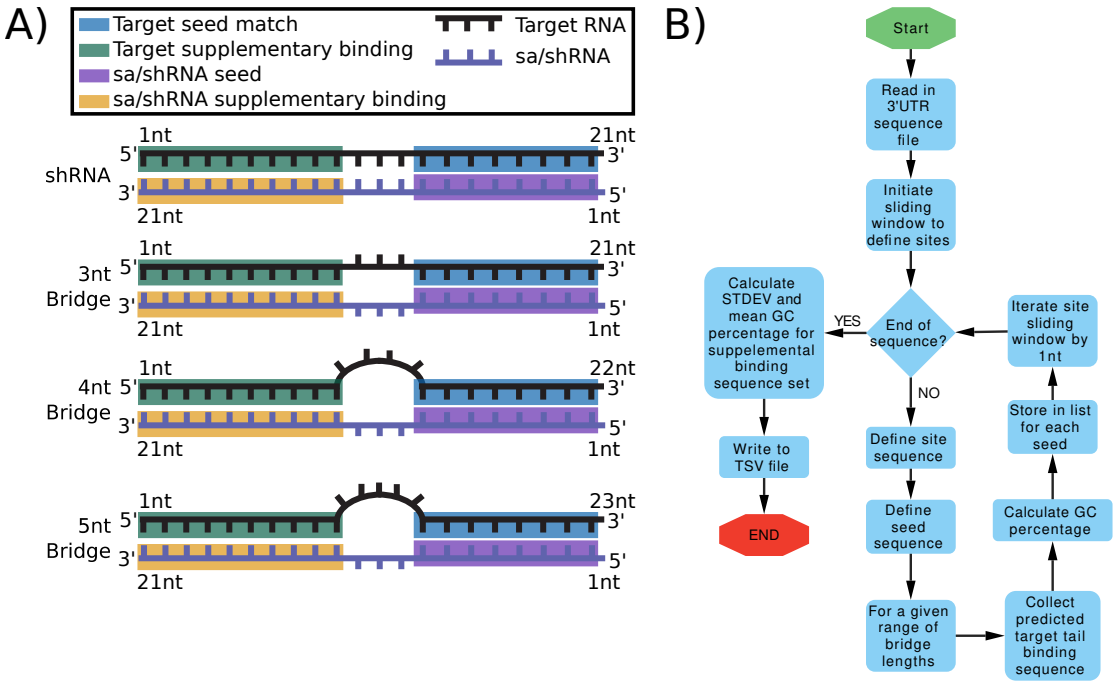

Figure 44: Algorithm for saRNA site choice

A) Example of set of RNAi molecules to be designed for a given site. Top strand corresponds to target sequence, bottom strand corresponds to sa/shRNA sequence **B**) Python algorithm for choosing NT'3UTR saRNA site sets.

Given the required saRNA set tail sequence variance for any given site, we needed to carefully consider which sites of the NT3'UTR sequence to test. Because all members of a given

site set would contain the same seed sequence, it was ignored. Primary consideration was given to the GC content for each saRNA supplemental binding region. Ideally, all saRNA supplemental binding regions would contain the exact same GC content, if not variation should be minimized. Given that we were only considering Watson-Crick binding, these sites were investigated by solely considering the target supplemental binding region. To find such sites, I wrote an algorithm in Python which scans the NT3'UTR sequence using a double sliding window approach (**Figure 44B**). The first sliding window scanned the NT3'UTR sequence one nucleotide at a time in chunks of sequence which represented a potential site. Sites are defined as the target RNA region needed for the largest saRNA:target duplex structure we were attempting to create. For our largest duplex structure, the 5nt bridge structure, this translates to a 23nt target sequence region. Each site was then evaluated using a second sliding window, where the target supplemental binding sequence for each sa/shRNA structure in the set was extracted, and then had its GC percentage is calculated. For the shRNA, the entire shRNA tail sequence (shRNA nts 9-21) were used for GC percentage calculations, whereas the relevant 10nt window for each saRNA were used for saRNA GC percentage calculations. Once the GC content for a site's target supplemental regions was collected, the average GC content and GC standard deviation are calculated and written to a flat text file for manual analysis.

Analysis of the output of this algorithm revealed that there were no sites which contained zero GC percentage variance for target supplementary binding sequences. We therefore chose sites with minimal GC content variance across the target supplemental binding regions. To avoid any given saRNA set from binding to unintended regions of the NT3'UTR sequence, we ensured the seed sequences for these sites were unique to the reporter mRNA sequence (GFP CDS and NT3'UTR sequences).

We chose three sites across the NT3'UTR sequence for testing (**Figure 45A**). The first two sites were relatively close to the GFP stop codon, and overlapped each other. The closest edges to the stop codon were 14nt for site one and 33nt for site 2 respectively. Site 3 was further down on the NT3'UTR sequence, and provided variance in the relative position of the sites across the NT3'UTR sequence. Given the correlation between relative position of the RNAi site across a 3'UTR and knock-down efficacy, site 3 provides some needed variance to the site

positions, and helps control for potential knock-down inhibition as a result of local secondary structures around sites 1 and 2[413]. For each of these sites, a control shRNA and three saRNA molecules were designed with specific duplex structures in mind (**Figure 45B**).

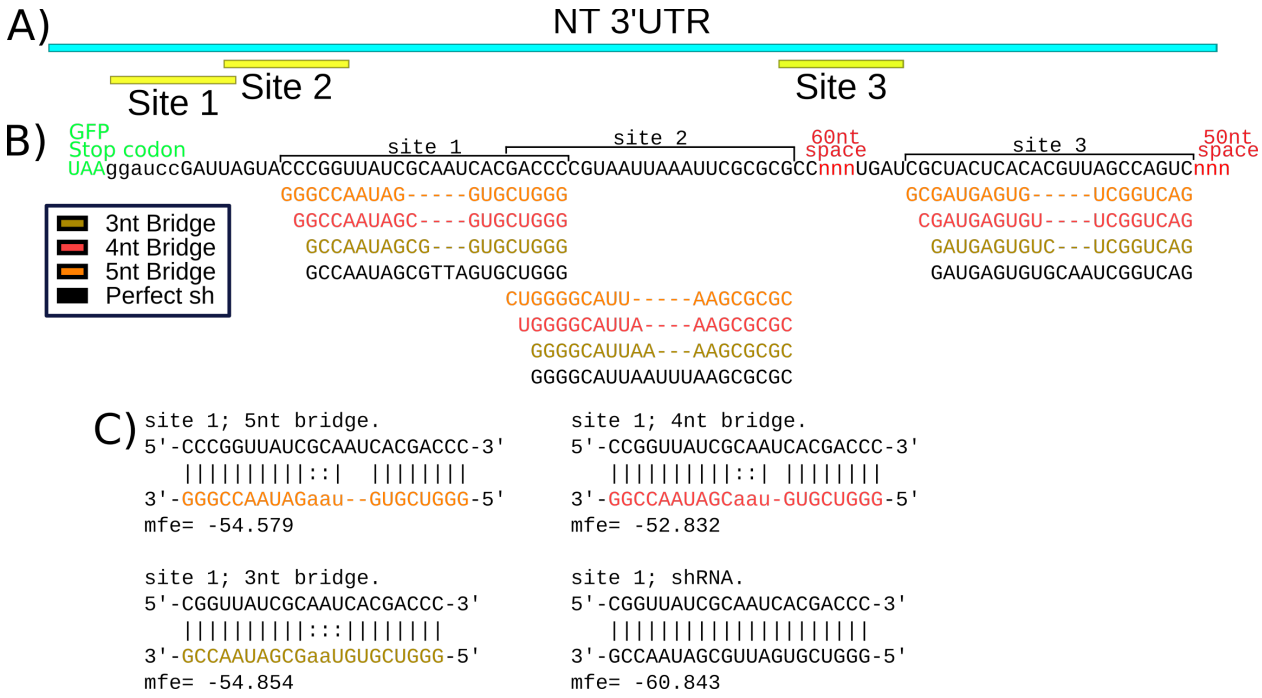

Figure 45: Tested saRNA bridge designs

A) Overview of site locations across the NT3'UTR sequence. **B**) Details of sa/shRNA intended hybridizations for all sites. **C**) MCFF minimal free energy (MFE) folding for site 1 sa/shRNAs. From left to right, top to bottom: 5nt bridge folding, 4nt bridge folding, 3nt bridge folding, shRNA folding. Top strands of each duplex corresponds to target sequence, bottom strand corresponds to sa/shRNA sequence. MFE for each duplex is shown beneath the duplex, in kcal/mol. "|" is Watson-Crick pairing; ":" is non-canonical pairing as determined by MCFF

Each site provided unique testing parameters for these saRNA designs. For site 1, the GC content of the tail sequence directly adjacent to the Abox decreased as the bridge length increased (**Figure 45A&C**). Should high GC content in this region be critical for saRNA activity, it would affect site 1 saRNA knock-down levels. For site 2, there was no change in the GC content

of any of the constructs, as the GCs of the site were located at the edges. As bridge lengths increased, only A or U were lost and gained (**Figure 45A**). Unlike sites 1 and 3, there are no GCs adjacent to the Abox in site 2, providing a negative control for the GC content hypothesis proposed for site 1. For site 3, GC content was spread out relatively evenly across the supplemental binding region, acting as a middle ground to the relative GC content concentrations of sites 1 and 2 (**Figure 45A**). The complete list of these constructs, including their Abox sequences can be found in **table 4** of this thesis. An example of the intended saRNA:target duplex structures are shown in **Figure 45C**. As exemplified in these structures, all saRNAs are less energetically favourable, as represented by the ~10% increase in MFE relative to the shRNA control. This decrease in stability is seen for all saRNA designs, and the complete table of folded structures tested can be found in **Table 18** of this thesis, in the supplemental material section. These foldings also predicted non-canonical pairing or minimal Watson-Crick pairing between the Abox and bridge sequence for certain saRNAs. For all designs, only one A:U Watson-crick bond is present at any time. Overall, we believed that these interactions were negligible, and did not modify the Abox sequences to further reduce them.

5.3.1.3 Bulged RNAi structures are effective at knocking-down mRNA

To test these designs, the intended active strands were cloned into the DOX-inducible pLKO lentiviral system as RNA hairpins (**Figure 46A**). In addition to the sa/shRNA constructs, a negative control scramble RNA was also constructed in a similar manner. Using these plasmids, we transduced low GFP expressing sorted DU145-GFP-NT3'UTR cells with one sa/shRNA construct. These doubly transduced cell lines would only express their given RNAi molecule when in the presence of DOX. Due to the Ametrine-T2A-Puro protein sequence in the plasmid, we could establish a pure cell line by selection or cell sorting. We opted to generate our cell lines by selection, and validate cell line purity by FACS. The fluorescent marker also allowed us to verify the purity of the cell lines over time, which never substantially changed over their time in culture. Given that all sa/shRNA constructs were of equal lengths and were given the same leading sequence, loop, and RNA Pol-III terminator sequence, we assumed that all constructs are equivalently processed.

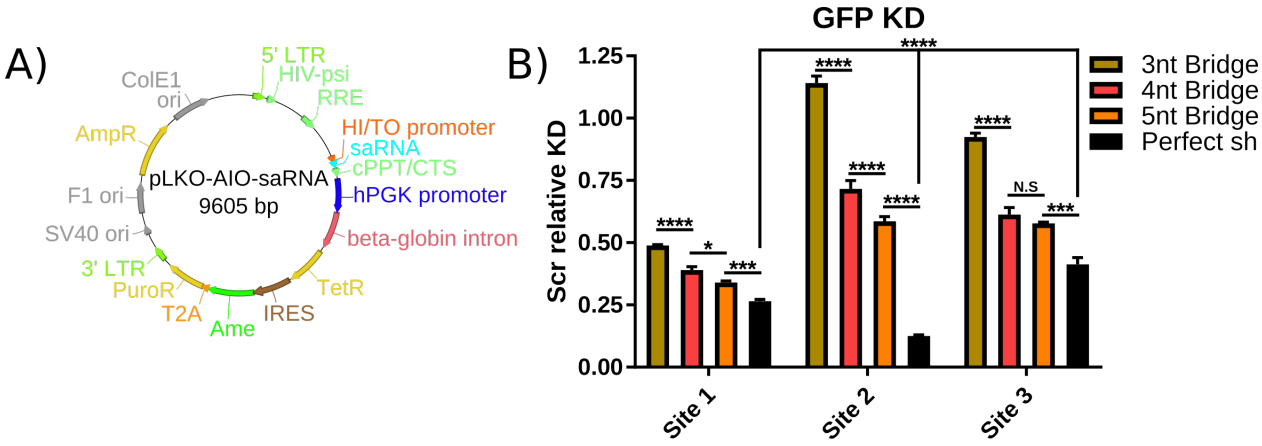

Figure 46: Knock-down efficacy of saRNA designs

A) Plasmid map for DOX inducible sa/shRNA constructs **B**) Knock-down efficiencies of sa/shRNA constructs. Knock-down values are normalized to knock-down effect of Scr expressing DU145-GFPNT3'UTR cells. Error bars represent SD of biological triplicates. Statistical significance calculated using a two-way ANOVA and Tukey's multiple comparison test. *: $p<0.05$; ***: $p<0.0005$; ****: $p<0.0001$; N.S: Not-significant

Once DU145-GFP-NT3'UTR cells were transduced with an sa/shRNA expression plasmid and selected for, cells were incubated with DOX to induce GFP knock-down. This ensured that GFP knock-down began simultaneously for all cell lines. Relative GFP levels for all constructs were collected by FACS after 6 days of DOX incubation, to allow for GFP protein turnover[700]. GFP knock-down was calculated using GFP geometric MFI levels, and normalized to the Scramble control construct (**Figure 46B**). Contrary to our original hypothesis where the 3nt bridge sequences would be the most efficacious, the most effective bridge constructs were the 5nt bridges for all three sites tested. While exact knock-down values for each construct varied from

site to site, the pattern was always consistent; as bridge lengths increased, so did knock-down efficacy.

Between the shRNA controls, there was significant variance between their knock-down efficiencies (**Figure 46B**). This highlighted the importance of including the shRNA controls, as they provided context for how much potential knock-down ability was remaining in the site. Thanks to these controls we could see the 5nt bridge of site 1 was approaching the same knock-down efficacy as the shRNA for site 1, while there was more ground to gain for sites 2 and 3. The relative differences between the knock-down efficacy of the shRNAs can be explained by multiple mechanisms. The differences in knock-down efficacy may be the result of site occlusion from shRNA-loaded AGO2, which can occur due to secondary structure surrounding the shRNA target site, or due to other RNA binding proteins shielding the site directly or indirectly[701,702]. Alternatively, the differences in knock-down efficacy may be due to differences in transcription and processing of the different hairpin RNA sequences[703,704].

To test the local secondary structure hypothesis, the secondary structure surrounding the sites can be determined through folding simulations with MCFF, or can be tested for directly using chemical probing[705]. Occlusion of the target site by RNA binding proteins can be probed using the following modified RNA IP protocol. First, the 3'UTR sequence would be modified by inserting synthetic sequences at the beginning and the end of the 3'UTR. These sequences would correspond to CAS13 crRNA binding sequences. With the expression of this modified 3'UTR, the targeting crRNA, and a nuclease deficient CAS13 protein, an RNA IP is performed on CAS13 after cross-linking[706]. This protocol should purify CAS13, the 3'UTR RNA sequence, and any RNA binding proteins associated to the sequence. Once RNA binding proteins are identified, their binding sites can be identified through deletion analyses of the 3'UTR sequence[707].

Focusing on sites 1 and 3, their respective 5nt bridge constructs were approaching the same knock-down efficacy as the shRNAs for their site (**Figure 46B**). For site 1, all of the saRNAs were more effective than the other site's saRNAs, and the knock-down difference within site 1's saRNAs are much tighter than the other two sites. A potential explanation for this is the exact sequence composition of the saRNAs of site 1. The tail sequence directly adjacent to the A box

in the 3nt bridge construct of site 1 was entirely made of GCs. The end of the seed sequence for site 1 was also relatively GC rich, and ended with a G. This high GC content specifically adjacent to the Abox sequence may act as a sort of GC clamp[708,709]. This may provide sufficient stability to overcome the steric hindrance of a 3nt bridge structure and allow for proper saRNA:target duplex formation and therefore function. If we look at the GC content of the sequences surrounding the 3nt bridge structures across the three sites; we can see that site 1 contains the highest GC content, with site 3 contains the second most, and site 2 contains zero GCs surrounding the Abox sequence. From this, we can hypothesize that Abox adjacent GC content can promote knock-down efficacy in potentially sterically hindered structures such as a 3nt bridge structure. As shown in the work of Gruttadauria et al, increased GC content can promote an increase in maximal bridge length[421]. In addition to the 3nt bridge structure of site 1, the 4 and 5nt bridge structures also maintain a guanine or cytosine directly adjacent to the Abox, further supporting this hypothesis.

Site 2, while maintaining the same overall knock-down pattern of the other two sites, was somewhat of an outlier (**Figure 46B**). While knock-down efficacy did increase with bridge length, 3nt bridge construct for site 2 was particularly ineffective. Upon closer inspection of the 3nt bridge construct for site 2, we realized that by mirroring the target sequence for the Abox region, we introduced a 7nt stretch of adenines (shown in **Table 4**). This adenine stretch, when converted to a hairpin RNA sequence for RNA Pol-III expression, is translated to a 7nt stretch of thymines, which is a strong transcription termination signal for RNA pol-III[710]. It was therefore unlikely that the 3nt bridge construct for site 2 was being properly transcribed, and the lack of knock-down efficacy was due to its lack of processing and not its lack of activity. This expression defect was not as strong in the other saRNA members of site 2 as their Abox sequences were changed to 5'-aaa-3' 5'-aat-3' to prevent interactions with the thymines in the target sequence. This change broke what would have been a 6 thymine stretch in the hairpin sequence into a four thymine stretch. While a four thymine stretch may not be as effective as the classical 6nt polyU stretch, a 4nt stretch has been shown to be 25% as effective[710]. These premature RNA Pol-III termination concerns do not apply to the shRNA for site 2, as its sequence reduced these polyT stretches to a maximum of three, which is insufficient for RNA Pol-III termination[710].

This provided a potential explanation for the knock-down efficacy gap between the 5nt bridge construct for site 2 and the shRNA for site 2, which was exceptionally wide compared to the other sites. It may be the overall expression levels for the saRNAs of site 2 are not comparable to the shRNA levels, resulting in a falsely perceived reduction of knock-down efficacy. To address this, the Abox of these constructs could be changed to a different sequence, or the original saRNAs sequences could be expressed through an alterative method. Alternative expression methods could be achieved using alternative shRNA expression constructs which use RNA Pol-II to drive expression, such as a miR-30 based expression cassette, or through transfection of chemically synthesized RNA. Both methods bypass the RNA Pol-III premature termination issue.

Given the large gap in knock-down efficiency between the 5nt bridge and shRNA constructs for site 2, we saw this as an opportunity to improve our designs by optimizing the Abox sequence. The original Abox sequence for the 5nt bridge construct for site 2 was 5'-aat-3'. This sequence showed some interactions between the bridging sequence, creating an A:U pairing. As we believe that minimizing Abox interactions would lead to increased knock-down efficacy, we designed two alternative 5nt bridge site 2 Abox sequences: CCC and CGC. The CCC Abox sequence removed the possibility of generating any Watson-Crick pairing between the Abox and the bridge sequence, whereas the CGC Abox allowed for a potential G:U wobble base pair to form between the Abox and the bridge sequence. Additionally, both of these new Abox sequences removed the 4nt polyA stretch from the 5nt bridge site 2 design, eliminating any concerns about premature RNA Pol-III transcription termination.

When tested in isolation, there was a clear yet moderate benefit to changing the Abox sequence from the original AAT to either of the new designs (**Figure 47**). However there was no difference between either of the modified Abox sequences, they applied a similar benefit to the knock-down efficacy. Despite this increase in efficacy, there was still significant ground to cover before the saRNAs could reach similar efficacy to the shRNA for this site. This result suggested that minimizing Abox interaction is not a universal rule for increasing knock-down efficacy. Given the low GC content of the centre of the 5nt bridge sequence and its surrounding sequence, perhaps increasing or even maximizing interactions between the saRNA and target

will result in a knock-down efficacy that is more in line with the shRNA for the site. Whether or not to minimize or maximize Abox interactions may depend on the relative GC content of the local target sequence.

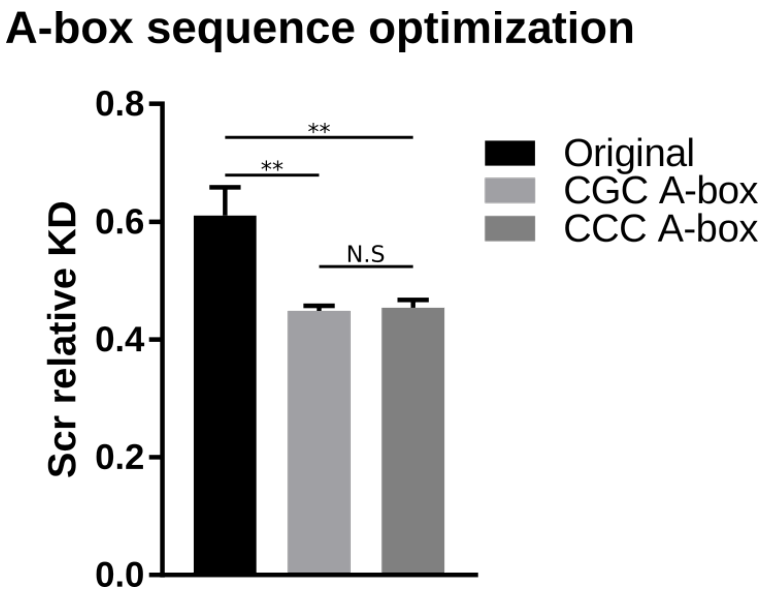

Figure 47: Abox sequence optimization

Effect of modifying the Abox sequence for the 5nt bridge structure of site 2. Knock-down induction and quantification is calculated as in previous figure. Error bars represent SD of biological triplicates. Statistical significance calculated using a one-way ANOVA. **: $p<0.005$; N.S: Non-Significant.

5.3.2 synmiR sites can be identified programatically

Building off our saRNA results, we next sought to apply these findings in a multi-targeted RNAi context. If we could find a set of target sites with the same seed sequence and the right supplemental binding sequences, we could imagine having a single RNAi sequence which binds to all target sites, but forms different duplex structures that are all functional. These target sites could potentially be across multiple target sequences, or all found on the same target sequence. Since we were trying to generate single sequences which formed shRNA-like structures on some

sites, and saRNA-like structures on other, we decided to term these new RNAi molecules synthetic miRs, abbreviated to synmiRs.

As a testing-ground to find and validate potential synmiR designs, we chose to work with the HIV genome. We chose HIV as our design space as the HIV genome is long enough to contain a large variety of potential sites. Additionally, RNAi of HIV has be shown to be ineffective in the past[711,712]. This is due to a combination of mechanisms: site inaccessibility, inhibition of the RNAi mechanisms within host cells, and the high mutational rate during HIV genome replication[712–714]. We hypothesized that synmiR RNAi would be more effective at reducing HIV genome replication than a single shRNA, as the RNAi across multiple sites reduces the probability of a sufficiently mutated HIV genome existing to escape RNAi existing within a single cell. Just as a combinatory approach to small molecule drugs has increased efficacy at controlling HIV infections, so to could combinatory RNAi help to make RNAi of HIV effective. We are attempting combinatory RNAi using a single synmiR to treat HIV.

To generate a reporter construct of HIV transcription and translation, we modified the pNL-R-E- Luc plasmid to express mCherry at the 3' end of the HIV genome (**Figure 48**). This new plasmid was termed pNL-mCherry. The mCherry CDS was inserted into and replaced the Luciferase CDS, which was originally inserted into the HIV genomic location containing *Nef*. The placement of the mCherry CDS ensured the mCherry would be expressed regardless of how the HIV genome was being spliced. The insertion of the mCherry CDS in combination with the existing mutations within the *ENV* and *VPR* sequences made the construct self-transcribing, but non-infectious.

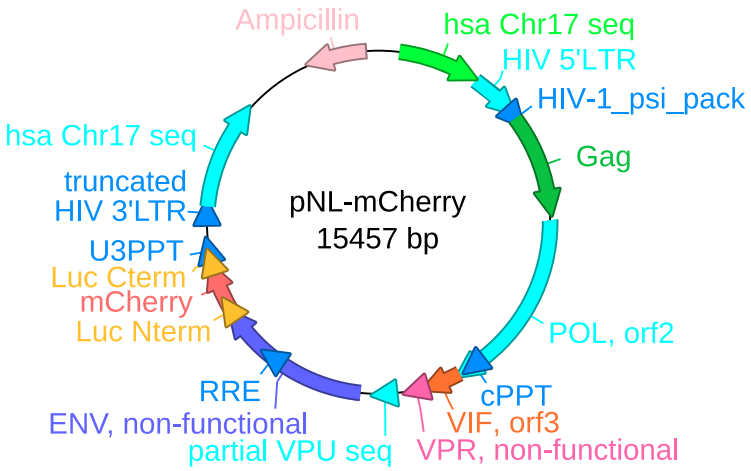

Figure 48: pNL-mCherry reporter plasmid

Plasmid map for non-infective pNL-mCherry reporter plasmid

To design our synmiR sequence, we wrote two different algorithms built upon two different synmiR design philosophies. The first was a minimal-mutation algorithm, starting with an shRNA sequence, the algorithm mutates the shRNA tail sequence such that it retains sufficient binding on the original site, but enhances binding to other sites (**Figure 49A**). The second algorithm was a ration-design algorithm where supplemental binding sequences conserved across a given seed's sites are used to generate synmiR sequences using user-defined parameters (**Figure 49B**).

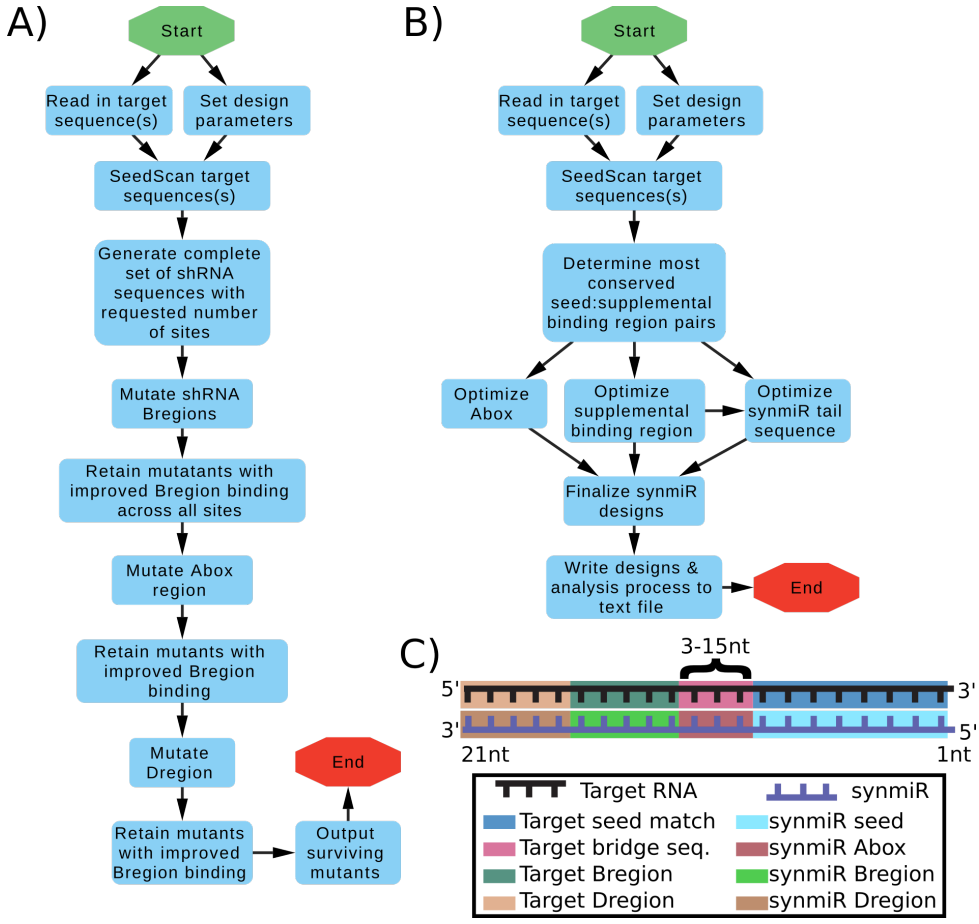

Figure 49: Overview of synmiR design algorithms

A) Flowchart of general design process for Minimal-Mutation synmiR design algorithm. **B**) Flowchart of general design process for Rational-Design algorithm for synmiR design process. **C**) Breakdown of target site sequence and synmiR sections relevant for synmiR design algorithms.

5.3.2.1 Description of the minimal-mutation synmiR design algorithm

The minimal-mutation algorithm, which is written in Java, functions as follows. To begin, a seed analysis is done across the entire target sequence. Based on a user-defined number of sites to target simultaneously, seed sequences which are not frequent enough are discarded. For seed sequences which are more frequent than the requested number of sites, these seeds are re-contextualized as a set of all combinations of their sites with the combination size set to the requested site size.

Using this list of seed:sites combinations, perfect shRNAs are designed for all sites. Each shRNA is then mutated across the Bregion to increase binding at other sites (**Figure 49C**).

Sufficient binding is evaluated by MCFF folding. Once a mutated shRNA contains canonical or non-canonical binding across a contiguous stretch of target sequence for all target sites, the Bregion is no longer mutated. These Bregion mutated shRNAs then undergo a second round of mutation within the Abox (**Figure 49C**). Mutations which further decrease average MFE are retained, so long as they do not disrupt the seed or Bregion binding profile. Doubly mutated shRNAs which pass this filter are used for the final round of mutation on the Dregion (**Figure 49C**). Like with the Abox, Dregion mutations which lower the average MFE across sites while maintaining seed and Bregion binding are kept. Based on our saRNA designs, we set the seed sequence to be synmiR nts 2-8, the Abox 9-11, the Bregion 12-17, and the Dregion 18-21 (**Figure 49C**).

5.3.2.2 Description of the rational-design synmiR design algorithm

The second design algorithm is the ration design algorithm written in Python. Rather than mutating and filtering shRNA sequences, this algorithm finds the set of sites with the most conserved supplemental binding region given a user define maximal bridge sequence length. The algorithm then designs the synmiR sequence in independent sections: the section meant to bind to the conserved supplemental binding region, termed the Bregion; the Abox, which is the synmiR sequence between the seed and the Bregion; and the synmiR Dregion, which is the remaining synmiR tail sequence past the Bregion. The positions of these different regions on the synmiR are all user defined. Based on our saRNA designs, we set the seed sequence to be synmiR nts 2-8, the Abox 9-11, the Bregion 12-16, and the Dregion 17-21 (**Figure 49C**). Pseudo-code and source-code for this algorithm can be found in the supplemental material of this thesis.

The algorithm begins by parsing the target sequence(s) for the identification of these conserved seed:target_Bregion pairs. After reading the target sequence(s) into memory, the algorithm uses a function called seedScan and the target sequence(s) to generate a dictionary of dictionaries; where the top level dictionary key is the seed sequence, and value is dictionary where the key is the target sequence name and the value is a list containing tuples of site data (**Figure 50A**). Examples of this data-structure can be found in the pseudo-code for the ration-design algorithm, found in the supplemental material section of this thesis. Sites are defined as

they were in the saRNA site finder: the maximal sequence for a saRNA with a given bridge sequence range. Sites are also investigated in a similar fashion to the saRNA site finder, using a double-sliding window to generate seed sequence defined lists of potential Bregion binding sequences.

Once the seedScan dictionary is constructed, the algorithm decides how many sites to target simultaneously. If the user provided a multi-FASTA file, the algorithm will design synmiRs which bind to one site per target. If only one sequence is provided, the algorithm will design synmiRs to target sites with the exact number of seed sequence occurrences as the user specified. This means that if the user is designing to target one sequence, they must specify how many sites they want to simultaneously target. Only seeds with the specified number of sites in target sequence will be considered. Seeds which contain a polyA or poly U stretch greater than three are discarded as they will prematurely terminate RNA Pol-III transcription[710]. Based on these design parameters, the seedScan dictionary is filtered down to the valid seeds.

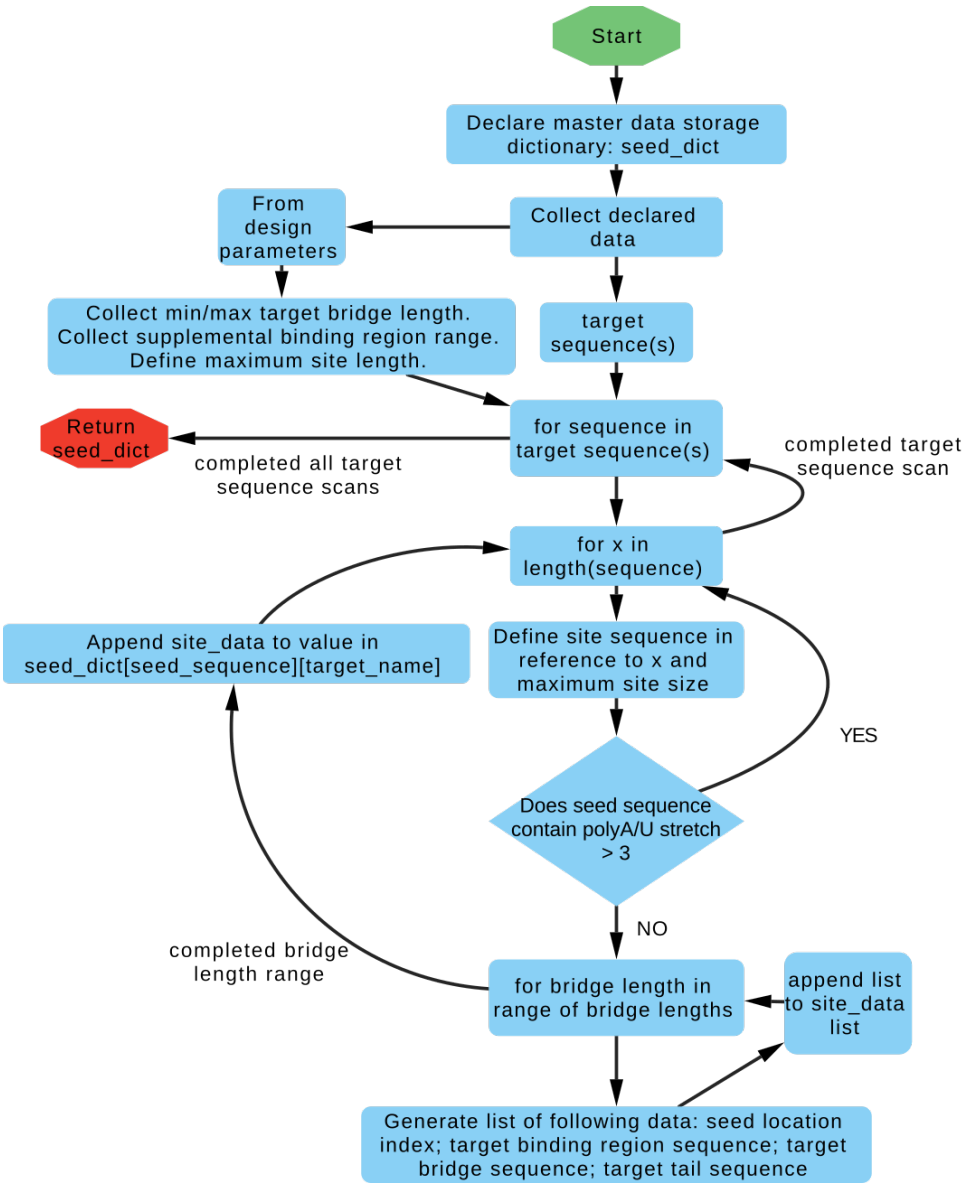

Figure 50: Rational-Design synmiR algorithm SeedScan function

Flowchart of Ration-Design synmiR design algorithm seedScan function.

To determine which seeds contain the most conserved Bregion binding sequences, the algorithm employs a function called combinationExplorer (**Figure 51**). For each seed, combinationExplorer generates all combinations of potential Bregions for a given seed's sites. This means as the number of sites to target increases, the number of combinations to explore becomes exponentially larger. The total number of combinations for a given seed is

Bridge_length_range^number_of_sites. As we are testing these algorithms using the pNL-mCherry reporter for RNAi of the HIV genome, we chose to only target the portion of the HIV genomic RNA corresponding to HIV. A breakdown of number of seeds, their frequency, and the corresponding total number of Bregion binding sequences combinations to evaluate can be found in **Table 17**. To keep the number of sites to validate manageable as well as reduce computation time we chose to target three sites simultaneously, meaning that the algorithm needed to compute and test 743,040 total Bregion sequence combinations. To manage this, the algorithm uses Python's itertools to create a generator for the total number of combinations for a given seed. This generator is then split among a user-defined number of processes in a Python multiprocess pool. This allows for parallel exploration of the total number of combinations while minimizing memory usage, significantly reducing computation requirements.

Table 17: Distribution of pNL-mCherry potential synmiR seed sites

BR = Bridge Range. Maximum Bridge length considered minus minimal bridge length considered.

Sites/seed	Seed frequency in pNL-mCherry	Total number of combinations (BR=12)
1	3608	43,296
2	1217	175,248
3	430	743,040
4	167	3,462,912
5	71	17,667,072
6	34	101,523,456
7	13	465,813,504
8	6	2,579,890,176
9	5	25,798,901,760
10	1	61,917,364,224
11	1	743,008,370,688

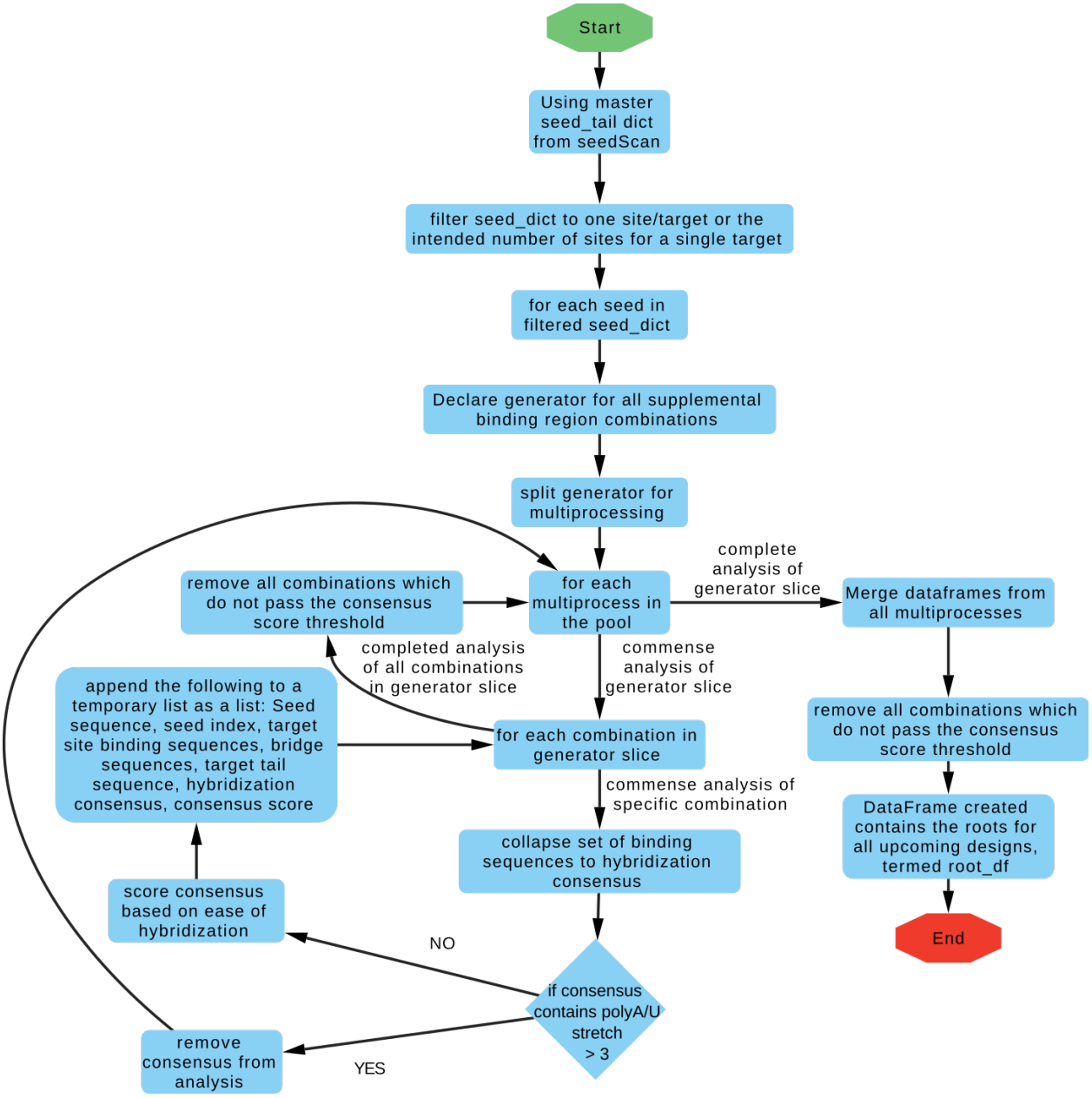

Figure 51: Rational-Design synmiR algorithm combination space exploration

Flowchart of Rational-Design synmiR design algorithm combination space evaluating function.

For each combination of B region binding sequences, a consensus sequence for hybridization to the set of Bregions is constructed (**Figure 52**). This is done through string analysis, where if all of the Bregion binding sequences contain the same nucleotide at the same position, the complementary nucleotide is added to the hybridization consensus. If the nucleotides at the same position are exclusively GA, then U is added to the hybridization consensus. This is because U can bind to A as Watson-Crick, and to G as a GU wobble. Similarly, if the nucleotides at the same position are exclusively CU, then G is added to the hybridization consensus for the same reasons. Otherwise, the complement to the set of nucleotides at the same position is added to the hybridization consensus using IUPAC ambiguous DNA notation.

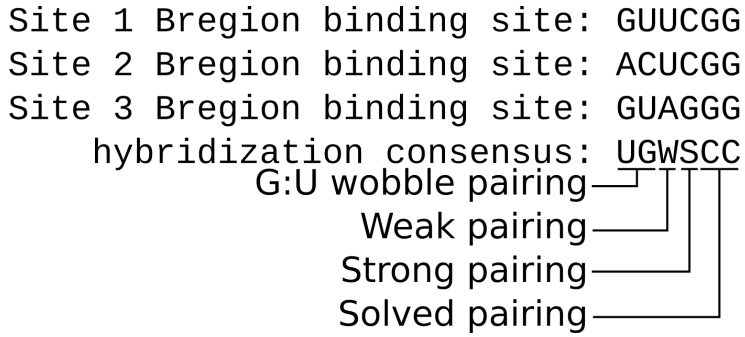

Figure 52: Ration-Design synmiR algorithm Bregion hybridization consensus

Example of hybridization consensus generation for a three-site targeting synmiR. Proposed target Bregions binding sites are aligned, and a consensus sequence which is able to hybridize across all Bregion sites is generated.

As the hybridization consensus sequence is being constructed, it is also simultaneously being scored using a scoring function. Starting with a score of 1.0, the score is multiplied by the weights assigned to each type of pairing added to the positional weight. Nucleotide composition weights are assigned in the following order GC > AU > GU > N2 > N3 > N4 where N2-4 are ambiguous DNA notations meant for two, three, or four nucleotide possibilities respectively[715]. These weights are added to a positional modifier, which increases linearly as the consensus approaches the seed-proximal side of the consensus. The score at each nucleotide are multiplied together for the final score. Exact weights used can be found in the ration-design pseudo-code in the supplemental material of this thesis. As a result, this scoring function heavily

prioritizes GC bonds, and adds additional priority to the seed-proximal side of the consensus. For each Bregion combination, the seed sequence, position of each site in each target, the bridge sequences, the Bregion binding sequences, the hybridization consensus, and the hybridization consensus score are appended to a Pandas DataFrame.

Once the total number of combinations has been scored, only the top scoring combinations are retained for further designs. This is a user-defined cutoff, with a default value set to the top 15% scoring combinations. For each of these seed:Bregion-binding-consensus combinations, the Abox is then designed (**Figure 53**). Essentially, the algorithm attempts to maximize Abox binding when the bridge sequences are equal in length to the Abox sequence. This represents the 3nt bridge structures and shRNAs we tested previously, and based on those results, binding should be maximized when possible, to mimic shRNA binding structure. Binding maximization is determined using MCFF. Using flanking GC sequences and a mask to force the binding of the GC sequences, all possible Abox sequences are tested against all bridge sequences. Maximum binding is measured by minimal average MFE across all bridge sequences. If all bridge sequences are longer than the Abox sequence, then the Abox sequence with the maximum average MFE across all bridge sequences. If the bridge sequences are a mixture of lengths relative to the Abox sequence, then an initial folding of all possible Abox sequences is done, the Abox sequences which maximize binding to the shorter bridge sequences are retained. Of those Abox sequences, the one with the maximum average MFE for non-binding bridge sequences is kept as the solution.

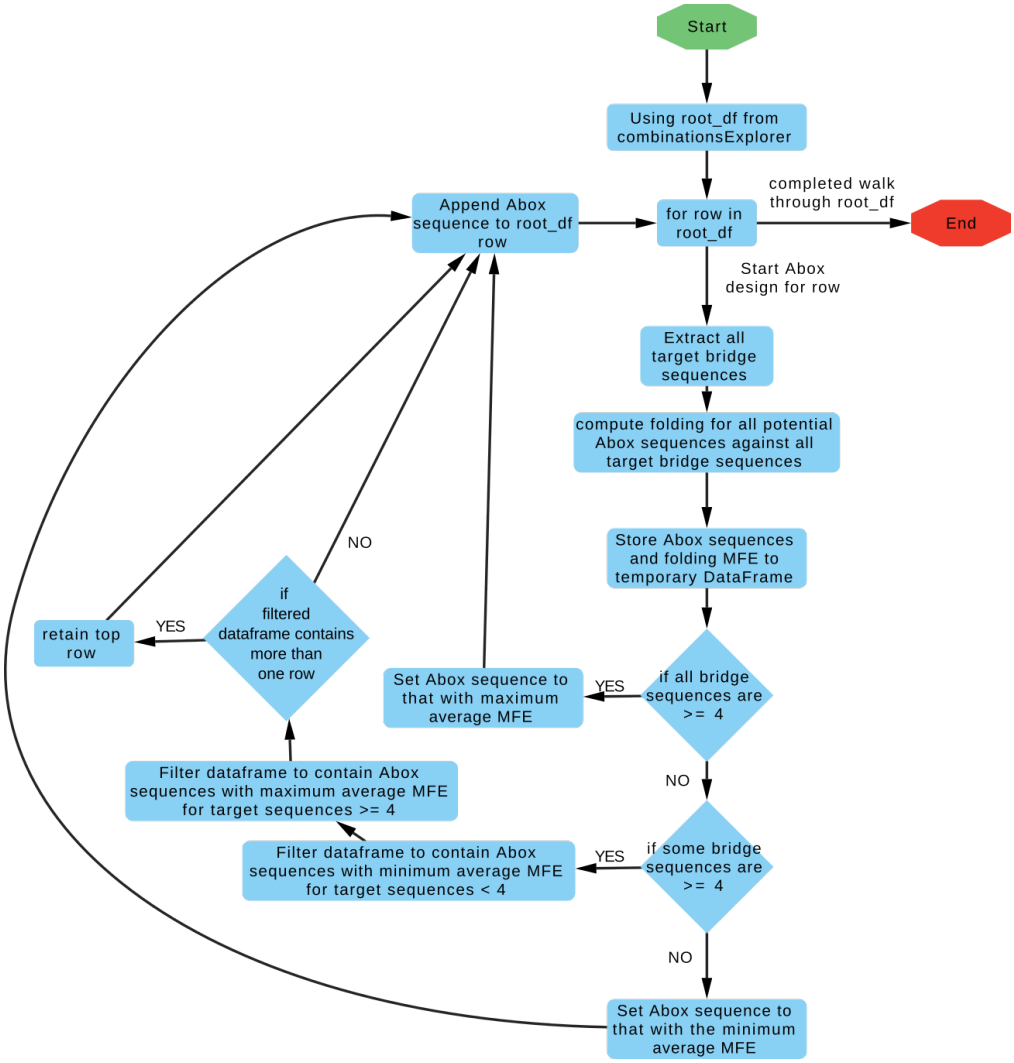

Figure 53: Rational-Design synmiR algorithm Abox design function

Flowchart of Rational-Design synmiR algorithm Abox design function

After determining the Abox sequence, the Bregion is designed (**Figure 54**). If the hybridization consensus does not contain any ambiguous DNA notation, the hybridization consensus is set as the Bregion sequence. Otherwise, the hybridization consensus is de-convoluted into a list of precise DNA sequences. Each potential Bregion sequence is evaluated against all intended Bregion binding sequences, and the ones with the maximum amount of Watson-Crick or GU wobble pairing is retained. In the event of a tie, all Bregion sequences are returned and are used to generate different synmiRs.

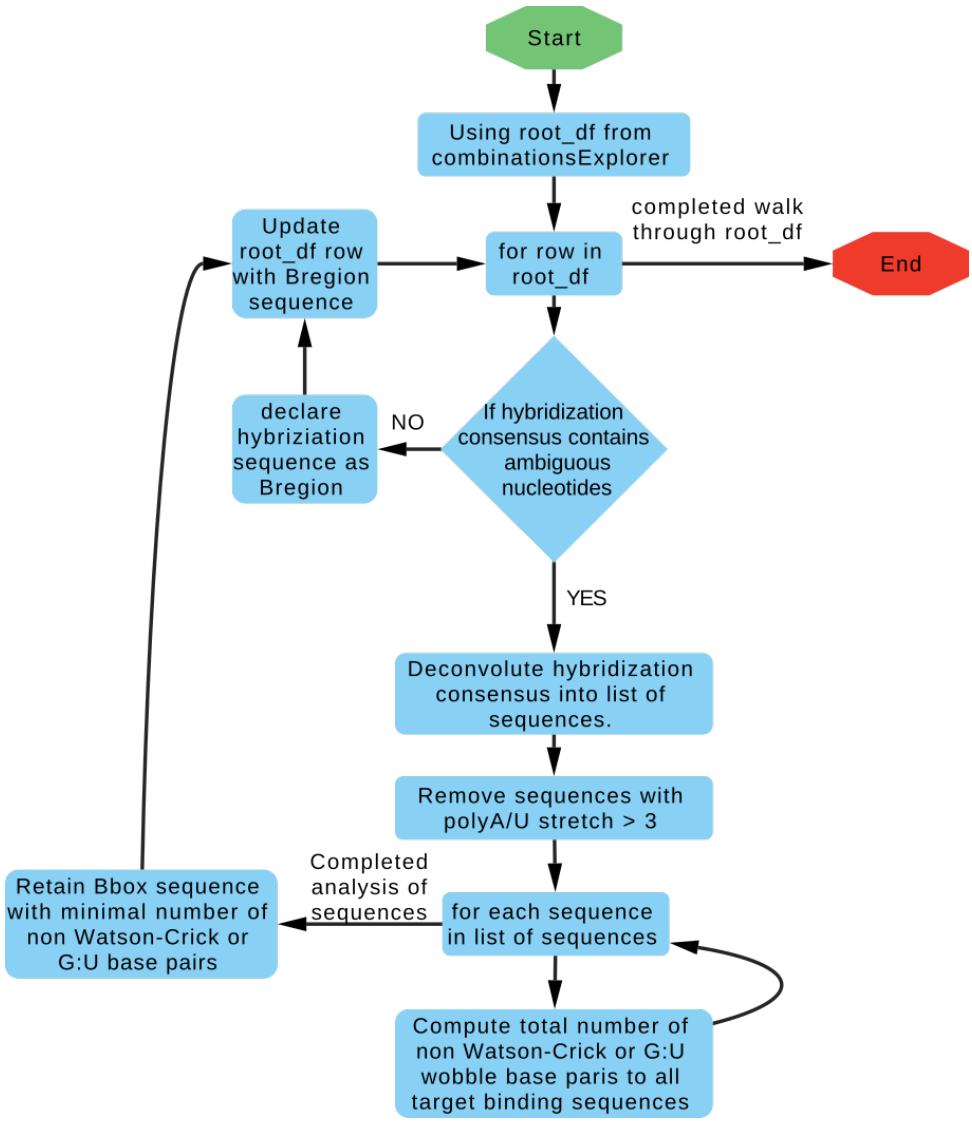

Figure 54: Rational-Design synmiR algorithm Bregion design function

Flowchart of the Rational-Design synmiR algorithm Bregion design function

After designing the Abox and Bregion, the Dregion is designed (**Figure 55**). The Dregion design function takes the Bregion and the intended Dregion binding sequences as inputs, and attempts to maximize binding across all targets. If the Bregion binds well to all Bregion binding sequences (at least 5bp of contiguous binding), the Dregion with the maximum amount of binding across all target sequences is chosen. If any of the Bregion binding sites result in less

than 5nt of contiguous binding to the Bregion, the Dregion sequence is prioritized to bind to that target sequence, to compensate for the weaker Bregion.

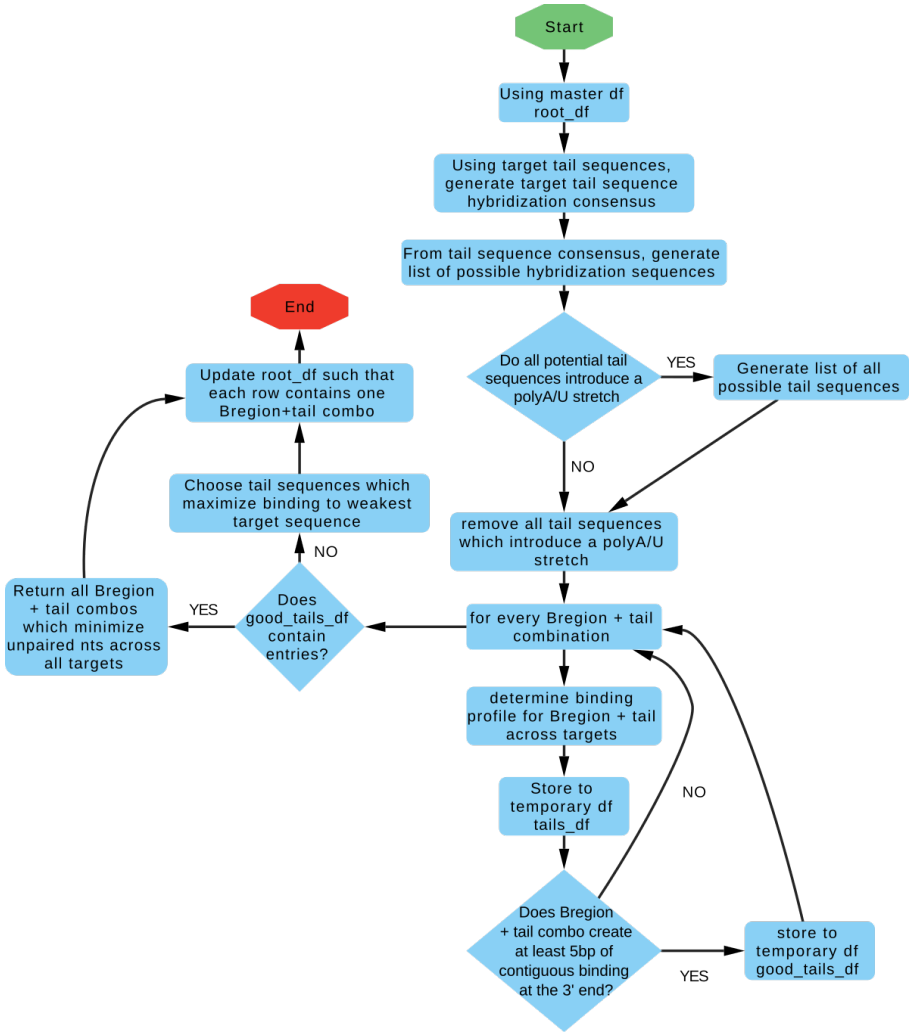

Figure 55: Rational-Design synmiR algorithm Dregion design function

Flowchart of Rational-Design synmiR algorithm Dregion design function

After the Dregion is designed, the entire synmiR sequence is complete. The final design is folded to all intended target sites using MCFF and hybridization diagrams are made in a flat text file for manual evaluation. Note that throughout the design process, any sequence addition that would introduce a polyA or polyU stretch are discarded from the process.

5.3.3 synmiRs show RNAi efficacy at multiple HIV sites

To test our two algorithms, we decided to design synmiRs targeting the same seed sequence. The seed sequence is ACUAUUC, and has three binding sites in the HIV genome (**Figure 56A**). The first site is within *POL*, and is only present in genomic and GAG/POL encoding mRNAs. The second and third sites are within the end of *ENV*, and are found in all HIV RNA species. This seed sequence results in alternative targeting of the HIV genome in comparison to miB, a site which has been shown to be effective at initially controlling HIV infection, before HIV escapes (**Figure 56A**)[711].

After evaluating the outputs the algorithms on these sites, we decided to test two designs from the minimal-mutation algorithm, termed MM-1 and MM-2; and one from the rational design algorithm, termed RD-1 (**Figure 56B&C**). Despite being designed to target the same sites, the tail sequences for all three synmiRs are very different, yet they all contain a bound Bregion across all sites. Regarding MM-1, it contains a shRNA-like structure on site 1, while containing saRNA-like structures on sites 2 and 3 (**Figure 56B**, top). MM-2 contains saRNA-like structures for all three sites, with 9nt long bridge sequence on site 3 (**Figure 56B**, bottom). RD-1 adopts a extremely shRNA-like structure on site 1, a 12nt long bridge sequence on site 2, and a Abox-bound bridged structure on site 3 (**Figure 53C**). synmiR sequences as well as control shRNAs for each site were cloned into pLKO-Ametrine and were expressed as shRNAs under a U6 promoter.

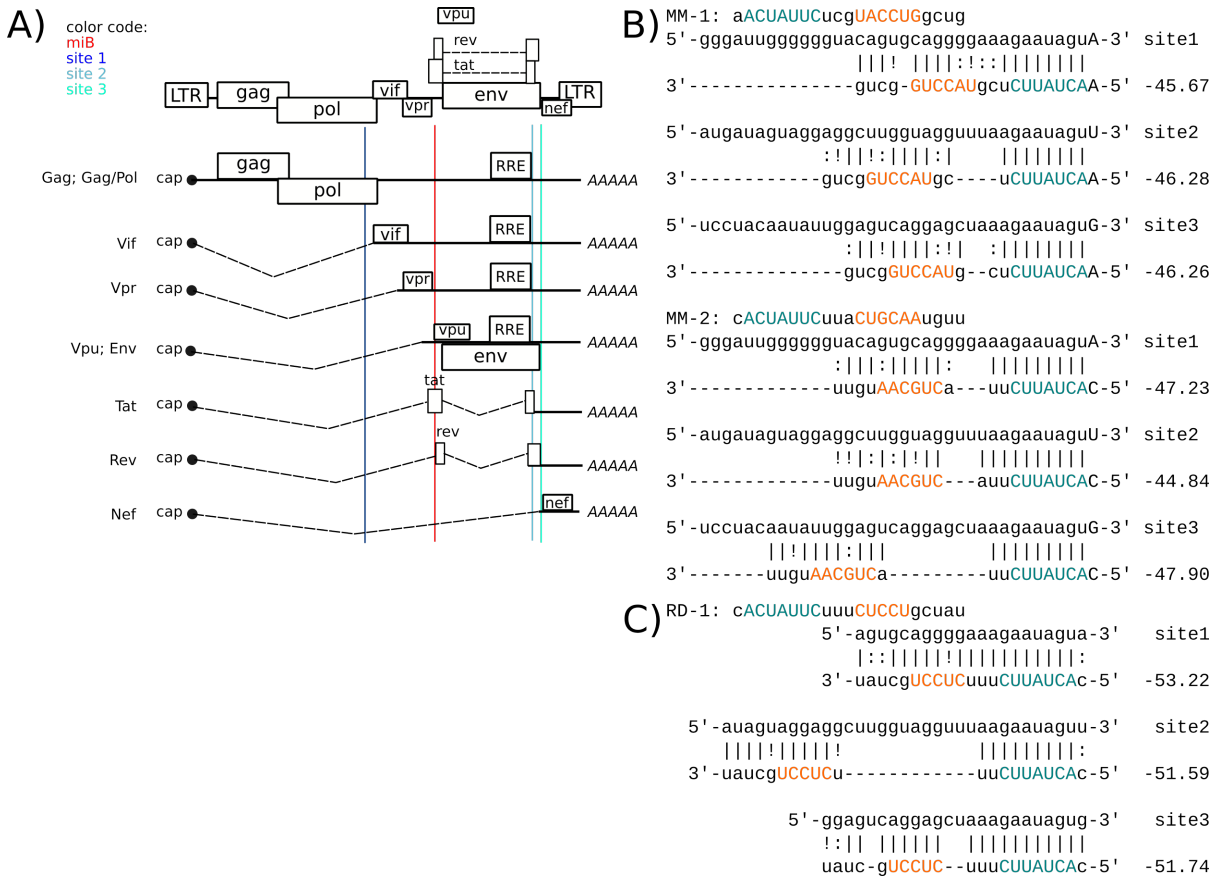

Figure 56: synmiR genomic sites and duplex structures

A) synmiR site locations for HIV genome and mRNA isoforms. **B**) MCFF folded structures for Minimal-Mutation (MM) symiR designs **C**) MCFF folded structures for Rational-Design symiR design **B,C**) Top strand corresponds to site sequence. SynmiR regions colour-coded: leading nt, Abox, and Dregion; black. Seed sequence; teal. Bregion; orange. dG of folding shown beneath site number in kcal/mol. "|" denotes Watson-crick pairing; "!" denotes GU wobble pairing; ":" denotes all other non-canonical pairing.

To determine the relative contribution of each site to the overall knock-down efficacy of the synmiRs, we generated variants of the pNL-mCherry construct which contained a additional repeat of one of the intended target sites (**Figure 57A&B**). These site sequences were inserted into the pNL-mCherry construct directly before the mCherry CDS (**Figure 57A**). This area was chosen for its ease of access while cloning and proximity to the mCherry CDS. Based on the pNL-

mCherry sequence, mCherry translation is likely beginning from a upstream in-frame start codon, as this start codon was originally used to drive translation of Luciferase. Insertion of the site sequences therefore needed to maintain the ORF and no introduce any premature stop codons. For sites 1 and 2 this caused no issue, however site 3 needed to be inserted at an offset to avoid inserting a stop codon (**Figure 57B**). Site sequences 33nt long were inserted, providing more sequence than needed based on the MCFF foldings. As a negative control, a non-targeting sequence was also designed using a specially coded Python algorithm. This sequence was derived of codons not found in any of the synmiR sequences, ensuring that only 2nt of contiguous binding to any synmiR could ever occur.

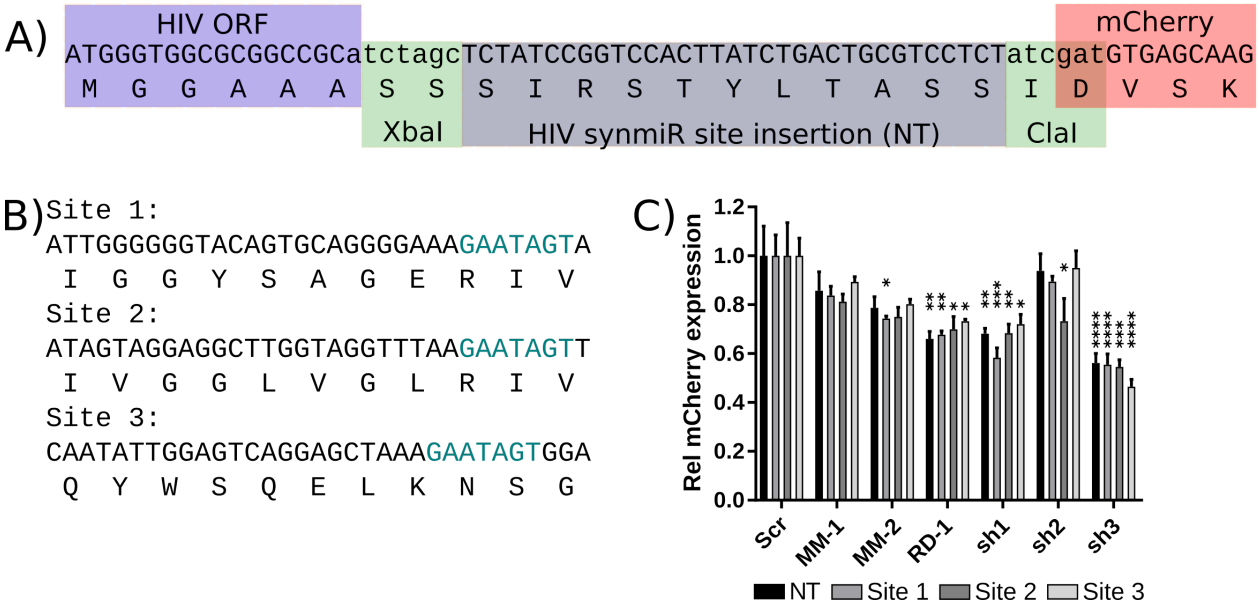

Figure 57: Insertion of synmiR sites and pNL-mCherry knock-down efficiencies

A) Description of synmiR site insertion into the mCherry ORF. Sites are inserted between the XbaI/ClaI sites, NT sequence is shown as an example **B**) synmiR site insertion sequences fro sites 1, 2, and 3. Seed sequence is highlighted in teal **C**) Relative knock-down abilities of synmiR designs, normalized to Scr negative control construct. Stars indicate statistical significance relative to the same pNL construct with Scr transfection, as calculated by two-way ANOVA with Tukey multiple comparisons test. *: p<0.05, **: p<0.005, ***: p<0.0005, ****: p<0.0001

Once the modified pNL-mCherry reporters were built, HEK-WT cells were transfected with pLKO-Ametrine synmiR. The following day, they were transfected with a specific pNL-

mCherry construct. 48 hours later, mCherry and Ametrine levels were measured by FACS. To determine HIV knock-down, a ratio of mCherry MFI in Ame+/Ame- cells is made, and normalized to the same ratio derived from cells transfected with a negative control scramble shRNA.

From this experiment, we can see that all synmiRs and control shRNAs were able to knock-down HIV to varying degrees (**Figure 57C**). For the control shRNAs we can see a increase in knock-down efficacy when the shRNA site is added to the mCherry CDS, confirming the shRNA are specific to their respective sites. While this increase in knock-down efficacy was not determined to be significant, the pattern of control shRNAs having increased knock-down efficacy specifically when their local site is added to the mCherry CDS is nonetheless striking. The shRNA results also determine the relative availability of each site in its endogenous context. For example, we can see the shRNA for site 2 only shows knock-down effect when the site is introduced into the mCherry CDS (**Figure 57C**). In all other pNL-mCherry constructs, in which site 2 is only present in the endogenous HIV genome, sh2 has no effect. This is in contrast with site 3, which shows the best knock-down under endogenous site expression conditions, and a slight additive effect when the site is re-introduced in the mCherry CDS. This is due to site 3 being present on all HIV mRNA species (**Figure 56A**), and is therefore the most abundant and targetable.

All of the synmiRs showed moderate knock-down ability of the HIV genome. However, only RD-1 showed consistently statistically significant knock-down effects across all targets. While MM-1 and MM-2 did knock-down the various pNL constructs by ~10-20%, these results were not statistically significant relative to the Scr control. The only time the MM series synmiRs achieved statistical significance was with MM-2 targeting site 1. This is despite the fact that the knock-down efficiencies for MM-2 are relatively similar. What separates the MM-2:pNL-site-1 knock-down values from the rest of the MM-2 knock-down values is its relatively low variance. This suggests that if replicate variance could be reduced, MM-2 and perhaps MM-1 may reach statistical significance. This could be addressed through the addition of more replicates.

Unlike the shRNAs, the addition of a new site did not substantially increase the synmiR knock-down efficacy. This lack of increased knock-down efficacy may be because the additive

benefit of a single additional site is too minor to quantify in this experimental context. When control shRNAs knock-down pNL constructs with an additional site, the increased knock-down effect is nowhere close to additive. The added benefit is approximately an additional 10-20% knock-down efficacy depending on the site. If this is the benefit observed in the "best-case" scenario of RNAi, it may be that the effects on synmiR activity are beyond the observable resolution of this experiment. To address this, we would need to be targeting sites with a higher knock-down efficacy, in order to see more subtle effects. These could be determined using shRNAs as an initial screen, and then designing synmiRs based on these highly targetable sites.

While there were moderate increases in knock-down ability of MM-1 and MM-2 when sites 1 and 2 are introduced into the mCherry CDS, these increases are non-significant. The most effective synmiR design was RD-1, whose average knock-down efficacy was similar to that of sh-1. As the sequence of RD-1 was very similar to sh-1, with only three nucleotides differentiating them, this was not very surprising (**Figure 56C**). However, given the lack of increased efficacy of RD-1 when site 1 is added to the mCherry CDS, it suggests that RD-1 was not causing target sequence cleavage but mRNA repression, like a miR.

5.4 Chapter Summary & Discussion

In this chapter, we began by building on previous work done in the Major lab demonstrating specific nucleotide regions within the shRNA tail sequence had differing importance for RNAi activity. These results showed the central region of the saRNA, nucleotides 9-11, were relatively dispensable with regards to RNAi[419]. This was in line with previous analyses of miRNA:target duplex structures[416,417]. To refine this model, we investigated if the shRNA tail sequence from nt 12-21 could bind further away on the target sequence than the 3nt found in a shRNA structure. Due to our presumption that shRNA:target duplexes are the ideal RNAi structure, we assumed that as the target bridging sequence increased, knock-down efficacy would decrease. After testing these structures in a 3'UTR context, we found the opposite to be true. While shRNA structures remained the most effective, we found a consistent pattern where as the target bridge sequence increased, knock-down efficacy increased.

These results raised several questions. First, through what mechanism of action are the saRNAs generating their knock-down effect? As already established in the literature, AGO2 RNAi can occur through two mechanisms of action: direct target sequence cleavage or through translational repression and degradation of the target mRNA via the recruitment of endogenous factors. The quickest way to determine which of these mechanisms is at play is to repeat the experiment and collect the cellular RNA for qPCR. Using a qPCR probe-set upstream, spanning, and downstream of the various target sites, we can determine if saRNA expression results in a significant reduction of GFPNT 3'UTR mRNA. Should this be the case, it would suggest that the saRNAs are exhibiting their effects through target cleavage rather than repression and degradation.

A direct way to determine the mechanism of action for saRNA effect would be translate the experiment into a purely *in vitro* setting. Purified AGO2 protein loaded with different sa/shRNA molecules are allowed to incubate with *in vitro* transcribed GFPNT3'UTR RNA. After incubation, GFPNT3'UTR RNA cleavage is quantified via qPCR using probe sets which span the relevant target sites. A GFP CDS binding probe-set can be used as a internal control for overall GFP3'UTR quantity. While more technically demanding, this experiment would allow for direct detection of AGO2 cleavage.

In addition to the mechanism of action at play, there are also other interesting questions, such as the limits to target bridge lengths in these structures. Recent work by Sheu-Gruttadauria et al and McGeary et al have investigated this aspect of saRNA structures through differing methods[418,421]. Through *in vitro* studies of AGO2, Sheu-Gruttadauria et al demonstrated that these bridging sequences can be up to 15nt long, considerably longer than what we tested[421]. McGeary et al found similar results, however they added the context of reduced affinity as bridge lengths increase past a certain point[418]. What both of these studies lacked were evidence of the efficacy of these structures *in vivo*, something we establish through our saRNA experiment. Recent work by Duan et al, also describe a similar model of miR-tail functionality to us, however we are approaching the problem from a predictive rather than descriptive perspective[420].

A natural extension of our saRNA experiment would be to test increasing saRNA bridge lengths up to 15nt long to determine the full extent of bridge length on RNAi in an *in vivo* setting. With longer bridge sequences in mind, stable secondary structures within the bridge sequence become a possibility. These structures may hinder RNAi as it may block AGO2 cleavage or generate enough steric hindrance to block miR-tail binding. Alternatively, a secondary structure in the bridging sequence may allow for miR-tail binding to occur further down from the seed sequence than predicted. Testing these structures using our model would require modifying the NT3'UTR sequence. By using our most stable site, site 1, we could generate variants of the site where we maintain the same seed and supplemental binding sequences, but we alter the bridge sequence length and composition. Using this new model, we can test bridge sequences which form hairpin structures, and determine how these affect knock-down efficacy. Length of the bridge sequence hairpin and GC content should be considered as well.

For our synmiR design algorithms, while our initial experiments did demonstrate reasonable efficacy across multiple sites, further work is needed. To gain insight into the efficiencies of these designs, their target sites should be placed in the context of a 3'UTR, as was done in the saRNA experiment. This would likely increase the efficacy of our designs, as they are most probably acting through the recruitment of translational repression and mRNA decay factors due to their non-shRNA-like structures[361].

To test our synmiRs in a 3'UTR context, a complete re-design of the reporter system is needed. Rather than using HIV as our target, a single lentiviral plasmid would be constructed, encoding three fluorescent reporters (Ametrine, GFP, mCherry) on separate mRNAs, each with their own 3'UTR. The HIV target sequences would then be cloned into the different 3'UTRs, with one HIV target site per UTR. Using this method, we would be able to more precisely determine which sites are more efficiently targeted by our synmiR designs. We may also see heightened efficiencies in our synmiR designs by shifting their sites to a 3'UTR context. Alternatively, applying our algorithms to design synmiRs targeting endogenous genes will allow us to further validate these programs as well as explore potential therapeutic opportunities. However, regulatory feedback loops on endogenous targets may cast doubt onto any knock-down results.

Endogenous targets should be carefully chosen based on their known regulatory mechanisms, as well as the effect of their knock-down through shRNAs.

To increase the efficacy of our designs on the HIV genome itself, we should re-evaluate our seed choice. Our seed was chosen based on the synmiR sequences and structures it generated. Instead, we should have chosen sites which have already shown to be biologically relevant, such as the miB site, sections of the HIV genome which are known to mutate slowly, or the LTR sequences[711,716–718]. By choosing sites which are known to be endogenously available for RNAi, or not involved in protein translation, we may be able to boost the efficacy of our synmiRs.

In future synmiR experiments, in addition to shRNA positive controls as well as a Scr negative control, a synmiR-tail Scr control should also be added. This control RNAi molecule would have the same seed sequence as the synmiRs being tested, but contain a random tail sequence. By including this control, we can distinguish the specific effect of synmiR tails designs have over a non-binding tail sequence. Designing an optimal non-binding tail sequence across multiple sites might be difficult, so using multiple random tail sequences may be necessary. Through the results we have already generated, we are contributing to our understanding of the mechanism of action for miRNAs, the potential off-target effects of shRNAs, and are helping to advance the field of RNA therapeutics.

Chapter 6 – Discussion & Conclusions

6.1 Executive Summary

Given the known role of hypoxia adaptation in exacerbating a patient's cancer prognosis, it is vital to gain a deeper understanding of how this adaptive response functions. A better understanding of the breadth of hypoxic responses as well as their limits will result in a higher efficacy of addressing them in a therapeutic context. The hypoxic microenvironment *in vivo* never occurs in isolation, and always correlates with other physiological stresses such as metabolic stress. *In vivo* hypoxia also develops over an extended period of time and reaches anoxic levels. To parse out how these dimensions affect hypoxia adaptation, we developed two novel *in vitro* hypoxic incubation protocols, which revealed distinct adaptation patterns which could be distinguished by distinct DEG and DEmiR signatures. As the differential expression of microRNAs are also a core aspect of hypoxia adaptation, we also investigated miR tail sequence functionality in a fashion which will not only lead to a better understanding of the miR mechanism of action, but will directly lead to novel RNA-based therapeutics to target these cancer specific adaptions.

This PhD thesis contains several novel findings stemming from attempting to understand and counter the hypoxia adaptation response seen in cancer. Using our hypoxia adaptation reporter cell line, we demonstrated how hypoxic incubation in conjunction with metabolic stress leads to an enhanced level of HIF1α stabilization. That this cell culture protocol induces a hypoxia adaptation response that correlates with hypoxia adaptation responses seen in patient samples, such as the downregulation of miR-7a and the enrichment of cell-motility based phenotypes.

To better simulate the gradual development of the hypoxic microenvironment *in vivo* we established the LTHY protocol, which induced a novel and spontaneous EMT-like response in our cells. Based on transcriptomic analyses of our data, we found WT1 to be a potential driver of our EMT program. By adjusting the hypoxic incubation protocol, we found that WT1 upregulation required pre-incubation at milder hypoxic environments as well as hypoxia levels below 1% O_2. To date, this is the first *in vitro* hypoxic incubation protocol that combines long-

term incubation with progressive lowering of oxygen levels, which produced unique cellular adaptation yet to be described in the literature.

Through RNAseq investigations, we discovered that our cells were expressing a novel transcript of WT1, tWT1, something which has not been described in mice to date. Molecular biology experiments on this novel transcript proved that it did generate protein, that it had proper subcellular localization, and retained its DNA binding ability to cell motility genes. However, due to an inability to generate a WT1 deficient cell line, we could not validate these genes as genuine targets of our tWT1. Identification of the human ortholog of this tWT1 expressed in Leucegene patient samples strengthened the validity of this isoform, and suggests a potential common drive to express this isoform across both cancer types and host species.

To further our understanding of RNAi and the function of the tail sequence in shRNAs and miRNA, we investigated how tail sequences can be manipulated while still maintaining functionality in an *in vivo* context. We demonstrated not only can shRNA tail sequences bind at positive offset from the seed, but these structures are more effective at reducing gene expression when their binding offset is increased. Using this as a foundation, we wrote algorithms which attempt to design multi-targeting RNAi molecules based on ours and others works[413,418,419,421,710].

6.2 Discussion of short-term hypoxic adaptation in conjunction with metabolic stress

6.2.1 Development of *in vivo* competent hypoxia reporter cells

As the core goal of this project was to more precisely delineate the tumoural hypoxia response, a natural extension of our work would be to purify hypoxic tumour cells from a primary tumour for transcriptomic analyses. Through the analyses of these cells, we would be able to use our *in vitro* RNAseq as signatures to parse out the relative types of hypoxic adaptations. Our *in vitro* hypoxic incubation protocols already provided us insight with two different axes of hypoxic adaptation: metabolic adaptation and hypoxia induced EMT.

We did attempt to sort hypoxic tumour cells from B16-HG primary tumours grown in mice, however the overall level of GFP MFI was lower than cells grown under tissue culture conditions. This was likely due to the overall melanin content of the primary tumour cells, which was significantly greater than in cells grown under tissue culture conditions. This increased melanin content more significantly affected GFP fluorescence levels than mCherry, as melanin blocks more radiation at the higher end of the spectrum than at the lower end.

An additional hurdle when using our B16-HG cell line *in vivo* was the overall dynamic range of the GFP signal. Under standard hypoxic incubation conditions, the GFP signal only increased two-fold relative to standard tissue culture levels. When examining pure populations of hypoxic or normoxic cells, this degree of separation is sufficient for analysis. However when mixed together, this range of signalling is insufficient to accurately separate the two populations. Therefore an *in vivo* hypoxia adaption reporter needs higher dynamic range than we were able to produce.

There are several avenues we can take to achieve a *in vivo* capable hypoxia reporter cell line. As shown in other works, *in vivo* hypoxia reporter cell lines can be established using HRE driven reporter constructs[719,720]. The added signal amplification of the HRE constructs should allow for greater dynamic range under a spectrum of *in vivo* hypoxic conditions and allow for accurate cell sorting based on GFP expression. An additional consideration is the oxygen-requirement of GFP to generate its chromophore[721,722]. While this requirement does not affect GFP fluorescence levels at *in vitro* hypoxic levels, it may become a factor under *in vivo* conditions, where tumour cores are frequently anoxic[26,719].

Instead, the fluorophore should be changed to one which does not require oxygen for proper function, such as mCherry or UnaG[723]. By swapping to UnaG, not only do we eliminate the possibility of sorting GFP- cells which are in fact expressing GFP, but we also reduce the overall size of the construct, as UnaG is about half the size of eGFP[723]. This would allow us to include mRNA stabilizing sequences such as a polyA tail, repeat the reporter construct multiple times in the same lentiviral construct to increase signal output, or both. These modifications would lead to increased signalling under hypoxic conditions due to increased mRNA stability due

to inclusion of a polyA tail or Woodchuck hepatitis virus Post-transcriptional Regulatory Element (WPRE), reduced induction time due to the shortened mRNA sequence, and increased mRNA output due to repeats of the reporter construct. Because of the lack of fluorescence when the construct is not induced, a second marker must be included to differentiate the normoxic tumour cells from the host cells. This can be accomplished using a second plasmid like we used for our existing hypoxia adaptation reporter cell lines, or could be incorporated into the HRE plasmid if there is sufficient space. Surface markers should not be used to mark tumour cells, as this may affect their interactions with the tumour microenvironment and other cells.

An ideal tumour hypoxia reporter system would include a read-out of HIF1α protein stability as well as transcriptional activity. Having such a reporter would allow us to differentiate between tumour cells in a HIF1 stabilizing microenvironment and those which are actively engaging in hypoxic adaptation, a distinction which may be critical to accurately dissecting the hypoxia adaptation landscape. Using the modified HRE construct as previously described, cells would be transduced a second and third time using a HIF1α-ODD-mCherry reporter construct as an additional tracker for HIF1α protein stability. As a general marker for tumour cells, these dual reporter cells would be transduced a final time for steady-state expression of Ametrine. Ametrine should be used as the steady state marker, as it has the lowest prevalence for fluorescent spillover into GFP or mCherry detection. Using this triple fluorescent reporter cell line, we would be able to distinguish between tumour cells that are able to stabilize HIF1, those which are actively adapting to hypoxia, and those which are experiencing normoxia. We would also be able to detect tumour cells transcribing HRE sensitive genes in a HIF1α negative context.

All of the reporter constructs thus described report on if a cell is actively engaging in hypoxia adaptation. Recently, Godet et al developed a hypoxia adaptation tracking system[174]. In this system, cells express a CRE recombinase system linked to the HIF1α ODD domain, as well as a LoxP flanked mCherry CDS and a eGFP CDS downstream of the second LoxP sequence. Under normoxic conditions, CRE is degraded and cells are mCherry+/GFP-. Under hypoxic conditions, CRE is stabilized and recombines the two LoxP sites, removing the mCherry CDS, causing the cells to permanently shift from mCherry+/GFP- to mCherry-/GFP+. As such, this system allows for the tracking of cells which have been exposed to a hypoxic microenvironment at any point.

Through a rearrangement of fluorophores, we could combine this system with our active hypoxia-adaptation reporters. By changing the CRE sensitive fluorescent switch from mCherry/GFP to mCherry/IFP, we keep a large amount of spectral space free for other markers, such as our HRE-UnaG reporter system. This cell line would allow us to distinguish between tumoural cells adapting to hypoxia for the first time, those which have already experienced it, and those which are actively adapting to it. This would allow us to test whether hypoxia adaptation changes in cells after they have already been exposed to it. It is known that tumoural hypoxia can be sporadic based on the ineffective local vasculature, however little is known on how cyclic hypoxia effect tumour cells *in vivo*[255]. It is not inconceivable that repeated exposure to hypoxic environments would render the exposed cells increasingly resistant to hypoxia and standard therapeutics. Understanding how they reach this functional state is vital to developing truly effective anti-cancer therapeutics.

To elucidate how cyclic hypoxia influences tumour cell adaptation in a more controlled setting, we can use our *in vitro* hypoxia incubation chamber. As this system is completely isolated, we can grow cancer cells under a variety of different hypoxic environments, and switch between hypoxic and normoxic oxygen levels in a matter of minutes. This gives us a decent ability to recreate the *in vivo* intermittent hypoxia[255]. Determining what oxygen levels to cycle between, the length of time at each level, and whether or not to include metabolic stress in conjunction with hypoxic stress cycles all need to be considered for this experiment. To best simulate *in vivo* intermittent hypoxia, cells should first be incubated at milder hypoxic levels as was done in our LTHY protocol. After this timepoint, cells would then enter a hypoxic cycle of 0.5% O_2 for 24 hours in RPMI 0% FBS pre-equilibrated to 0.5% O_2.

This would combine both stresses which occur during *in vivo* hypoxia: a lack of oxygen, nutrients and growth factors simultaneously. After 24 hours, oxygen would then be raised back to 7% and media would be replaced with complete media for another cycle of tissue normoxia. These cycles would then be repeated for two weeks, with cells collected after every cycle for analysis. Ideally, cells would be analyzed through deep transcriptome profiling with a 21% O_2 grown sample as a control. In the absence of this, qPCR on a panel of hypoxia adaptation genes as well as EMT markers can be used to determine relative hypoxia-adaptation.

Rather than generate an *in vivo*-competent cell line, an alternative strategy would be to detect transcriptional adaptions *in vivo* using spatial transcriptomics. Using spatial transcriptomic techniques such as seqFISH, it would be possible to identify hypoxic regions of the tumor using the RNA of the HRE-based reporter (as well as canonical hypoxia markers such as CAR9), and then correlate those regions with the metabolic changes found in the MHY treatments[724]. Using the RHY-specific adaptations as a reference, we could identify sub-regions of the tumor that are experiencing both hypoxia and growth-factor starvation. This would aide in validating the RHY incubation protocol as being more representative of *in vivo* tumor microenvironments, as well as provide biomarkers for these more stressed cells in patient samples.

6.2.2 Improvements and modifications to the MHY protocol

In Chapter 3, we demonstrated how the inclusion of metabolic stress in conjunction with hypoxic incubation can lead to increased levels of HIF1α stabilization. Transcriptomic analyses of cells exposed to these conditions demonstrated that this protocol induced distinct metabolic changes, such as the upregulation of the Rapoport-Luebering Glycolytic shunt and cytosolic isocitrate metabolism. This metabolic stress was accomplished through changing the media used from DMEM to RPMI and removing all FBS. This shift was meant to simulate the deficit of nutrients and growth factors found in the hypoxic microenvironment of the tumour. While it is clear that this media change resulted in distinct metabolic changes, the reasons why, and the exact thresholds for inducing these changes are not known. A natural extension of this experiment would be to determine the exact amount of growth factors needed for the synergistic effect seen in our data (**Figure 11**). While we know that 0% FBS induced the changes we have observed, we do not know if that is the necessary threshold to induce those changes.

Additionally, we do not know if the contribution of the FBS is due to its growth hormone content or its nutrient content. There is a body of literature suggesting that if glucose levels drop below a critical level, HIF1α protein levels and transcriptional activity is abolished[178,725–727]. Vordermark et al determined this glucose level to be between 0.55-5.5mM, about half of the listed glucose concentration in RPMI (ThermoFisher)[725]. Based on this work, we can assume the

the enhanced HIF1α levels observed in the RHY condition are the result of the lack of growth factors provided by the FBS, not the nutritional content. The issue with examining this is inherent growth factor and nutrient variability between FBS batches. This would require determining optimal levels of the major growth factors and hormones found in FBS (TGF1, GGF, IGFBP2, IGFII, CAA33746.1, bFGF, IGFBP2, IGFBP4, EGF, PDGF, IGF-I/II, FGF, IL-1, IL-6, Hydrocortisone, Insulin, Triiodothyronine, Thyroxine)[728,729] and then determining which of these factors are essential for regulating out observed metabolic changes. Quite the panel if a brute-force combinatorial approach is taken.

Extending this line of experimentation, we could also consider more specifically modulating the nutrient composition of the media during hypoxic incubation. It is known that glucose is a key energy providing nutrient under standard and hypoxic incubation conditions. We have shown that B16 cells can survive short-term hypoxic incubation when given less than half of their usual amount of glucose, as well as the host of other nutrients provided by FBS supplementation. This begs the question: What are the limits of metabolic stress that can be tolerated by these cells under these conditions? What are the limits of glucose deprivation that can be tolerated in this setting? How much do other micronutrients play a role under these conditions? We observed RHY specific upregulation of cholesterol biosynthesis and zinc import, under media conditions where these two nutrients are completely absent, as both zinc and cholesterol are not found in RPMI media[729]. As trace amounts of zinc may be released from cells during incubation via ZNT activity or apoptosis, a true test of zinc deprivation on these cells would be to add a extracellular zinc chelator such as DTPA to the media, to prevent zinc re-uptake[730]. Exploring the adaptation boundaries of these cells would deepen our understanding of extremities of the environments tumour cells can adapt for. Once armed with this knowledge, we will be better equipped to design novel therapeutics that push tumour cells beyond their capacity to adapt.

6.2.3 Investigation into the Rappont Luebering shunt and other specific metabolic adaptions

While our work did describe the transcriptomic changes occurring during HY an RHY incubation conditions, we still lack an understanding of the metabolic flux through these pathways. In order to solidify the description of these metabolic adaptations, metabalomics should be performed using heavy isotopes for differing nutrients. By replacing specific nutrients in the cell media with heavy carbon isotope versions, we could investigate the carbon flux of these cells under MHY conditions. Glucose should be prioritized, due to both its known role in hypoxia adaption and our RHY specific glucose metabolism findings. Investigating the metabolic flux through the Rappont Luebering shunt in glycolysis, would be particularly interesting. We could probe for the relative flux through this shunt used heavy glucose replaced media under RHY conditions, and probing for 2,3 bisphosphglycerate, the metabolic product of the first step in the Rappont Luebering shunt.

Alternatively, we could knock-down BPGM in B16 cells to see if that has a differential effect on RHY survivability. Its unclear why this shunt is upregulated, as it results in a drop of ATP generation by the glycolytic pathway which seems counter productive for cancer survival. It is unlikely that the shunt is being used for the generation of 2,3 bisphosphglycerate, as there are no known metabolic functions for it. The best known function of 2,3 bisphosphglycerate is in stabilizing hemoglobin in the deoxygenated state[731]. Perhaps the cancer cells are generating 2,3 bisphosphglycerate to secrete into the extracellular space with the intention of it being absorbed by passing red blood cells. If so, it would prevent the hemoglobin within the passing red blood cells from reabsorbing O_2, and potentially lead to a larger local delivery of oxygen. Therefore the upregulation of the Rappont Luebering shunt may be an attempt by the cancer cells to alleviate hypoxic conditions This could be tested by determining if the 2,3 bisphosphglycerate remains intracellular or is being secreted during RHY incubation. That alone would be a step in understanding the role of 2,3 bisphosphglycerate and the Rappont Luebering shunt in this context.

6.3.4 Characterizing RHY induced cell motility

Another axis of tumour cell adaptation to RHY conditions was the higher enrichment of motility associated phenotypes, such as cell adhesion and cell migration. Follow-up experiments to measure the motility rate of these cells compared to HY grown cells would help to characterize and differentiate these two cell adaptation states. Cell migration assays such as wound-healing assays can be done to determine if the these enriched phenotypes correspond with the expected cell activity, or if RHY causes a dysregulation in motility[732]. Given that RHY exposed cells are likely increasing their motility to find a better food source, chemotaxis assay should be performed as well, using both hypoxic and normoxic nutrient-complete media as chemotaxis sources[733]. While it is unlikely that the oxygen levels of the media will affect chemotaxis signalling, it would be interesting to see if it has a synergistic effect on chemotaxis, or if tumour cells are solely attracted to nutrients and growth factors.

6.2.5 Investigating the role of miRs in MHY adaptations

As demonstrated by the smallRNAseq dataset of MHY, there are several DEmiRs common between and unique to either of the hypoxic conditions. As microRNA activity is primarily to inhibit target mRNA protein output, further characterizing the roles of these miRs would lead to a deeper understanding of which mRNAs need to be removed from the transcriptome as quickly as possible to adapt to hypoxia. Additionally, target identification of miRs which are down-regulated under hypoxic conditions should be characterized as they represent ideal targets for countering hypoxic adaptation. The targets of miRs can be determined using Luciferase assays, where the miR binding site is introduced into the 3'UTR of the Luciferase mRNA and differential luciferase levels shows miR targeting. Alternatively, if the miR binds to a non CDS region of the target, CRISPR can be employed to mutate the binding site. Loss of regulation of the target would suggest that the miR was acting through this site.

Of the miRs with a loss of expression under hypoxic conditions, miR-7 and miR-211 are two candidates of primary interest. They represent RHY specific markers with evidence of being negative prognosis indicators in patients[451,459,493,604–608,627]. The expected targets of these miRs in the RHY transcriptome can be predicted using miRBooking, a software which allows for the

simulation of miR overexpression and predicts targets using seed sequence binding and relative expression of target genes[411]. Overexpression of these microRNAs can be accomplished through transfection of microRNA mimetics, or the development of microRNA expression constructs.

To test how the repression of hypoxia-induced miRs would effect hypoxic adaptation, we can overexpress reporter constructs expressing miR sponging sequences. These sponging sequences will reduce the effect of the targeted miR(s), allowing us to measure the effect of their absence on predicted targets. Much like how we can predict targets of over-expressed miRs in a given transcriptome, we can predict the targets of hypoxia-induced miRs by simulating their overexpression under tissue culture conditions in miRBooking[411]. These predicted targets can then be validated by measuring their relative overexpression under miR-sponging conditions.

6.2.6 Targeting MHY adaptations

Using our existing analyses, we can find several potential vulnerabilities which if inhibited, would counter the cellular metabolic response to hypoxia. For example, targeting either GPI1, GAPDH, PKM, or SLC16A3 would significantly impair the glycolysis pathway (**Figure 17**). By targeting GPI1, GAPDH, or PKM, we are targeting single-gene bottlenecks in the pathway which make them potential points of failure should they be significantly inhibited. GAPDH would make for the best target, as its place in the pathway is after the energy investment phase of glycolysis, increasing the negative effect of its suppression.

Similarly, inhibition of PKM would result in the prevention of pyruvate generation. This would effect both glycolysis and TCA cycle activity (**Figure 19**). Pyruvate is needed to regenerate NAD+ for the pathway. It is possible the TCA cycle would not be effected by PKM inhibition under hypoxic conditions, as we have evidence for pyruvate generation from malate based on the increased expression of mitochondrially localized isoforms of Malic Enzyme (**Figure 19**). By inhibiting the expression of SLC16A3, we block hypoxic glycolysis at the very end of the pathway: the exporting of lactate. The export of lactate is essential for maintaining a high glycolytic flux, and the inhibition of MCT4 (SLC16A3) leads to cell death[734].

Instead of targeting key elements of a metabolic pathway, targeting upregulated transporters may also be effective. Across the examined metabolic pathways, there are eight glucose importers expressed and 22 amino acid transporters expressed in our cells (**Figures 17&18**). Sufficient inhibitions of these transporters may be enough to prevent a pathway from running by restricting its input, as with the glucose transporters in glycolysis. Alternatively, inhibition of intracellular transporters like SLC25A1 may block the essential flow of intermediate metabolites leading to toxic outcomes.

Given the wide range and functions of genes that need to be targeted for this therapeutic approach, it is impossible for a single small molecule to have such a diverse set of intended activities. Similarly, a single RNAi molecule would not be able to target this diverse set of genes simultaneously. But by using a synmiR, it is possible to target all or some of these genes. Ultimately, the goal is to push hypoxic cancer cells beyond the limit of their capacity to adapt. By accomplishing this, we will have successfully treated the more therapy resistant cancer cells of the tumour.

6.2.7 Pitfalls and improvements to MHY RNAseq data analysis pipeline

While there certainly is a wealth of data and conclusions which have been drawn from the MHY RNAseq data, there are still some pitfalls and improvements which could be made to the experiment as a whole. Beginning with the overall design of the experiment, expanding from biological duplicates to triplicates would add additional certainty to our identification of DEGs and DEmiRs. The MHY experiment was performed in 2016, and at the time, we decided to prioritize sequencing depth over replicates to ensure detection of lowly expressed transcript. Looking back at the MHY analyses, we could have sacrificed sequencing depth to include triplicates without affecting the presented analyses. Alternatively we could have stayed with biological duplicates and include a fourth condition for analysis, such as B16-HG cells grown in RPMI without FBS under normoxia. FACS results suggested this culture condition stabilized HIF1α, and would have made for an interesting additional control to the TC condition.

A core aspect of the MHY experiment is the nature of the library preparation. By choosing rRNA depletion, we allow for detection of lncRNAs, pre-mRNAs, and other long non-

polyA tailed cellular RNAs. While lncRNAs were not addressed in this thesis, maximizing the diversity of detected RNA species in an RNAseq experiment aides in discovering novel aspects of transcriptomic adaption. Given the increased relevance of lncRNAs as functional components of the hypoxia response, examining them within our datasets is warrented[115,735]. Another aspect of the RNAseq library preparation is it requires bulk quantities of RNA. By the nature of being a bulk-RNAseq experiment, we automatically lose heterogeneity in gene expression between individual cells. This could be addressed by performing single-cell RNAseq on cells grown under MHY conditions. However, recall at the time of the experiment (2016), single-cell RNAseq was still in its early stage of adoption. Additionally, we wanted to capture the miRNA signatures of these cells, making single-cell RNAseq out of the question. Nevertheless, determining the heterogeneity of the metabolic adaptations remains an interesting question. This could be addressed while maintaining the spacial layout of the cells by probing for differentially expressed metabolic genes identified in Chapter 3 at the protein level using high-throughput Immunohistochemistry, or at the RNA level using spatially resolved transcriptomics[736,737].

Regarding our RNAseq read processing pipeline, the tools used reflect the time period when the data was acquired. In 2016, Tophat was regarded as adequate for aligning RNSeq reads to genomes or exomes[738,739]. Since 2016, several new read aligners have been developed with higher true-positive read alignments and better multi-mapping read algorithms[740,741]. In our data processing pipeline, multi-mapped reads were discarded from analysis, as the Tophat alignment tool was not well equipped to handle them at the time. This represents a high loss of potential data, as multi-mapping reads can account anywhere from 5-40% of reads, and skew gene quantification[740]. Fortunately this should be mitigated in our MHY dataset thanks to our high sequencing depth per sample, and our focus on inter-sample comparisons rather than intra-sample.

To verify that our outdated RNAseq read processing pipeline does not significantly affect our differential expression analyses, several steps should be taken. Re-alignment of our reads to the murine genome needs to be done with a more accepted or modern aligner, such as STAR or HISAT2[742,743]. Once re-aligned and re-processed by DESeq2, we can compare the DEG sets between both pipelines. Should there be large differences in the DEG sets between the two

pipelines, many of the bioinformatic analyses presented in Chapter 3 would need to be re-done. Expression patterns should be compared on a gene-by-gene basis, to see if any expression profile patterns (ie RHY>HY>TC) change. If these expression patterns are stable, reanalysis may not be necessary. Given that our analyses are primarily focused on differential expression of the same gene between samples, this kind of analysis should be relatively insensitive to multi-mapping reads.

Given that the B16-HG cell line was derived from a single clone, there is a higher probability that the B16-HG genome contains SNPs or other genomic mutations relative to polyclonal cell lines. This can be investigated by using tools such as *km*, and can be used to identify exonic mutation. Alternatively, whole genome sequencing of the cell line could be performed. This would be the most insightful, as it would provide locations of SNPs and larger genomic mutations, potentially explain the mechanism behind non-expressed genes like HIF2α, and provide the genomic locations for the integration of both lentiviral constructs.

6.3 Discussion of Long-Term Hypoxic incubation results

6.3.1 Inquiry into the nature of the LTHY protocol

To date, our LTHY protocol does not exist in the literature. Several groups have done extended hypoxic incubation or cyclic hypoxic incubation, but we are the first to report results arising from a extended hypoxic incubation protocol with progressive hypoxic stress[255]. While aspects of the LTHY protocol were explored, such as the effects of direct incubation at deeper hypoxic levels and extended incubation at milder hypoxic levels (**Figure 30**) questions still remain surrounding the steps in the protocol.

Further investigation into the effects of each step of the LTHY protocol are needed. For example, the upregulation of WT1 required preincubation at milder hypoxic levels for optimal induction. Why was that the case? Are the milder hypoxic phases of incubation priming the cells for these transcriptomic changes? One potential hypothesis is that the milder hypoxic phases allow the cells to sufficiently alter their epigenetic state to allow for the drastic transcriptomic changes seen at 0.5% O_2 and below. This is supported by the bimodal regulation of histone

genes at 1% and 0.1% O_2 (**Figure 27**, cluster 5). Differential expression of these genes suggests a change in the epigenome of the cells, and could explain why preincubation at milder hypoxic levels is needed. This could be tested for by growing B16 cells under the following conditions, and then processing them all to ATACseq: B16s grown under normoxic conditions; B16s subjected to the LTHY protocol, 0.5% O_2 timepoint; and B16s incubated directly at 5%, 1%, and 0.5% O_2 for 72 hours. By comparing the epigenetic landscape of the cells through ATACseq, we would be able to determine if the LTHY protocol induces unique epigenetic changes associated with increased transcription of our EMT markers, as well as determine if these epigenetic changes are unique to the LTHY protocol or if they can be induced by direct incubation at specific hypoxic levels.

Along this same line of questioning, it is also unclear what are the limits for each phase of the LTHY protocol are. We need to determine the minimal times and oxygen levels needed to induce the same transcriptional response observed in our original protocol. The transcriptional response can be estimated using a small panel of genes identified in our LTHY DEG analysis, such as WT1, CYR61, FN1, COL8A1 and others. We would start by minimizing the early steps of the protocol until we no longer have induction of these reporter genes at 0.5% O_2. Once that limit is discovered, we can optimize the later steps of the protocol by minimizing or eliminating incubation phases.

Once fully optimized, adding metabolic stress during the incubation phases can be introduced as well. This will allow for a more complete simulation of the tumour microenvironment, as metabolic stress often correlates with the hypoxic microenvironment. Optimization of the LTHY protocol will give us a more detailed understanding of the limits of hypoxic adaptation in cancer, and will provide context for hypoxic adaptation *in vivo*. For example, when sections of tumours become necrotic due to excessively hypoxic environments, we know those environments exceeded this optimized LTHY protocol.

6.3.2 Investigation of other potential EMT drivers in LTHY

As mentioned in the results and discussion surrounding **Figure 28**, there are other candidate EMT driving transcription factors which could explain our phenotype in the LTHY

dataset. Few canonical EMT drivers are expressed in our dataset. The B16-HG cell line stably expressed EZH2 and ZEB2, and these genes could be driving our EMT signature. While they were de-prioritized due to their lack of differential expression, these genes could be regulated at the protein level leading to an induction of EMT without having a differential expression at the mRNA level. As EZH2 is an inhibitor of EMT, therefore if EZH2 is driving our observed EMT, we would expect to see reduced levels of EZH2 protein under oxygen levels at 0.5% O_2 or 0.1% O_2[744]. Western blots to probe for the overall levels of EZH2 as well as ubiquitination status would allow us to investigate this. Alternatively, GSEA analyses could be done using a set of known EZH2 regulated genes. If EZH2 activity is being inhibited, we would expect to see an enrichment for these genes. We could perform similar experiments on ZEB2, keeping in mind that ZEB2 is an EMT promoter[293]. Therefore we would be looking for Increased levels of ZEB2 protein levels and positive GSEA enrichment for known ZEB2 regulated genes.

Alternatively, we could take and RNAi approach, and attempt to knock-down EZH2 or ZEB2 using shRNAs. With stable knock-down of EZH2, we would expect spontaneous EMT in our B16-HG cells, regardless of oxygen status. We can control for shRNA off-target effects by exogenously overexpressing an shRNA-resistant version of EZH2 in an attempt to recover the original phenotype. If the over expression rescues the phenotype, we would have two lines of evidence suggesting loss of EZH2 is driving our phenotype. Similar experiments can be performed on ZEB2 as well. If ZEB2 knock-down reduces our hypoxia induced EMT signature, that would suggest ZEB2 is acting as the main driver. Overepxression of ZEB2 may not yield phenotypic rescue results because as a positive EMT driver, ZEB2 is already not differentially regulated across LTHY conditions. This suggest that if ZEB2 is driven EMT, its not regulated at the mRNA abundance level.

6.3.3 Determining tWT1 transcription initiation factor(s)

While we did demonstrate evidence for HIF1α binding within the WT1 intron 5 sequence by ChIP-qCPR, the enriched HREs were far too distal to be directly driving transcription. To identify if HIF1 is acting as an enhancer to tWT1 transcription through these more distal interactions, performing a chromatin interaction analysis by paired-end tag sequencing (ChIA-

PET) on HIF1 in B16 cells undergoing to LTHY protocol could be done[745,746]. Should HIF1 bind to both the distal HRE elements identified by our ChIP-qPCR data (**Figure 32**) as well as the tWT1 TSS through direct or indirect interaction, ChIA-PET should be able to detect this.

Regardless, the precise mechanism of action for tWT1 transcription remains unanswered. To address this, there are two experiments which must be performed. The first is to verify the genomic sequence of *WT1* intron 5. It is possible that due to a deletion of genomic sequence in our B16, the enriched HREs could be located more proximally to the tWT1 TSS, providing a concise explanation for tWT1 transcription. While we did verify the genomic sequence immediately upstream of the TSS via genomic PCR and Sanger sequencing to be the correct genomic sequence, the extreme length of the intron prevented us from exhaustively verifying the *WT1* genomic locus. However, sequence verification can still be done by Sanger sequencing of multiple PCR products. Alternatively, next-generation sequencing can be done on B16 genomic DNA to determine the sequence as well.

Should the genomic sequence of *WT1* intron 5 not contain any deletions, more work is needed to profile the tWT1 TSS. One such experiment could be a series of ChIP-qPCRs on the transcription factors with binding sites that are the most proximal to the tWT1 TSS such as TCF12, ALX1, ATF3, and others (**Figure 32**). Through this series of ChIP-qPCR experiments, we would be able to determine the primary transcription factor driving tWT1 expression. Alternatively, we could develop promoter reporters for the tWT1 TSS to see if this region is capable of driving transcription of a reporter gene. Should this reporter work, we could then proceed to mutate different regions of the promoter to determine which sub-regions, and more precisely which transcription factors, are relevant to tWT1 transcription. To complement this experiment, we could attempt to determine the precise TSS of the tWT1 promoter, given that our RNAseq data only provides an estimate of the exact TSS. To address this, since we know the transcript sequence, we can use 5' Rapid Amplification of cDNA Ends (RACE) to determine the exact tWT1 TSS[747]. Alternatively, we could use a more global technique like CAGEseq to determine the tWT1 TSS, as well as TSSs across the genome[748].

6.3.4 Determining function of exon-6-skip KTS- isoform

Through our work characterizing the exon-6-skip KTS- isoform of tWT1, we have confirmed the protein's ability to bind to DNA using both ChIPseq and ChIP-qPCR. Due to the difficulties with generating a WT1 deficient cell line, we were unable to validate the activity of tWT1 as an activator or repressor for the various target genes identified in our ChIPseq data. Using a combination of shRNAs to target tWT1, or generating a tWT1 KO cell line using CRISPR would be ideal. One alternative method to verify exon-6-skip KTS- tWT1 target genes would be to overexpress exon-6-skip KTS- tWT1 in B16s and probe for differences in prospective target gene expression by qPCR.

Besides transcriptional control of target genes, tWT1 could have other functions which warrant investigation. One such function would be interacting with other proteins. To address this, we attempted to generate a HIS-tagged version of exon-6-skip KTS- tWT1 for protein complex pull-down assays. This experiment would allow us to determine what proteins and protein complexes tWT1 was associating with. Unfortunately, these experiments were unsuccessful. Initially there were issues with expression of HIS-tagged tWT1, which was eventually solved by moving the HIS tag from the C terminus to the N terminus. We believe that the C terminal HIS tag was interfering with proper folding of the zinc fingers, as the HIS tag has affinity for Zn^{2+} and has been shown to alter the conformation of C terminal zinc finger transcription factors[749,750]. This improper folding would then result in proteasomal degradation. After solving the construct stability issue, the next issue was producing sufficient amounts of protein for HIS-tagged pull-downs. Initial experiments failed to induce sufficient levels of HIS-tagged tWT1 for pull-down, and further tWT1 production optimization is required. Unfortunately direct IP of unlabelled tWT1 has become more difficult, as antibody suppliers no longer carry WT1 C-terminus binding antibodies. They used to generate such antibodies, but removed them from production due to "non-specific banding". It may be that these antibodies were in fact labelling tWT1 isoforms. C-terminal WT1 antibodies could be generated in-house to circumvent this issue.

An alternative method to determine tWT1 protein binding partners would be to use BioID to directly enrich for tWT1 proximal proteins. In this method, a biotin ligase tWT1 fusion

would be generated[751,752]. Once the fusion protein is expressed under hypoxic conditions, biotin is added to the cell media, allowing the biotin ligase to biotinylate proteins within the vicinity of tWT1. Once biotinylated, these proteins can be purified using Streptavidin pull-down methods, and identified via mass spectrometry. Based on our ChIPseq data, we do not expect other transcription factors to be binding to tWT1, as there were no other significantly enriched transcription factor binding motifs identified by HOMER analysis. However these analyses were performed on the entire list of ChIPseq peaks. If we focus our analyses on peaks without the WT1 binding motif, or peaks associated with genes involved in cell adhesion, we may identify potential WT1 protein complex members.

Based on previous works identifying WT1 protein-protein interactors, we can predict which known proteins should be enriched[326]. As we expressed mainly the zinc fingers of WT1, we can predict any WT1 binding partners which interact with the WT1 zinc fingers as potential interactors. This would include p53, p73, WTX, CBP, SRY, potentially Par-4, and others. We would not expect interaction partners which bind to the N-terminus of WT to be interaction partners with tWT1. This includes STAT3, HtrA2, BASP1, Hsp70, and others[326]. The tWT1 N terminal protein sequence is unlikely to be involved in protein-protein interaction, as its sequence is not homologous to anything in the human or murine proteome (as validated by BLAST)[753].

In addition to protein-protein interactions, it is also possible the exon-6-skip KTS- tWT1 has a role as a RNA binding protein. Recent work by Nishikawa et al has demonstrated that both the KTS- and KTS+ zinc fingers of WT1 are capable of binding to RNA[341]. They show that the KTS- isoform of WT zinc fingers are capable of binding to a 29mer RNA sequence which forms a hairpin. This raises the possibility that the function of tWT1 during hypoxic adaptation is to function as a RNA binding protein rather than a DNA-binding or protein-binding factor. By binding to specific RNA sequences, tWT1 may be blocking miRNA binding sites and prevent degradation of specific transcripts. Alternatively, tWT1 may affect splicing patterns for targeted primary mRNA transcripts. By performing an RNA IP on tWT1, we would be able determine what are the RNA binding targets of tWT1 in the hypoxic transcriptome, which will help prioritize

these hypotheses. It may be that the role of tWT1 in hypoxic adaptation is much more subtle than anticipated, resulting in reduced mRNA turn-over or shifting mRNA isoform ratios.

In this same line of thinking, we have yet to investigate the effects of the exon-6-skip KTS+ isoform in hypoxic adaptation. Based on our RNAseq data, the KTS+ and KTS- isoforms are expressed relatively equally. Given the known functional distinctions between the KTS+ and KTS- isoforms of WT1, it is likely that exon-6-skip KTS+ tWT1 also has a distinct functional role in hypoxic adaptation. Therefore, an important extension of the work presented in this thesis is to investigate the role of exon-6-skip KTS+ tWT1 in hypoxic adaptation. All experiments performed and proposed to delineate the function of exon-6-skip KTS- tWT1 should be repeated on the KTS+ isoform. Given the preference of KTS+ WT1 for binding to RNA, RNA binding experiments should be prioritized over DNA binding.

An additional question surrounding these tWT1 transcripts is their function in a tumor setting. Given the higher complexity of the tumor microenvironment compared to tissue culture conditions, analyses which maintain a single-cell resolution and spatial arrangement of the tumor, such as spatial transcriptomics, would be ideal. Using techniques such as seqFISH+ or msFISH, it would be possible to identify the exon-6-skip tWT1 transcripts using isoform-specific probes[754,755]. With these techniques, it would be possible to identify which tumor cells *in situ* are expressing these tWT1 transcripts, as well as other hypoxia/EMT markers. If we find these hypoxia/EMT markers localized in single cells with a mesenchymal morphology oriented away from hypoxic subregions of the tumor, it would suggest these cells are the metastases-forming subpopulation of the tumor[174]. Should tWT1 also colocalize within these cells, it would support our hypothesis of tWT1 driving EMT in in vivo. Should tWT1 transcripts not localize within these mesenchymal cells, the transcriptomic information provided by these spatial transcriptomic techniques would allow us to make better hypotheses for the role of tWT1 in tumour biology.

6.3.5 Validation tWT1 expression in patient samples

An interesting finding presented in this thesis was the discovery and description of tWT1 isoforms in human leukemia samples. Using a kmer-based analysis of patient samples, we were able to identify samples which contained expression of a known human truncated WT1 isoform

known as isoform G, as well as a less described isoform which we termed isoform I. While these isoforms were found in the Leucegene dataset, their rarity made our survival analyses inconclusive. To address this, additional tWT1 expressing cancer samples are needed. Unfortunately, as these isoforms are not identified in standard gene expression analysis pipelines, we require access to patient data at the level of RNAseq reads, or a reanalysis which includes transcripts considered to be "non-productive", to properly identify expression of these isoforms. Ideally, we would be given access to The Cancer Genome Atlas (TCGA) database, which contains over 20,000 primary samples across 33 cancer types. With this larger dataset, not only will we be able to identify cancer types with a higher propensity for tWT1 expression, but we will be able to perform more conclusive survival analyses. We can also correlate isoform G/I expression with a panel of hypoxia induced mRNAs such as CA9, VEGFA, and high expression of GAPDH as a proxy for hypoxic adaptation in the sample. This will allow us to correlate isoform G/I expression with our observed mechanism of induction.

As we did with the identification of the exon-6-skip tWT1 isoforms in mice, a follow-up experiment would be to identify the isoform G/I transcripts in human cancer samples. This can be done through PCR of the unique 5' splicing pattern which distinguishes isoforms G/I from each other and other WT1 isoforms. These PCRs would be done using cDNA derived from the human melanoma and breast cancer cell lines used in this thesis, which demonstrated WT1 upregulation under hypoxic conditions. This will also validate if isoform G/I are expressed in these cell lines which was suggested by the 3' probes used in our qPCR experiments (**Figure 29**) but requires more concrete evidence. Should these cell lines not express isoforms G and/or I, primary AML patient samples will be needed, as they have a chance of expressing these isoforms.

Once the expression of isoforms G and/or I has been confirmed in human cancers, they would be subjected to the same set of characterization experiments as performed on the murine orthologs. This would begin with relative level of protein production and subcellular localization. Based on our findings when comparing the exon-6-skip and exon-6 tWT1 mRNAs in mice, it would be likely that isoform I would the more productive and functional of the two isoforms (**Figure 41B**). This is because isoform I begins translation from a stronger start codon, a

codon with a similar kozak context to the exon-6-skip isoform. For isoform G to form a functional polypeptide, it requires translation initation from a non-canonical start codon[350].

Assuming that isoform I is expressed in cancer samples, it would demonstrate a convergence of mechanisms to express a truncated isoform of WT1 in multiple cancer types in multiple species. Both human isoform I and murine exon-6-skip are capable of generating similarly functional proteins, using different mechanisms for translation initiation. In mice, a novel splicing route is used to bring a start codon into frame. Due to these diverse mechanisms whose only similarly is the goal of expressing tWT1, it is unlikely that tWT1 expression is a spandrel in the evolution of the tumour. The tumour wants to express tWT1, the remaining question is why. We've chipped away at that question, but much work still remains. I trust that we've demonstrated the need for further investigation into this novel protein.

6.5.6 Issues and improvements to the LTHY RNAseq data analysis pipeline

Similar to the MHY RNAseq dataset, the LTHY RNAseq data was acquired early in the project. The LTHY data was acquired in early 2017, and was processed using the same pipeline as the MHY dataset. Therefore, all the pitfalls and improvements that apply to the MHY dataset also apply to the LTHY dataset. What separates the LTHY data analysis from the MHY dataset is the investigation into gene isoforms, particularity tWT1. Because of this, there would have been a benefit of acquiring the mRNAseq using third-generation sequencing techniques such as nanopore-based sequencing. The benefit of having the longer reads provided by this technology would allow for more accurate measurements of the exon-6 to exon-6-skip isoform ratio, as well as their relation to the relative abundance of the KTS motif[756]. Additionally, third-generation sequencing technologies offer more precise detection of the TSSs, which would allow for a better definition of the 5' end of the tWT1 transcripts[757].

6.4 Discussion on the development of synthetic microRNAs

Since the discovery of microRNAs, the focus of the literature has been largely descriptive. Research focused on elucidating the mechanism by which miRs exhibited their function. With the recent work done to describe the functionality of the tail sequence of miRs,

we are closing in on a model for the role of the tail sequence in miR function. However, to complete the model of microRNA functionality, the research being done needs to shift from being descriptive to prescriptive. Our work represents a large step out into this direction, where we are attempting to determine the mechanism of action of microRNAs through the design of synthetic miRs.

When the field of synmiR design matures, it will represent the dawn of an entirely new type of therapeutic. Synthetic microRNAs have unique therapeutic potential and can be applied in multiple disease contexts. They can be designed to target multiple transcripts simultaneously, with varying efficacy across the set of targets. The differential efficacy can be leveraged to cleave priority targets, repress translation of secondary targets, and even block degradation of tertiary targets. Tertiary targets can be protected from endogenous miRs by intentionally designing the synmiR to block effective endogenous miR sites, and have the miR tail sequence be ineffective at cleaving or repressing translation on that site. This makes synthetic microRNAs a very sophisticated therapeutic, as no other single therapeutic can have such a diversity of intended effects.

Once design parameters are settled, the question of design philosophy must be answered. When designing synmiRs is it more therapeutically effective to target multiple transcripts of a given pathway, or is focusing targeting on one or two bottlenecks more effective? Depending on the answer to this question, synmiRs may need to be specialized to target specific pathways or will be more flexible and be able to inhibit different pathways simultaneously. Based on the number of targets endogenous miRs are able to effect, synmiRs may be able to inhibit multiple pathways simultaneously.

Applied to a therapeutic context, this provides several possibilities. In a cancer context, synmiRs can be designed to target key hypoxic adaptations to kill hypoxic tumour cells. Targeting key genes in glycolysis may be sufficient to reach this goal due to the energy requirement of hypoxic tumour cells. Alternatively, targeting both metabolic and motility genes induced by hypoxia would prevent the escape of hypoxic cells to a more ameliorable microenvironment and reduce metastases. These are just examples of synmiRs applied to a

cancer therapeutic context. The ability to tune gene expression can turn the tide in a wide variety of disease contexts.

6.5 Future directions

Given the wide diversity of work done for this project, several main avenues can be pursued. Of the directions discussed, further characterization of tWT1 should be prioritized. Knowledge of this novel isoform and its function in cancer has the highest probability of impacting patient survival. The novel peptide sequence from tWT1 isoform I could be used as a potential cancer vaccine target, assuming it is capable of generating a Major Histocompatibility Complex (MHC) peptide for presentation to the immune system and is absent in immune education. The expression of tWT1 may be useful as a potential biomarker. Knowledge of this isoform will help inform future research, and may help explain previously unexpected or contradictory results. For these reasons, tWT1 characterization, particularly in human cancers, should be prioritized. Secondarily, additional work should be done on progressing our synmiR designs. The therapeutic potential for synthetic microRNAs as well as contribution to the definition the mechanism of action of miRs makes any progress of this work highly significant.

6.6 Conclusions

In this thesis, we presented two new *in vitro* hypoxic incubation protocols. The design purpose of each incubation protocol was to consider key aspects of the *in vivo* hypoxic microenvironment and simulate them *in vitro*. Through our RNAseq analyses, we demonstrated how these protocols enhanced hypoxia adaptation to better match phenotypes seen in patients. When we implemented more aspects of *in vivo* hypoxia, we got drastically different DEmiR and DEG signatures that are more representative of patient data. These incubation protocols allowed for better investigation into potential adaptation pathways tumour cells can make in the hypoxic microenvironment.

Arguably, a core issue with current cancer treatments is they do not account for the variety of cancer cell phenotypes present in the tumour. This inability causes therapeutics to miss specific sub-types of cancer cells within the tumour, such as hypoxic tumour cells. This is

why understanding the adaptive capacity of tumour cells to hypoxia crucial to the design of better cancer therapeutics. And to better understand hypoxic adaptation capacity, we must push cancer cells to the limits of their adaptation. Our novel incubation protocols move the field in this direction. Currently in the literature, there is little effort in exploring hypoxic adaptations using non-standard hypoxic incubation protocols. We are among the first to incorporate additional stresses into our hypoxic incubation protocols.

In order to effectively target this large variety of potential adaptations, single target therapeutics are not enough. The recurring theme in cancer therapy is a high initial rate of success of the therapeutic followed by tumour escape. In order to account for this adaptive potential, a multi-targeted approach is essential. This thesis presents a novel findings regarding designing single-sequence, multi-targeting RNAi molecules in the hopes to address this issue. We have broken new ground in describing non-standard RNAi duplexes in a prescriptive manner which is in line with recently published descriptive findings. We pushed this concept to the point of designing synthetic microRNAs, something which has never been attempted to date. Synthetic microRNAs offer a new model for therapeutics, where multiple effector genes are targeted rather than a single driver. This strategy may be favourable as it by-passes alternative activation mechanisms for the same phenotype, and will be effective across a larger section of the cancer adaptive landscape. With the increasing interest in RNA therapeutics and specifically microRNA based therapies, our work only adds to this burgeoning field.

References

1. Moore, G. E., Gerner, R. E. & Franklin, H. A. Culture of normal human leukocytes. *JAMA* **199**, 519–524 (1967).
2. Babcock, G. T. How oxygen is activated and reduced in respiration. *Proceedings of the National Academy of Sciences* **96**, 12971–12973 (1999).
3. Rolfe, D. F. & Brown, G. C. Cellular energy utilization and molecular origin of standard metabolic rate in mammals. *Physiological Reviews* **77**, 731–758 (1997).
4. Webby, C. J. *et al*. Jmjd6 catalyses lysyl-hydroxylation of U2AF65, a protein associated with RNA splicing. *Science* **325**, 90–93 (2009).
5. Lando, D. *et al*. FIH-1 is an asparaginyl hydroxylase enzyme that regulates the transcriptional activity of hypoxia-inducible factor. *Genes Dev* **16**, 1466–1471 (2002).
6. Yang, M. *et al*. Asparagine and aspartate hydroxylation of the cytoskeletal ankyrin family is catalyzed by factor-inhibiting hypoxia-inducible factor. *J Biol Chem* **286**, 7648–7660 (2011).
7. Ge, W. *et al*. Oxygenase-catalyzed ribosome hydroxylation occurs in prokaryotes and humans. *Nat Chem Biol* **8**, 960–962 (2012).
8. Zurlo, G., Guo, J., Takada, M., Wei, W. & Zhang, Q. New Insights into Protein Hydroxylation and Its Important Role in Human Diseases. *Biochim Biophys Acta* **1866**, 208–220 (2016).
9. Aktan, F. iNOS-mediated nitric oxide production and its regulation. *Life Sciences* **75**, 639–653 (2004).
10. Bertout, J. A., Patel, S. A. & Simon, M. C. The impact of O2 availability on human cancer. *Nat Rev Cancer* **8**, 967–975 (2008).
11. Simon, M. C. & Keith, B. The role of oxygen availability in embryonic development and stem cell function. *Nat Rev Mol Cell Biol* **9**, 285–296 (2008).
12. Carreau, A., El Hafny-Rahbi, B., Matejuk, A., Grillon, C. & Kieda, C. Why is the partial oxygen pressure of human tissues a crucial parameter? Small molecules and hypoxia. *J Cell Mol Med* **15**, 1239–1253 (2011).

13. Lartigau, E. *et al*. Intratumoral oxygen tension in metastatic melanoma. *Melanoma Res* **7**, 400–406 (1997).

14. Wang, W., Winlove, C. P. & Michel, C. C. Oxygen partial pressure in outer layers of skin of human finger nail folds. *J Physiol* **549**, 855–863 (2003).

15. McKeown, S. R. Defining normoxia, physoxia and hypoxia in tumours—implications for treatment response. *Br J Radiol* **87**, 20130676 (2014).

16. Brezis, M., Agmon, Y. & Epstein, F. H. Determinants of intrarenal oxygenation. I. Effects of diuretics. *Am J Physiol* **267**, F1059-1062 (1994).

17. Brezis, M., Heyman, S. N. & Epstein, F. H. Determinants of intrarenal oxygenation. II. Hemodynamic effects. *Am J Physiol* **267**, F1063-1068 (1994).

18. Braun, R. D., Lanzen, J. L., Snyder, S. A. & Dewhirst, M. W. Comparison of tumor and normal tissue oxygen tension measurements using OxyLite or microelectrodes in rodents. *Am J Physiol Heart Circ Physiol* **280**, H2533-2544 (2001).

19. Hale, L. P., Braun, R. D., Gwinn, W. M., Greer, P. K. & Dewhirst, M. W. Hypoxia in the thymus: role of oxygen tension in thymocyte survival. *American Journal of Physiology-Heart and Circulatory Physiology* **282**, H1467–H1477 (2002).

20. Spencer, J. A. *et al*. Direct measurement of local oxygen concentration in the bone marrow of live animals. *Nature* **508**, 269–273 (2014).

21. Warburg, O. On the origin of cancer cells. *Science* **123**, 309–314 (1956).

22. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).

23. Lacassagne A. *Compt. Rend. Acad. Sci.* **ccxv**, 231 (1942).

24. Gray, L. H., Conger, A. D., Ebert, M., Hornsey, S. & Scott, O. C. The concentration of oxygen dissolved in tissues at the time of irradiation as a factor in radiotherapy. *Br J Radiol* **26**, 638–648 (1953).

25. Thomlinson, R. H. & Gray, L. H. The histological structure of some human lung cancers and the possible implications for radiotherapy. *Br J Cancer* **9**, 539–549 (1955).

26. Powers, W. E. & Tolmach, L. J. DEMONSTRATION OF AN ANOXIC COMPONENT IN A MOUSE TUMOR-CELL POPULATION BY IN VIVO ASSAY OF SURVIVAL FOLLOWING IRRADIATION. *Radiology* **83**, 328–336 (1964).

27. Churchill-Davidson, I. OXYGENATION IN RADIOTHERAPY OF MALIGNANT DISEASE OF THE UPPER AIR PASSAGES. THE OXYGEN EFFECT OF RADIOTHERAPY. *Proc R Soc Med* **57**, 635–638 (1964).

28. Cater, D. B. & Silver, I. A. Quantitative measurements of oxygen tension in normal tissues and in the tumours of patients before and after radiotherapy. *Acta radiol* **53**, 233–256 (1960).

29. Gatenby, R. A. *et al*. Oxygen tension in human tumors: in vivo mapping using CT-guided probes. *Radiology* **156**, 211–214 (1985).

30. Kolstad, P. Intercapillary distance, oxygen tension and local recurrence in cervix cancer. *Scand J Clin Lab Invest Suppl* **106**, 145–157 (1968).

31. Bergsjö, P. & Evans, J. C. Oxygen tension of cervical carcinoma during the early phase of external irradiation. II. Measurements with bare platinum micro electrodes. *Scand J Clin Lab Invest* **27**, 71–82 (1971).

32. Badib, A. O. & Webster, J. H. Changes in tumor oxygen tension during radiation therapy. *Acta Radiol Ther Phys Biol* **8**, 247–257 (1969).

33. Urbach, F. & Noell, W. K. Effects of oxygen breathing on tumor oxygen measured polarographically. *J Appl Physiol* **13**, 61–65 (1958).

34. Urbach, F. Pathophysiology of malignancy. I. Tissue oxygen tension of benign and malignant tumors of the skin. *Proc Soc Exp Biol Med* **92**, 644–649 (1956).

35. Thomlinson, R. H. Tumour anoxia and the response to radiation. *Sci Basis Med Annu Rev* 74–90 (1965).

36. Brown, J. M. Tumor microenvironment and the response to anticancer therapy. *Cancer Biol Ther* **1**, 453–458 (2002).

37. Blumenson, L. E. & Bross, I. D. A possible mechanism for enhancement of increased production of tumor angiogenic factor. *Growth* **40**, 205–209 (1976).

38. van den Brenk, H. A., Moore, V., Sharpington, C. & Orton, C. Production of metastases by a primary tumour irradiated under aerobic and anaerobic conditions in vivo. *Br J Cancer* **26**, 402–412 (1972).

39. Young, S. D., Marshall, R. S. & Hill, R. P. Hypoxia induces DNA overreplication and enhances metastatic potential of murine tumor cells. *Proc Natl Acad Sci U S A* **85**, 9533–9537 (1988).

40. Heacock, C. S. & Sutherland, R. M. Enhanced synthesis of stress proteins caused by hypoxia and relation to altered cell growth and metabolism. *Br J Cancer* **62**, 217–225 (1990).

41. Sciandra, J. J., Subjeck, J. R. & Hughes, C. S. Induction of glucose-regulated proteins during anaerobic exposure and of heat-shock proteins after reoxygenation. *Proc Natl Acad Sci U S A* **81**, 4843–4847 (1984).

42. Semenza, G. L. & Wang, G. L. A nuclear factor induced by hypoxia via de novo protein synthesis binds to the human erythropoietin gene enhancer at a site required for transcriptional activation. *Mol Cell Biol* **12**, 5447–5454 (1992).

43. Goldberg, M. A., Dunning, S. P. & Bunn, H. F. Regulation of the erythropoietin gene: evidence that the oxygen sensor is a heme protein. *Science* **242**, 1412–1415 (1988).

44. Beck, I., Ramirez, S., Weinmann, R. & Caro, J. Enhancer element at the 3'-flanking region controls transcriptional response to hypoxia in the human erythropoietin gene. *J Biol Chem* **266**, 15563–15566 (1991).

45. Pugh, C. W., Tan, C. C., Jones, R. W. & Ratcliffe, P. J. Functional analysis of an oxygen-regulated transcriptional enhancer lying 3' to the mouse erythropoietin gene. *Proc Natl Acad Sci U S A* **88**, 10553–10557 (1991).

46. Semenza, G. L., Nejfelt, M. K., Chi, S. M. & Antonarakis, S. E. Hypoxia-inducible nuclear factors bind to an enhancer element located 3' to the human erythropoietin gene. *Proc Natl Acad Sci U S A* **88**, 5680–5684 (1991).

47. Madan, A. & Curtin, P. T. A 24-base-pair sequence 3' to the human erythropoietin gene contains a hypoxia-responsive transcriptional enhancer. *Proc Natl Acad Sci U S A* **90**, 3928–3932 (1993).

48. Wang, G. L., Jiang, B. H., Rue, E. A. & Semenza, G. L. Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O2 tension. *Proc Natl Acad Sci U S A* **92**, 5510–5514 (1995).

49. Reyes, H., Reisz-Porszasz, S. & Hankinson, O. Identification of the Ah receptor nuclear translocator protein (Arnt) as a component of the DNA binding form of the Ah receptor. *Science* **256**, 1193–1195 (1992).

50. Hoffman, E. C. *et al*. Cloning of a factor required for activity of the Ah (dioxin) receptor. *Science* **252**, 954–958 (1991).

51. Jiang, B. H., Semenza, G. L., Bauer, C. & Marti, H. H. Hypoxia-inducible factor 1 levels vary exponentially over a physiologically relevant range of O2 tension. *Am J Physiol* **271**, C1172-1180 (1996).

52. Ratcliffe, P. J. HIF-1 and HIF-2: working alone or together in hypoxia? *J Clin Invest* **117**, 862–865 (2007).

53. Graham, A. M. & Presnell, J. S. Hypoxia Inducible Factor (HIF) transcription factor family expansion, diversification, divergence and selection in eukaryotes. *PLOS ONE* **12**, e0179545 (2017).

54. Rytkönen, K. T., Williams, T. A., Renshaw, G. M., Primmer, C. R. & Nikinmaa, M. Molecular Evolution of the Metazoan PHD–HIF Oxygen-Sensing System. *Molecular Biology and Evolution* **28**, 1913–1926 (2011).

55. Inoue, J. G. *et al*. Evolutionary Origin and Phylogeny of the Modern Holocephalans (Chondrichthyes: Chimaeriformes): A Mitogenomic Perspective. *Molecular Biology and Evolution* **27**, 2576–2586 (2010).

56. Hedges, S. B. & Kumar, S. *The Timetree of Life*. (OUP Oxford, 2009).

57. Gu, Y. Z., Moran, S. M., Hogenesch, J. B., Wartman, L. & Bradfield, C. A. Molecular characterization and chromosomal localization of a third alpha-class hypoxia inducible factor subunit, HIF3alpha. *Gene Expr* **7**, 205–213 (1998).

58. Wiesener, M. S. *et al*. Widespread hypoxia-inducible expression of HIF-2alpha in distinct cell populations of different organs. *FASEB J* **17**, 271–273 (2003).

59. Mole, D. R. *et al*. Genome-wide association of hypoxia-inducible factor (HIF)-1alpha and HIF-2alpha DNA binding with expression profiling of hypoxia-inducible transcripts. *J Biol Chem* **284**, 16767–16775 (2009).

60. Rankin, E. B. *et al*. Hypoxia-inducible factor 2 regulates hepatic lipid metabolism. *Mol Cell Biol* **29**, 4527–4538 (2009).

61. Makino, Y. *et al*. Inhibitory PAS domain protein is a negative regulator of hypoxia-inducible gene expression. *Nature* **414**, 550–554 (2001).

62. Salceda, S. & Caro, J. Hypoxia-inducible factor 1alpha (HIF-1alpha) protein is rapidly degraded by the ubiquitin-proteasome system under normoxic conditions. Its stabilization by hypoxia depends on redox-induced changes. *J Biol Chem* **272**, 22642–22647 (1997).

63. Huang, L. E., Gu, J., Schau, M. & Bunn, H. F. Regulation of hypoxia-inducible factor 1alpha is mediated by an O2-dependent degradation domain via the ubiquitin-proteasome pathway. *Proc Natl Acad Sci U S A* **95**, 7987–7992 (1998).

64. Kallio, P. J., Wilson, W. J., O'Brien, S., Makino, Y. & Poellinger, L. Regulation of the hypoxia-inducible transcription factor 1alpha by the ubiquitin-proteasome pathway. *J Biol Chem* **274**, 6519–6525 (1999).

65. Maxwell, P. H. *et al*. The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature* **399**, 271–275 (1999).

66. Gossage, L., Eisen, T. & Maher, E. R. VHL, the story of a tumour suppressor gene. *Nat Rev Cancer* **15**, 55–64 (2015).

67. Jaakkola, P. *et al*. Targeting of HIF-alpha to the von Hippel-Lindau ubiquitylation complex by O2-regulated prolyl hydroxylation. *Science* **292**, 468–472 (2001).

68. Ivan, M. *et al*. HIFalpha targeted for VHL-mediated destruction by proline hydroxylation: implications for O2 sensing. *Science* **292**, 464–468 (2001).

69. Yu, F., White, S. B., Zhao, Q. & Lee, F. S. HIF-1alpha binding to VHL is regulated by stimulus-sensitive proline hydroxylation. *Proc Natl Acad Sci U S A* **98**, 9630–9635 (2001).

70. Bruick, R. K. & McKnight, S. L. A conserved family of prolyl-4-hydroxylases that modify HIF. *Science* **294**, 1337–1340 (2001).

71. Epstein, A. C. *et al*. C. elegans EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. *Cell* **107**, 43–54 (2001).

72. Ivan, M. *et al*. Biochemical purification and pharmacological inhibition of a mammalian prolyl hydroxylase acting on hypoxia-inducible factor. *Proc Natl Acad Sci U S A* **99**, 13459–13464 (2002).

73. Oehme, F. *et al*. Overexpression of PH-4, a novel putative proline 4-hydroxylase, modulates activity of hypoxia-inducible transcription factors. *Biochem Biophys Res Commun* **296**, 343–349 (2002).

74. Myllyharju, J. & Kivirikko, K. I. Characterization of the iron- and 2-oxoglutarate-binding sites of human prolyl 4-hydroxylase. *EMBO J* **16**, 1173–1180 (1997).

75. Myllyharju, J. Prolyl 4-hydroxylases, key enzymes in the synthesis of collagens and regulation of the response to hypoxia, and their roles as treatment targets. *Ann Med* **40**, 402–417 (2008).

76. Kaelin, W. G. & Ratcliffe, P. J. Oxygen sensing by metazoans: the central role of the HIF hydroxylase pathway. *Mol Cell* **30**, 393–402 (2008).

77. Loenarz, C. & Schofield, C. J. Physiological and biochemical aspects of hydroxylations and demethylations catalyzed by human 2-oxoglutarate oxygenases. *Trends Biochem Sci* **36**, 7–18 (2011).

78. Pan, Y. *et al*. Multiple factors affecting cellular redox status and energy metabolism modulate hypoxia-inducible factor prolyl hydroxylase activity in vivo and in vitro. *Mol Cell Biol* **27**, 912–925 (2007).

79. Mahon, P. C., Hirota, K. & Semenza, G. L. FIH-1: a novel protein that interacts with HIF-1alpha and VHL to mediate repression of HIF-1 transcriptional activity. *Genes Dev* **15**, 2675–2686 (2001).

80. Lando, D., Peet, D. J., Whelan, D. A., Gorman, J. J. & Whitelaw, M. L. Asparagine hydroxylation of the HIF transactivation domain a hypoxic switch. *Science* **295**, 858–861 (2002).

81. Huang, L. E., Arany, Z., Livingston, D. M. & Bunn, H. F. Activation of hypoxia-inducible transcription factor depends primarily upon redox-sensitive stabilization of its alpha subunit. *J Biol Chem* **271**, 32253–32259 (1996).

82. Semenza, G. L. Targeting HIF-1 for cancer therapy. *Nat Rev Cancer* **3**, 721–732 (2003).

83. Gu, J., Milligan, J. & Huang, L. E. Molecular Mechanism of Hypoxia-inducible Factor 1α-p300 Interaction: A LEUCINE-RICH INTERFACE REGULATED BY A SINGLE CYSTEINE *. *Journal of Biological Chemistry* **276**, 3550–3554 (2001).

84. Albanese, A., Daly, L. A., Mennerich, D., Kietzmann, T. & Sée, V. The Role of Hypoxia-Inducible Factor Post-Translational Modifications in Regulating Its Localisation, Stability, and Activity. *Int J Mol Sci* **22**, E268 (2020).

85. Flügel, D., Görlach, A., Michiels, C. & Kietzmann, T. Glycogen synthase kinase 3 phosphorylates hypoxia-inducible factor 1alpha and mediates its destabilization in a VHL-independent manner. *Mol Cell Biol* **27**, 3253–3265 (2007).

86. Cassavaugh, J. M. *et al*. Negative regulation of HIF-1α by an FBW7-mediated degradation pathway during hypoxia. *J Cell Biochem* **112**, 3882–3890 (2011).

87. Flügel, D., Görlach, A. & Kietzmann, T. GSK-3β regulates cell growth, migration, and angiogenesis via Fbw7 and USP28-dependent degradation of HIF-1α. *Blood* **119**, 1292–1301 (2012).

88. Xu, D., Yao, Y., Lu, L., Costa, M. & Dai, W. Plk3 functions as an essential component of the hypoxia regulatory pathway by direct phosphorylation of HIF-1alpha. *J Biol Chem* **285**, 38944–38950 (2010).

89. Lacher, S. E., Levings, D. C., Freeman, S. & Slattery, M. Identification of a functional antioxidant response element at the HIF1A locus. *Redox Biol* **19**, 401–411 (2018).

90. Lee, J.-Y. *et al*. LSD1 demethylates HIF1α to inhibit hydroxylation and ubiquitin-mediated degradation in tumor angiogenesis. *Oncogene* **36**, 5512–5521 (2017).

91. Geng, H. *et al*. HIF1α protein stability is increased by acetylation at lysine 709. *J Biol Chem* **287**, 35496–35505 (2012).

92. Bullen, J. W. *et al*. Protein kinase A-dependent phosphorylation stimulates the transcriptional activity of hypoxia-inducible factor 1. *Sci Signal* **9**, ra56 (2016).

93. Lim, J.-H. *et al*. Sirtuin 1 modulates cellular responses to hypoxia by deacetylating hypoxia-inducible factor 1alpha. *Mol Cell* **38**, 864–878 (2010).

94. Cam, H., Easton, J. B., High, A. & Houghton, P. J. mTORC1 signaling under hypoxic conditions is controlled by ATM-dependent phosphorylation of HIF-1α. *Mol Cell* **40**, 509–520 (2010).

95. Geng, H. *et al*. HDAC4 protein regulates HIF1α protein lysine acetylation and cancer cell response to hypoxia. *J Biol Chem* **286**, 38095–38102 (2011).

96. Gradin, K., Takasaki, C., Fujii-Kuriyama, Y. & Sogawa, K. The transcriptional activation function of the HIF-like factor requires phosphorylation at a conserved threonine. *J Biol Chem* **277**, 23508–23514 (2002).

97. Lancaster, D. E. *et al*. Disruption of dimerization and substrate phosphorylation inhibit factor inhibiting hypoxia-inducible factor (FIH) activity. *Biochem J* **383**, 429–437 (2004).

98. Cho, H., Ahn, D.-R., Park, H. & Yang, E. G. Modulation of p300 binding by posttranslational modifications of the C-terminal activation domain of hypoxia-inducible factor-1alpha. *FEBS Lett* **581**, 1542–1548 (2007).

99. Herzog, J. *et al*. Cyclin-dependent kinase 5 stabilizes hypoxia-inducible factor-1α: a novel approach for inhibiting angiogenesis in hepatocellular carcinoma. *Oncotarget* **7**, 27108–27121 (2016).

100. Hubbi, M. E. *et al*. Cyclin-dependent kinases regulate lysosomal degradation of hypoxia-inducible factor 1α to promote cell-cycle progression. *Proc Natl Acad Sci U S A* **111**, E3325-3334 (2014).

101. Warfel, N. A., Dolloff, N. G., Dicker, D. T., Malysz, J. & El-Deiry, W. S. CDK1 stabilizes HIF-1α via direct phosphorylation of Ser668 to promote tumor growth. *Cell Cycle* **12**, 3689–3701 (2013).

102. Daly, L. A. *et al*. Oxygen-dependent changes in binding partners and post-translational modifications regulate the abundance and activity of HIF-1α/2α. *Sci Signal* **14**, eabf6685 (2021).

103. Kalousi, A. *et al*. Casein kinase 1 regulates human hypoxia-inducible factor HIF-1. *J Cell Sci* **123**, 2976–2986 (2010).

104. Bao, L. *et al*. Methylation of hypoxia-inducible factor (HIF)-1α by G9a/GLP inhibits HIF-1 transcriptional activity and cell migration. *Nucleic Acids Res* **46**, 6576–6591 (2018).

105. Antonsson, C., Arulampalam, V., Whitelaw, M. L., Pettersson, S. & Poellinger, L. Constitutive function of the basic helix-loop-helix/PAS factor Arnt. Regulation of target promoters via the E box motif. *J Biol Chem* **270**, 13968–13972 (1995).

106. Swanson, H. I., Chan, W. K. & Bradfield, C. A. DNA binding specificities and pairing rules of the Ah receptor, ARNT, and SIM proteins. *J Biol Chem* **270**, 26292–26302 (1995).

107. Wenger, R. H., Stiehl, D. P. & Camenisch, G. Integration of oxygen signaling at the consensus HRE. *Sci STKE* **2005**, re12 (2005).

108. Cornish-Bowden, A. Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984. *Nucleic Acids Res* **13**, 3021–3030 (1985).

109. Rashid, I. *et al*. Genome-Wide Comparative Analysis of HIF Binding Sites in Cyprinus Carpio for In Silico Identification of Functional Hypoxia Response Elements. *Front Genet* **10**, 659 (2019).

110. Sabò, A. & Amati, B. Genome recognition by MYC. *Cold Spring Harb Perspect Med* **4**, a014191 (2014).

111. Bird, A. P. CpG-rich islands and the function of DNA methylation. *Nature* **321**, 209–213 (1986).

112. D'Anna, F. *et al*. DNA methylation repels binding of hypoxia-inducible transcription factors to maintain tumor immunotolerance. *Genome Biology* **21**, 182 (2020).

113. Schödel, J. *et al*. High-resolution genome-wide mapping of HIF-binding sites by ChIP-seq. *Blood* **117**, e207–e217 (2011).

114. Choudhry, H. *et al*. Extensive regulation of the non-coding transcriptome by hypoxia: role of HIF in releasing paused RNApol2. *EMBO Rep* **15**, 70–76 (2014).

115. S, I. & C, M. Molecular regulation of hypoxia through the lenses of noncoding RNAs and epitranscriptome. *Wiley interdisciplinary reviews. RNA* (2022) doi:10.1002/wrna.1750.

116. Zhao, Z. B., Chen, F. & Bai, X. F. Long Noncoding RNA MALAT1 Regulates Hepatocellular Carcinoma Growth Under Hypoxia via Sponging MicroRNA-200a. *Yonsei Med J* **60**, 727–734 (2019).

117. Zhang, Q. *et al*. Hypoxia-Induced lncRNA-NEAT1 Sustains the Growth of Hepatocellular Carcinoma via Regulation of miR-199a-3p/UCK2. *Front Oncol* **10**, 998 (2020).

118. Hayashi, Y., Yokota, A., Harada, H. & Huang, G. Hypoxia/pseudohypoxia-mediated activation of hypoxia-inducible factor-1α in cancer. *Cancer Sci* **110**, 1510–1517 (2019).

119. Dodd, K. M., Yang, J., Shen, M. H., Sampson, J. R. & Tee, A. R. mTORC1 drives HIF-1α and VEGF-A signalling via multiple mechanisms involving 4E-BP1, S6K1 and STAT3. *Oncogene* **34**, 2239–2250 (2015).

120. Masoud, G. N. & Li, W. HIF-1α pathway: role, regulation and intervention for cancer therapy. *Acta Pharm Sin B* **5**, 378–389 (2015).

121. Ohh, M. Tumor strengths and frailties: Cancer SUMmOns Otto's metabolism. *Nat Med* **18**, 30–31 (2012).

122. Jiang, B. H., Agani, F., Passaniti, A. & Semenza, G. L. V-SRC induces expression of hypoxia-inducible factor 1 (HIF-1) and transcription of genes encoding vascular endothelial growth

factor and enolase 1: involvement of HIF-1 in tumor progression. *Cancer Res* **57**, 5328–5335 (1997).

123. Harris, A. L. Hypoxia--a key regulatory factor in tumour growth. *Nat Rev Cancer* **2**, 38–47 (2002).

124. Berra, E. *et al*. Signaling angiogenesis via p42/p44 MAP kinase and hypoxia. *Biochem Pharmacol* **60**, 1171–1178 (2000).

125. Zundel, W. *et al*. Loss of PTEN facilitates HIF-1-mediated gene expression. *Genes Dev* **14**, 391–396 (2000).

126. Chen, C., Pore, N., Behrooz, A., Ismail-Beigi, F. & Maity, A. Regulation of glut1 mRNA by hypoxia-inducible factor-1. Interaction between H-ras and hypoxia. *J Biol Chem* **276**, 9519–9525 (2001).

127. Zelzer, E. *et al*. Insulin induces transcription of target genes through the hypoxia-inducible factor HIF-1alpha/ARNT. *EMBO J* **17**, 5085–5094 (1998).

128. Richard, D. E., Berra, E. & Pouyssegur, J. Nonhypoxic pathway mediates the induction of hypoxia-inducible factor 1alpha in vascular smooth muscle cells. *J Biol Chem* **275**, 26765–26771 (2000).

129. Laughner, E., Taghavi, P., Chiles, K., Mahon, P. C. & Semenza, G. L. HER2 (neu) signaling increases the rate of hypoxia-inducible factor 1alpha (HIF-1alpha) synthesis: novel mechanism for HIF-1-mediated vascular endothelial growth factor expression. *Mol Cell Biol* **21**, 3995–4004 (2001).

130. Sodhi, A. *et al*. The Kaposi's sarcoma-associated herpes virus G protein-coupled receptor up-regulates vascular endothelial growth factor expression and secretion through mitogen-activated protein kinase and p38 pathways acting on hypoxia-inducible factor 1alpha. *Cancer Res* **60**, 4873–4880 (2000).

131. Hirota, K. & Semenza, G. L. Rac1 activity is required for the activation of hypoxia-inducible factor 1. *J Biol Chem* **276**, 21166–21172 (2001).

132. Aragonés, J. *et al*. Evidence for the involvement of diacylglycerol kinase in the activation of hypoxia-inducible transcription factor 1 by low oxygen tension. *J Biol Chem* **276**, 10548–10555 (2001).

133. Lee, S. W. *et al*. Human hepatitis B virus X protein is a possible mediator of hypoxia-induced angiogenesis in hepatocarcinogenesis. *Biochem Biophys Res Commun* **268**, 456–461 (2000).

134. Kuiper, C. *et al*. Low ascorbate levels are associated with increased hypoxia-inducible factor-1 activity and an aggressive tumor phenotype in endometrial cancer. *Cancer Res* **70**, 5749–5758 (2010).

135. Fischer, A. P. & Miles, S. L. Ascorbic acid, but not dehydroascorbic acid increases intracellular vitamin C content to decrease Hypoxia Inducible Factor -1 alpha activity and reduce malignant potential in human melanoma. *Biomed Pharmacother* **86**, 502–513 (2017).

136. Vissers, M. C. M., Gunningham, S. P., Morrison, M. J., Dachs, G. U. & Currie, M. J. Modulation of hypoxia-inducible factor-1 alpha in cultured primary cells by intracellular ascorbate. *Free Radic Biol Med* **42**, 765–772 (2007).

137. Kuiper, C., Dachs, G. U., Currie, M. J. & Vissers, M. C. M. Intracellular ascorbate enhances hypoxia-inducible factor (HIF)-hydroxylase activity and preferentially suppresses the HIF-1 transcriptional response. *Free Radic Biol Med* **69**, 308–317 (2014).

138. Durán, R. V. *et al*. HIF-independent role of prolyl hydroxylases in the cellular response to amino acids. *Oncogene* **32**, 4549–4556 (2013).

139. Pagé, E. L., Chan, D. A., Giaccia, A. J., Levine, M. & Richard, D. E. Hypoxia-inducible factor-1alpha stabilization in nonhypoxic conditions: role of oxidation and intracellular ascorbate depletion. *Mol Biol Cell* **19**, 86–94 (2008).

140. Salnikow, K., An, W. G., Melillo, G., Blagosklonny, M. V. & Costa, M. Nickel-induced transformation shifts the balance between HIF-1 and p53 transcription factors. *Carcinogenesis* **20**, 1819–1823 (1999).

141. Muñoz-Sánchez, J. & Chánez-Cárdenas, M. E. The use of cobalt chloride as a chemical hypoxia model. *J Appl Toxicol* **39**, 556–570 (2019).

142. Selak, M. A. *et al*. Succinate links TCA cycle dysfunction to oncogenesis by inhibiting HIF-alpha prolyl hydroxylase. *Cancer Cell* **7**, 77–85 (2005).

143. Hewitson, K. S. *et al*. Structural and mechanistic studies on the inhibition of the hypoxia-inducible transcription factor hydroxylases by tricarboxylic acid cycle intermediates. *J Biol Chem* **282**, 3293–3301 (2007).

144. Lu, H., Forbes, R. A. & Verma, A. Hypoxia-inducible factor 1 activation by aerobic glycolysis implicates the Warburg effect in carcinogenesis. *J Biol Chem* **277**, 23111–23115 (2002).

145. Jung, S.-Y., Song, H. S., Park, S.-Y., Chung, S.-H. & Kim, Y.-J. Pyruvate promotes tumor angiogenesis through HIF-1-dependent PAI-1 expression. *Int J Oncol* **38**, 571–576 (2011).

146. De Saedeleer, C. J. *et al*. Lactate activates HIF-1 in oxidative but not in Warburg-phenotype human tumor cells. *PLoS One* **7**, e46571 (2012).

147. Losman, J.-A. & Kaelin, W. G. What a difference a hydroxyl makes: mutant IDH, (R)-2-hydroxyglutarate, and cancer. *Genes Dev* **27**, 836–852 (2013).

148. Ward, P. S. *et al*. The common feature of leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting alpha-ketoglutarate to 2-hydroxyglutarate. *Cancer Cell* **17**, 225–234 (2010).

149. Xu, W. *et al*. Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α-ketoglutarate-dependent dioxygenases. *Cancer Cell* **19**, 17–30 (2011).

150. Briggs, K. J. *et al*. Paracrine Induction of HIF by Glutamate in Breast Cancer: EglN1 Senses Cysteine. *Cell* **166**, 126–139 (2016).

151. Hes, F. J., Höppener, J. W., Luijt, R. B. van der & Lips, C. J. Von hippel-lindau disease. *Hered Cancer Clin Pract* **3**, 171–178 (2005).

152. Linehan, W. M., Rubin, J. S. & Bottaro, D. P. VHL loss of function and its impact on oncogenic signaling networks in clear cell renal cell carcinoma. *Int J Biochem Cell Biol* **41**, 753–756 (2009).

153. Banks, R. E. *et al*. Genetic and epigenetic analysis of von Hippel-Lindau (VHL) gene alterations and relationship with clinical variables in sporadic renal cancer. *Cancer Res* **66**, 2000–2011 (2006).

154. Kim, W. Y. & Kaelin, W. G. Role of VHL gene mutation in human cancer. *J Clin Oncol* **22**, 4991–5004 (2004).

155. Kondo, K. *et al*. Comprehensive mutational analysis of the VHL gene in sporadic renal cell carcinoma: relationship to clinicopathological parameters. *Genes Chromosomes Cancer* **34**, 58–68 (2002).

156. Lonser, R. R. *et al*. von Hippel-Lindau disease. *Lancet* **361**, 2059–2067 (2003).

157. Li, X. *et al*. Ubiquitin carboxyl-terminal hydrolase L1 promotes hypoxia-inducible factor 1-dependent tumor cell malignancy in spheroid models. *Cancer Sci* **111**, 239–252 (2020).

158. Kim, J. J. *et al*. WSB1 promotes tumor metastasis by inducing pVHL degradation. *Genes Dev* **29**, 2244–2257 (2015).

159. Doe, M. R., Ascano, J. M., Kaur, M. & Cole, M. D. Myc posttranscriptionally induces HIF1 protein and target gene expression in normal and cancer cells. *Cancer Res* **72**, 949–957 (2012).

160. Fu, R. *et al*. Wogonin inhibits multiple myeloma-stimulated angiogenesis via c-Myc/VHL/HIF-1α signaling axis. *Oncotarget* **7**, 5715–5727 (2016).

161. Gradin, K. *et al*. Functional interference between hypoxia and dioxin signal transduction pathways: competition for recruitment of the Arnt transcription factor. *Mol Cell Biol* **16**, 5221–5231 (1996).

162. Liu, Y. V. *et al*. RACK1 competes with HSP90 for binding to HIF-1alpha and is required for O(2)-independent and HSP90 inhibitor-induced degradation of HIF-1alpha. *Mol Cell* **25**, 207–217 (2007).

163. Ravi, R. *et al*. Regulation of tumor angiogenesis by p53-induced degradation of hypoxia-inducible factor 1alpha. *Genes Dev* **14**, 34–44 (2000).

164. Ryan, H. E., Lo, J. & Johnson, R. S. HIF-1 alpha is required for solid tumor formation and embryonic vascularization. *EMBO J* **17**, 3005–3015 (1998).

165. Maxwell, P. H. *et al*. Hypoxia-inducible factor-1 modulates gene expression in solid tumors and influences both angiogenesis and tumor growth. *Proc Natl Acad Sci U S A* **94**, 8104–8109 (1997).

166. Carmeliet, P. *et al*. Role of HIF-1alpha in hypoxia-mediated apoptosis, cell proliferation and tumour angiogenesis. *Nature* **394**, 485–490 (1998).

167. Dewhirst, M. W., Cao, Y. & Moeller, B. Cycling hypoxia and free radicals regulate angiogenesis and radiotherapy response. *Nat Rev Cancer* **8**, 425–437 (2008).

168. Zhong, H. *et al*. Overexpression of hypoxia-inducible factor 1alpha in common human cancers and their metastases. *Cancer Res* **59**, 5830–5835 (1999).

169. Goel, S. *et al*. Normalization of the vasculature for treatment of cancer and other diseases. *Physiol Rev* **91**, 1071–1121 (2011).

170. Folkman, J. Tumor angiogenesis: therapeutic implications. *N Engl J Med* **285**, 1182–1186 (1971).

171. Fu, Z., Mowday, A. M., Smaill, J. B., Hermans, I. F. & Patterson, A. V. Tumour Hypoxia-Mediated Immunosuppression: Mechanisms and Therapeutic Approaches to Improve Cancer Immunotherapy. *Cells* **10**, 1006 (2021).

172. Cosse, J.-P. & Michiels, C. Tumour hypoxia affects the responsiveness of cancer cells to chemotherapy and promotes cancer progression. *Anticancer Agents Med Chem* **8**, 790–797 (2008).

173. Vaupel, P., Mayer, A. & Höckel, M. Tumor hypoxia and malignant progression. *Methods Enzymol* **381**, 335–354 (2004).

174. Godet, I. *et al*. Fate-mapping post-hypoxic tumor cells reveals a ROS-resistant phenotype that promotes metastasis. *Nat Commun* **10**, 4862 (2019).

175. Horsman, M. R., Mortensen, L. S., Petersen, J. B., Busk, M. & Overgaard, J. Imaging hypoxia to improve radiotherapy outcome. *Nat Rev Clin Oncol* **9**, 674–687 (2012).

176. Muz, B., de la Puente, P., Azab, F. & Azab, A. K. The role of hypoxia in cancer progression, angiogenesis, metastasis, and resistance to therapy. *Hypoxia (Auckl)* **3**, 83–92 (2015).

177. Tate, J. G. *et al*. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* **47**, D941–D947 (2019).

178. Iommarini, L., Porcelli, A. M., Gasparre, G. & Kurelac, I. Non-Canonical Mechanisms Regulating Hypoxia-Inducible Factor 1 Alpha in Cancer. *Front Oncol* **7**, 286 (2017).

179. Akakura, N. *et al*. Constitutive expression of hypoxia-inducible factor-1alpha renders pancreatic cancer cells resistant to apoptosis induced by hypoxia and nutrient deprivation. *Cancer Res* **61**, 6548–6554 (2001).

180. Balamurugan, K. HIF-1 at the crossroads of hypoxia, inflammation, and cancer. *Int J Cancer* **138**, 1058–1066 (2016).

181. Ruan, K., Song, G. & Ouyang, G. Role of hypoxia in the hallmarks of human cancer. *J Cell Biochem* **107**, 1053–1062 (2009).

182. Rashid, M. *et al*. Up-down regulation of HIF-1α in cancer progression. *Gene* **798**, 145796 (2021).

183. Hirschhaeuser, F., Sattler, U. G. A. & Mueller-Klieser, W. Lactate: a metabolic key player in cancer. *Cancer Res* **71**, 6921–6925 (2011).

184. An, W. G. *et al*. Stabilization of wild-type p53 by hypoxia-inducible factor 1alpha. *Nature* **392**, 405–408 (1998).

185. Bruick, R. K. Expression of the gene encoding the proapoptotic Nip3 protein is induced by hypoxia. *Proc Natl Acad Sci U S A* **97**, 9082–9087 (2000).

186. Baek, J. H. *et al*. Hypoxia-induced VEGF enhances tumor survivability via suppression of serum deprivation-induced apoptosis. *Oncogene* **19**, 4621–4631 (2000).

187. Saikumar, P. *et al*. Role of hypoxia-induced Bax translocation and cytochrome c release in reoxygenation injury. *Oncogene* **17**, 3401–3415 (1998).

188. Dong, Z. *et al*. Gene promoter of apoptosis inhibitory protein IAP2: identification of enhancer elements and activation by severe hypoxia. *Biochem J* **364**, 413–421 (2002).

189. Graeber, T. G. *et al*. Hypoxia-mediated selection of cells with diminished apoptotic potential in solid tumours. *Nature* **379**, 88–91 (1996).

190. Greijer, A. E. & van der Wall, E. The role of hypoxia inducible factor 1 (HIF-1) in hypoxia induced apoptosis. *J Clin Pathol* **57**, 1009–1014 (2004).

191. Luoto, K. R., Kumareswaran, R. & Bristow, R. G. Tumor hypoxia as a driving force in genetic instability. *Genome Integr* **4**, 5 (2013).

192. Rofstad, E. K., Johnsen, N. M. & Lyng, H. Hypoxia-induced tetraploidisation of a diploid human melanoma cell line in vitro. *Br J Cancer Suppl* **27**, S136-139 (1996).

193. Coquelle, A., Toledo, F., Stern, S., Bieth, A. & Debatisse, M. A new role for hypoxia in tumor progression: induction of fragile site triggering genomic rearrangements and formation of complex DMs and HSRs. *Mol Cell* **2**, 259–265 (1998).

194. Bristow, R. G. & Hill, R. P. Hypoxia and metabolism. Hypoxia, DNA repair and genetic instability. *Nat Rev Cancer* **8**, 180–192 (2008).

195. Chan, N. & Bristow, R. G. 'Contextual' synthetic lethality and/or loss of heterozygosity: tumor hypoxia and modification of DNA repair. *Clin Cancer Res* **16**, 4553–4560 (2010).

196. Comerford, K. M. *et al*. Hypoxia-inducible factor-1-dependent regulation of the multidrug resistance (MDR1) gene. *Cancer Res* **62**, 3387–3394 (2002).

197. St Croix, B., Man, S. & Kerbel, R. S. Reversal of intrinsic and acquired forms of drug resistance by hyaluronidase treatment of solid tumors. *Cancer Lett* **131**, 35–44 (1998).

198. Lv, Y. *et al*. Hypoxia-inducible factor-1α induces multidrug resistance protein in colon cancer. *Onco Targets Ther* **8**, 1941–1948 (2015).

199. Tannock, I. F. The relation between cell proliferation and the vascular system in a transplanted mouse mammary tumour. *Br J Cancer* **22**, 258–273 (1968).

200. Tannock, I. Cell kinetics and chemotherapy: a critical review. *Cancer Treat Rep* **62**, 1117–1133 (1978).

201. Hirst, D. G. & Denekamp, J. Tumour cell proliferation in relation to the vasculature. *Cell Tissue Kinet* **12**, 31–42 (1979).

202. Ljungkvist, A. S. E. *et al*. Vascular architecture, hypoxia, and proliferation in first-generation xenografts of human head-and-neck squamous cell carcinomas. *Int J Radiat Oncol Biol Phys* **54**, 215–228 (2002).

203. Sugawara, I., Akiyama, S., Scheper, R. J. & Itoyama, S. Lung resistance protein (LRP) expression in human normal tissues in comparison with that of MDR1 and MRP. *Cancer Lett* **112**, 23–31 (1997).

204. Trédan, O., Galmarini, C. M., Patel, K. & Tannock, I. F. Drug resistance and the solid tumor microenvironment. *J Natl Cancer Inst* **99**, 1441–1454 (2007).

205. Pastorek, J. *et al*. Cloning and characterization of MN, a human tumor-associated protein with a domain homologous to carbonic anhydrase and a putative helix-loop-helix DNA binding segment. *Oncogene* **9**, 2877–2888 (1994).

206. Ivanov, S. V. *et al*. Down-regulation of transmembrane carbonic anhydrases in renal cell carcinoma cell lines by wild-type von Hippel-Lindau transgenes. *Proc Natl Acad Sci U S A* **95**, 12596–12601 (1998).

207. Tanaka, N. *et al*. Expression of carbonic anhydrase 9, a potential intrinsic marker of hypoxia, is associated with poor prognosis in oesophageal squamous cell carcinoma. *Br J Cancer* **99**, 1468–1475 (2008).

208. Vaupel, P. W., Frinak, S. & Bicher, H. I. Heterogeneous oxygen partial pressure and pH distribution in C3H mouse mammary adenocarcinoma. *Cancer Res* **41**, 2008–2013 (1981).

209. Vaupel, P. W. The influence of tumor blood flow and microenvironmental factors on the efficacy of radiation, drugs and localized hyperthermia. *Klin Padiatr* **209**, 243–249 (1997).

210. Fuchs, E. Scratching the surface of skin development. *Nature* **445**, 834–842 (2007).

211. Soeda, A. *et al*. Hypoxia promotes expansion of the CD133-positive glioma stem cells through activation of HIF-1alpha. *Oncogene* **28**, 3949–3959 (2009).

212. McCord, A. M. *et al*. Physiologic oxygen concentration enhances the stem-like properties of CD133+ human glioblastoma cells in vitro. *Mol Cancer Res* **7**, 489–497 (2009).

213. Heddleston, J. M., Li, Z., McLendon, R. E., Hjelmeland, A. B. & Rich, J. N. The hypoxic microenvironment maintains glioblastoma stem cells and promotes reprogramming towards a cancer stem cell phenotype. *Cell Cycle* **8**, 3274–3284 (2009).

214. Koh, M. Y. & Powis, G. Passing the baton: the HIF switch. *Trends Biochem Sci* **37**, 364–372 (2012).

215. Rakotomalala, A., Escande, A., Furlan, A., Meignan, S. & Lartigau, E. Hypoxia in Solid Tumors: How Low Oxygenation Impacts the 'Six Rs' of Radiotherapy. *Front Endocrinol (Lausanne)* **12**, 742215 (2021).

216. Barsoum, I. B., Smallwood, C. A., Siemens, D. R. & Graham, C. H. A mechanism of hypoxia-mediated escape from adaptive immunity in cancer cells. *Cancer Res* **74**, 665–674 (2014).

217. Siemens, D. R. *et al*. Hypoxia increases tumor cell shedding of MHC class I chain-related molecule: role of nitric oxide. *Cancer Res* **68**, 4746–4753 (2008).

218. Sceneay, J. *et al*. Primary tumor hypoxia recruits CD11b+/Ly6Cmed/Ly6G+ immune suppressor cells and compromises NK cell cytotoxicity in the premetastatic niche. *Cancer Res* **72**, 3906–3911 (2012).

219. Facciabene, A. *et al*. Tumour hypoxia promotes tolerance and angiogenesis via CCL28 and T(reg) cells. *Nature* **475**, 226–230 (2011).

220. Kouvaras, E. *et al*. Hypoxia-inducible factor 1-alpha and vascular endothelial growth factor in cartilage tumors. *Biotech Histochem* **94**, 283–289 (2019).

221. He, J., Hu, Y., Hu, M. & Li, B. Development of PD-1/PD-L1 Pathway in Tumor Immune Microenvironment and Treatment for Non-Small Cell Lung Cancer. *Sci Rep* **5**, 13110 (2015).

222. Fujii, T. *et al*. VEGF-A Is Associated With the Degree of TILs and PD-L1 Expression in Primary Breast Cancer. *In Vivo* **34**, 2641–2646 (2020).

223. López-Novoa, J. M. & Nieto, M. A. Inflammation and EMT: an alliance towards organ fibrosis and cancer progression. *EMBO Mol Med* **1**, 303–314 (2009).

224. Jiang, J., Tang, Y. & Liang, X. EMT: a new vision of hypoxia promoting cancer progression. *Cancer Biol Ther* **11**, 714–723 (2011).

225. Tang, N. *et al*. Loss of HIF-1alpha in endothelial cells disrupts a hypoxia-driven VEGF autocrine loop necessary for tumorigenesis. *Cancer Cell* **6**, 485–495 (2004).

226. Zhang, L. *et al*. Tumor-derived vascular endothelial growth factor up-regulates angiopoietin-2 in host endothelium and destabilizes host vasculature, supporting angiogenesis in ovarian cancer. *Cancer Res* **63**, 3403–3412 (2003).

227. Mazzieri, R. *et al*. Targeting the ANG2/TIE2 axis inhibits tumor growth and metastasis by impairing angiogenesis and disabling rebounds of proangiogenic myeloid cells. *Cancer Cell* **19**, 512–526 (2011).

228. Schito, L. *et al*. Hypoxia-inducible factor 1-dependent expression of platelet-derived growth factor B promotes lymphatic metastasis of hypoxic breast cancer cells. *Proc Natl Acad Sci U S A* **109**, E2707-2716 (2012).

229. Zahra, F. T., Sajib, M. S. & Mikelis, C. M. Role of bFGF in Acquired Resistance upon Anti-VEGF Therapy in Cancer. *Cancers (Basel)* **13**, 1422 (2021).

230. Krock, B. L., Skuli, N. & Simon, M. C. Hypoxia-induced angiogenesis: good and evil. *Genes Cancer* **2**, 1117–1133 (2011).

231. Vaupel, P. Tumor microenvironmental physiology and its implications for radiation oncology. *Semin Radiat Oncol* **14**, 198–206 (2004).

232. Hapke, R. Y. & Haake, S. M. Hypoxia-induced epithelial to mesenchymal transition in cancer. *Cancer Lett* **487**, 10–20 (2020).

233. Nagelkerke, A. *et al*. Hypoxia stimulates migration of breast cancer cells via the PERK/ATF4/LAMP3-arm of the unfolded protein response. *Breast Cancer Res* **15**, R2 (2013).

234. Chen, J., Imanaka, N., Chen, J. & Griffin, J. D. Hypoxia potentiates Notch signaling in breast cancer leading to decreased E-cadherin expression and increased cell migration and invasion. *Br J Cancer* **102**, 351–360 (2010).

235. Evans, A. J. *et al*. VHL promotes E2 box-dependent E-cadherin transcription by HIF-mediated regulation of SIP1 and snail. *Mol Cell Biol* **27**, 157–169 (2007).

236. Heldin, C.-H., Vanlandewijck, M. & Moustakas, A. Regulation of EMT by TGFβ in cancer. *FEBS Lett* **586**, 1959–1970 (2012).

237. Ben-Yosef, Y., Lahat, N., Shapiro, S., Bitterman, H. & Miller, A. Regulation of endothelial matrix metalloproteinase-2 by hypoxia/reoxygenation. *Circ Res* **90**, 784–791 (2002).

238. Caley, M. P., Martins, V. L. C. & O'Toole, E. A. Metalloproteinases and Wound Healing. *Adv Wound Care (New Rochelle)* **4**, 225–234 (2015).

239. Gilkes, D. M., Bajpai, S., Chaturvedi, P., Wirtz, D. & Semenza, G. L. Hypoxia-inducible factor 1 (HIF-1) promotes extracellular matrix remodeling under hypoxic conditions by inducing P4HA1, P4HA2, and PLOD2 expression in fibroblasts. *J Biol Chem* **288**, 10819–10829 (2013).

240. Erler, J. T. & Giaccia, A. J. Lysyl oxidase mediates hypoxic control of metastasis. *Cancer Res* **66**, 10238–10241 (2006).

241. Ji, K. *et al*. Expression of hypoxic markers and their prognostic significance in soft tissue sarcoma. *Oncology letters* **9**, (2015).

242. Dillekås, H., Rogers, M. S. & Straume, O. Are 90% of deaths from cancer caused by metastases? *Cancer Med* **8**, 5574–5576 (2019).

243. Kaczmarek, M. *et al*. Metal ions-stimulated iron oxidation in hydroxylases facilitates stabilization of HIF-1 alpha protein. *Toxicol Sci* **107**, 394–403 (2009).

244. Zhang, Z., Yan, J., Chang, Y., ShiDu Yan, S. & Shi, H. Hypoxia inducible factor-1 as a target for neurodegenerative diseases. *Curr Med Chem* **18**, 4335–4343 (2011).

245. Mole, D. R. *et al*. 2-oxoglutarate analogue inhibitors of HIF prolyl hydroxylase. *Bioorg Med Chem Lett* **13**, 2677–2680 (2003).

246. Harvey, A. J., Kind, K. L. & Thompson, J. G. Effect of the oxidative phosphorylation uncoupler 2,4-dinitrophenol on hypoxia-inducible factor-regulated gene expression in bovine blastocysts. *Reprod Fertil Dev* **16**, 665–673 (2004).

247. Kim, J., So, D., Shin, H.-W., Chun, Y.-S. & Park, J.-W. HIF-1α Upregulation due to Depletion of the Free Ubiquitin Pool. *J Korean Med Sci* **30**, 1388–1395 (2015).

248. Horev-Azaria, L. *et al*. Predictive toxicology of cobalt nanoparticles and ions: comparative in vitro study of different cellular models using methods of knowledge discovery from data. *Toxicol Sci* **122**, 489–501 (2011).

249. Semenza, G. L., Roth, P. H., Fang, H. M. & Wang, G. L. Transcriptional regulation of genes encoding glycolytic enzymes by hypoxia-inducible factor 1. *J Biol Chem* **269**, 23757–23763 (1994).

250. Ouiddir, A., Planès, C., Fernandes, I., VanHesse, A. & Clerici, C. Hypoxia upregulates activity and expression of the glucose transporter GLUT1 in alveolar epithelial cells. *Am J Respir Cell Mol Biol* **21**, 710–718 (1999).

251. Lendahl, U., Lee, K. L., Yang, H. & Poellinger, L. Generating specificity and diversity in the transcriptional response to hypoxia. *Nat Rev Genet* **10**, 821–832 (2009).

252. Zhigalova, N., Artemov, A., Mazur, A. & Prokhortchouk, E. Transcriptome sequencing revealed differences in the response of renal cancer cells to hypoxia and CoCl 2 treatment. *F1000Res* **4**, 1518 (2015).

253. Wu, D. & Yotnda, P. Induction and testing of hypoxia in cell culture. *J Vis Exp* 2899 (2011) doi:10.3791/2899.

254. Pilgaard, L. *et al*. Transcriptional signature of human adipose tissue-derived stem cells (hASCs) preconditioned for chondrogenesis in hypoxic conditions. *Exp Cell Res* **315**, 1937–1952 (2009).

255. Bader, S. B., Dewhirst, M. W. & Hammond, E. M. Cyclic Hypoxia: An Update on Its Characteristics, Methods to Measure It and Biological Implications in Cancer. *Cancers (Basel)* **13**, E23 (2020).

256. Pirkmajer, S. & Chibalin, A. V. Serum starvation: caveat emptor. *Am J Physiol Cell Physiol* **301**, C272-279 (2011).

257. Park, J. E. *et al*. Pulsed SILAC-based proteomic analysis unveils hypoxia- and serum starvation-induced de novo protein synthesis with PHD finger protein 14 (PHF14) as a hypoxia sensitive epigenetic regulator in cell cycle progression. *Oncotarget* **10**, 2136–2150 (2019).

258. Hay, E. D. Role of cell-matrix contacts in cell migration and epithelial-mesenchymal transformation. *Cell Differ Dev* **32**, 367–375 (1990).

259. Zeisberg, M. & Neilson, E. G. Biomarkers for epithelial-mesenchymal transitions. *J Clin Invest* **119**, 1429–1437 (2009).

260. Kalluri, R. & Weinberg, R. A. The basics of epithelial-mesenchymal transition. *J Clin Invest* **119**, 1420–1428 (2009).

261. Thiery, J. P. Epithelial-mesenchymal transitions in tumour progression. *Nat Rev Cancer* **2**, 442–454 (2002).

262. Fidler, I. J. The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. *Nat Rev Cancer* **3**, 453–458 (2003).

263. Brabletz, T. *et al*. Variable beta-catenin expression in colorectal cancers indicates tumor progression driven by the tumor environment. *Proc Natl Acad Sci U S A* **98**, 10356–10361 (2001).

264. Zeisberg, M., Shah, A. A. & Kalluri, R. Bone morphogenic protein-7 induces mesenchymal to epithelial transition in adult renal fibroblasts and facilitates regeneration of injured kidney. *J Biol Chem* **280**, 8094–8100 (2005).

265. Duband, J. L. & Thiery, J. P. Appearance and distribution of fibronectin during chick embryo gastrulation and neurulation. *Dev Biol* **94**, 337–350 (1982).

266. Sauka-Spengler, T. & Bronner-Fraser, M. A gene regulatory network orchestrates neural crest formation. *Nat Rev Mol Cell Biol* **9**, 557–568 (2008).

267. Knecht, A. K. & Bronner-Fraser, M. Induction of the neural crest: a multigene process. *Nat Rev Genet* **3**, 453–461 (2002).

268. Liem, K. F., Jessell, T. M. & Briscoe, J. Regulation of the neural patterning activity of sonic hedgehog by secreted BMP inhibitors expressed by notochord and somites. *Development* **127**, 4855–4866 (2000).

269. Villanueva, S., Glavic, A., Ruiz, P. & Mayor, R. Posteriorization by FGF, Wnt, and retinoic acid is required for neural crest induction. *Dev Biol* **241**, 289–301 (2002).

270. Karafiat, V., Dvorakova, M., Pajer, P., Cermak, V. & Dvorak, M. Melanocyte fate in neural crest is triggered by Myb proteins through activation of c-kit. *Cell Mol Life Sci* **64**, 2975–2984 (2007).

271. Sela-Donenfeld, D. & Kalcheim, C. Localized BMP4-noggin interactions generate the dynamic patterning of noggin expression in somites. *Dev Biol* **246**, 311–328 (2002).

272. Burstyn-Cohen, T., Stanleigh, J., Sela-Donenfeld, D. & Kalcheim, C. Canonical Wnt activity regulates trunk neural crest delamination linking BMP/noggin signaling with G1/S transition. *Development* **131**, 5327–5339 (2004).

273. Garcia-Martinez, V. & Schoenwolf, G. C. Primitive-streak origin of the cardiovascular system in avian embryos. *Dev Biol* **159**, 706–719 (1993).

274. Thiery, J. P., Acloque, H., Huang, R. Y. J. & Nieto, M. A. Epithelial-Mesenchymal Transitions in Development and Disease. *Cell* **139**, 871–890 (2009).

275. Di Bella, A. *et al*. An appraisal of intermediate filament expression in adult and developing pancreas: vimentin is expressed in alpha cells of rat and mouse embryos. *J Histochem Cytochem* **57**, 577–586 (2009).

276. Humphreys, B. D. *et al*. Fate tracing reveals the pericyte and not epithelial origin of myofibroblasts in kidney fibrosis. *Am J Pathol* **176**, 85–97 (2010).

277. Smith, B. N. & Bhowmick, N. A. Role of EMT in Metastasis and Therapy Resistance. *J Clin Med* **5**, E17 (2016).

278. Rout-Pitt, N., Farrow, N., Parsons, D. & Donnelley, M. Epithelial mesenchymal transition (EMT): a universal process in lung diseases with implications for cystic fibrosis pathophysiology. *Respir Res* **19**, 136 (2018).

279. Okada, H., Danoff, T. M., Kalluri, R. & Neilson, E. G. Early role of Fsp1 in epithelial-mesenchymal transformation. *Am J Physiol* **273**, F563-574 (1997).

280. Rastaldi, M. P. *et al*. Epithelial-mesenchymal transition of tubular epithelial cells in human renal biopsies. *Kidney Int* **62**, 137–146 (2002).

281. Walker, J. L., Bleaken, B. M., Romisher, A. R., Alnwibit, A. A. & Menko, A. S. In wound repair vimentin mediates the transition of mesenchymal leader cells to a myofibroblast phenotype. *Mol Biol Cell* **29**, 1555–1570 (2018).

282. Strutz, F. *et al*. Role of basic fibroblast growth factor-2 in epithelial-mesenchymal transformation. *Kidney Int* **61**, 1714–1728 (2002).

283. Pierce, G. F., Mustoe, T. A., Altrock, B. W., Deuel, T. F. & Thomason, A. Role of platelet-derived growth factor in wound healing. *J Cell Biochem* **45**, 319–326 (1991).

284. Wenczak, B. A., Lynch, J. B. & Nanney, L. B. Epidermal growth factor receptor distribution in burn wounds. Implications for growth factor-mediated repair. *J Clin Invest* **90**, 2392–2401 (1992).

285. Penn, J. W., Grobbelaar, A. O. & Rolfe, K. J. The role of the TGF-β family in wound healing, burns and scarring: a review. *Int J Burns Trauma* **2**, 18–28 (2012).

286. Yang, J. & Weinberg, R. A. Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. *Dev Cell* **14**, 818–829 (2008).

287. Medici, D., Hay, E. D. & Olsen, B. R. Snail and Slug promote epithelial-mesenchymal transition through beta-catenin-T-cell factor-4-dependent expression of transforming growth factor-beta3. *Mol Biol Cell* **19**, 4875–4887 (2008).

288. Savagner, P., Yamada, K. M. & Thiery, J. P. The zinc-finger protein slug causes desmosome dissociation, an initial and necessary step for growth factor-induced epithelial-mesenchymal transition. *J Cell Biol* **137**, 1403–1419 (1997).

289. Peinado, H., Olmeda, D. & Cano, A. Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? *Nat Rev Cancer* **7**, 415–428 (2007).

290. Aigner, K. *et al*. The transcription factor ZEB1 (deltaEF1) promotes tumour cell dedifferentiation by repressing master regulators of epithelial polarity. *Oncogene* **26**, 6979–6988 (2007).

291. Ren, J., Chen, Y., Song, H., Chen, L. & Wang, R. Inhibition of ZEB1 reverses EMT and chemoresistance in docetaxel-resistant human lung adenocarcinoma cell line. *J Cell Biochem* **114**, 1395–1403 (2013).

292. Van de Putte, T. *et al*. Mice lacking ZFHX1B, the gene that codes for Smad-interacting protein-1, reveal a role for multiple neural crest cell defects in the etiology of Hirschsprung disease-mental retardation syndrome. *Am J Hum Genet* **72**, 465–470 (2003).

293. Vandewalle, C., Van Roy, F. & Berx, G. The role of the ZEB family of transcription factors in development and disease. *Cell Mol Life Sci* **66**, 773–787 (2009).

294. Wang, Y., Liu, J., Ying, X., Lin, P. C. & Zhou, B. P. Twist-mediated Epithelial-mesenchymal Transition Promotes Breast Tumor Cell Invasion via Inhibition of Hippo Pathway. *Sci Rep* **6**, 24606 (2016).

295. Teng, Y. & Li, X. The roles of HLH transcription factors in epithelial mesenchymal transition and multiple molecular mechanisms. *Clin Exp Metastasis* **31**, 367–377 (2014).

296. Mani, S. A. *et al*. Mesenchyme Forkhead 1 (FOXC2) plays a key role in metastasis and is associated with aggressive basal-like breast cancers. *Proc Natl Acad Sci U S A* **104**, 10069–10074 (2007).

297. Hartwell, K. A. *et al*. The Spemann organizer gene, Goosecoid, promotes tumor metastasis. *Proc Natl Acad Sci U S A* **103**, 18969–18974 (2006).

298. Chung, V. Y. *et al*. GRHL2-miR-200-ZEB1 maintains the epithelial status of ovarian cancer through transcriptional regulation and histone modification. *Sci Rep* **6**, 19943 (2016).

299. Gan, L. *et al*. The polycomb group protein EZH2 induces epithelial-mesenchymal transition and pluripotent phenotype of gastric cancer cells by binding to PTEN promoter. *J Hematol Oncol* **11**, 9 (2018).

300. Hasegawa, S. *et al*. A crucial epithelial to mesenchymal transition regulator, Sox4/Ezh2 axis is closely related to the clinical outcome in pancreatic cancer patients. *Int J Oncol* **48**, 145–152 (2016).

301. Call, K. M. *et al*. Isolation and characterization of a zinc finger polypeptide gene at the human chromosome 11 Wilms' tumor locus. *Cell* **60**, 509–520 (1990).

302. Ruteshouser, E. C., Robinson, S. M. & Huff, V. Wilms tumor genetics: mutations in WT1, WTX, and CTNNB1 account for only about one-third of tumors. *Genes Chromosomes Cancer* **47**, 461–470 (2008).

303. Huff, V. Wilms' tumours: about tumour suppressor genes, an oncogene and a chameleon gene. *Nat Rev Cancer* **11**, 111–121 (2011).

304. Hohenstein, P. & Hastie, N. D. The many facets of the Wilms' tumour gene, WT1. *Hum Mol Genet* **15 Spec No 2**, R196-201 (2006).

305. Hastie, N. D. Wilms' tumour 1 (WT1) in development, homeostasis and disease. *Development* **144**, 2862–2872 (2017).

306. Kreidberg, J. A. *et al*. WT-1 is required for early kidney development. *Cell* **74**, 679–691 (1993).

307. Hartwig, S. *et al*. Genomic characterization of Wilms' tumor suppressor 1 targets in nephron progenitor cells during kidney development. *Development* **137**, 1189–1203 (2010).

308. Motamedi, F. J. *et al*. WT1 controls antagonistic FGF and BMP-pSMAD pathways in early renal progenitors. *Nat Commun* **5**, 4444 (2014).

309. Sim, E. U.-H. *et al*. Wnt-4 regulation by the Wilms' tumour suppressor gene, WT1. *Oncogene* **21**, 2948–2960 (2002).

310. Davies, J. A. *et al*. Development of an siRNA-based method for repressing specific genes in renal organ culture and its use to show that the Wt1 tumour suppressor is required for nephron differentiation. *Hum Mol Genet* **13**, 235–246 (2004).

311. Essafi, A. *et al*. A wt1-controlled chromatin switching mechanism underpins tissue-specific wnt4 activation and repression. *Dev Cell* **21**, 559–574 (2011).

312. Martínez-Estrada, O. M. *et al*. Wt1 is required for cardiovascular progenitor cell formation through transcriptional control of Snail and E-cadherin. *Nat Genet* **42**, 89–93 (2010).

313. Velecela, V. *et al*. WT1 regulates the expression of inhibitory chemokines during heart development. *Hum Mol Genet* **22**, 5083–5095 (2013).

314. Madden, S. L. *et al*. Transcriptional repression mediated by the WT1 Wilms tumor gene product. *Science* **253**, 1550–1553 (1991).

315. Koesters, R. *et al*. WT1 is a tumor-associated antigen in colon cancer that can be recognized by in vitro stimulated cytotoxic T cells. *Int J Cancer* **109**, 385–392 (2004).

316. Loeb, D. M. *et al*. Wilms' tumor suppressor gene (WT1) is expressed in primary breast tumors despite tumor-specific promoter methylation. *Cancer Res* **61**, 921–925 (2001).

317. Oji, Y. *et al*. Overexpression of the Wilms' tumor gene W T1 in primary astrocytic tumors. *Cancer Sci* **95**, 822–827 (2004).

318. Amini Nik, S. *et al*. Upregulation of Wilms' tumor gene 1 (WT1) in desmoid tumors. *Int J Cancer* **114**, 202–208 (2005).

319. Miyagi, T. *et al*. Expression of the candidate Wilm's tumor gene, WT1, in human leukemia cells. *Leukemia* **7**, 970–977 (1993).

320. Miwa, H., Beran, M. & Saunders, G. F. Expression of the Wilms' tumor gene (WT1) in human leukemias. *Leukemia* **6**, 405–409 (1992).

321. Menssen, H. D. *et al*. Presence of Wilms' tumor gene (wt1) transcripts and the WT1 nuclear protein in the majority of human acute leukemias. *Leukemia* **9**, 1060–1067 (1995).

322. T, Y. *et al*. Growth inhibition of human leukemic cells by WT1 (Wilms tumor gene) antisense oligodeoxynucleotides: implications for the involvement of WT1 in leukemogenesis. *Blood* **87**, (1996).

323. Ito, K. *et al*. Antiapoptotic function of 17AA(+)WT1 (Wilms' tumor gene) isoforms on the intrinsic apoptosis pathway. *Oncogene* **25**, 4217–4229 (2006).

324. Sampson, V. B. *et al*. Wilms' tumor protein induces an epithelial-mesenchymal hybrid differentiation state in clear cell renal cell carcinoma. *PLoS One* **9**, e102041 (2014).

325. Plönes, T. *et al*. Turning back the Wheel: Inducing Mesenchymal to Epithelial Transition via Wilms Tumor 1 Knockdown in Human Mesothelioma Cell Lines to Influence Proliferation, Invasiveness, and Chemotaxis. *Pathol Oncol Res* **23**, 723–730 (2017).

326. Toska, E. & Roberts, S. G. E. Mechanisms of transcriptional regulation by WT1 (Wilms' tumour 1). *Biochem J* **461**, 15–32 (2014).

327. Sakharkar, M. K., Perumal, B. S., Sakharkar, K. R. & Kangueane, P. An analysis on gene architecture in human and mouse genomes. *In Silico Biol* **5**, 347–365 (2005).

328. Hofmann, W., Royer, H. D., Drechsler, M., Schneider, S. & Royer-Pokora, B. Characterization of the transcriptional regulatory region of the human WT1 gene. *Oncogene* **8**, 3123–3132 (1993).

329. Eccles, M. R., Grubb, G., Ogawa, O., Szeto, J. & Reeve, A. E. Cloning of novel Wilms tumor gene (WT1) cDNAs; evidence for antisense transcription of WT1. *Oncogene* **9**, 2059–2063 (1994).

330. Malik, K. T., Wallace, J. I., Ivins, S. M. & Brown, K. W. Identification of an antisense WT1 promoter in intron 1: implications for WT1 gene regulation. *Oncogene* **11**, 1589–1595 (1995).

331. Moorwood, K. *et al*. Antisense WT1 transcription parallels sense mRNA and protein expression in fetal kidney and can elevate protein levels in vitro. *J Pathol* **185**, 352–359 (1998).

332. Dallosso, A. R. *et al*. Alternately spliced WT1 antisense transcripts interact with WT1 sense RNA and show epigenetic and splicing defects in cancer. *RNA* **13**, 2287–2299 (2007).

333. Zhang, Y., Fan, L.-J., Zhang, Y., Jiang, J. & Qi, X.-W. Long Non-coding Wilms Tumor 1 Antisense RNA in the Development and Progression of Malignant Tumors. *Front Oncol* **10**, 35 (2020).

334. Lv, L., Chen, G., Zhou, J., Li, J. & Gong, J. WT1-AS promotes cell apoptosis in hepatocellular carcinoma through down-regulating of WT1. *J Exp Clin Cancer Res* **34**, 119 (2015).

335. Du, T. *et al*. Decreased expression of long non-coding RNA WT1-AS promotes cell proliferation and invasion in gastric cancer. *Biochim Biophys Acta* **1862**, 12–19 (2016).

336. Carpenter, B. *et al*. BASP1 is a transcriptional cosuppressor for the Wilms' tumor suppressor protein WT1. *Mol Cell Biol* **24**, 537–549 (2004).

337. Maheswaran, S. *et al*. Inhibition of cellular proliferation by the Wilms tumor suppressor WT1 requires association with the inducible chaperone Hsp70. *Genes Dev* **12**, 1108–1120 (1998).

338. Yang, L., Han, Y., Suarez Saiz, F., Saurez Saiz, F. & Minden, M. D. A tumor suppressor and oncogene: the WT1 story. *Leukemia* **21**, 868–876 (2007).

339. Menke, A. L., van der Eb, A. J. & Jochemsen, A. G. The Wilms' tumor 1 gene: oncogene or tumor suppressor gene? *Int Rev Cytol* **181**, 151–212 (1998).

340. Bardeesy, N. & Pelletier, J. Overlapping RNA and DNA binding domains of the wt1 tumor suppressor gene product. *Nucleic Acids Res* **26**, 1784–1792 (1998).

341. Nishikawa, T., Wojciak, J. M., Dyson, H. J. & Wright, P. E. RNA Binding by the KTS Splice Variants of the Wilms' Tumor Suppressor Protein WT1. *Biochemistry* **59**, 3889–3901 (2020).

342. Bruening, W., Moffett, P., Chia, S., Heinrich, G. & Pelletier, J. Identification of nuclear localization signals within the zinc fingers of the WT1 tumor suppressor gene product. *FEBS Lett* **393**, 41–47 (1996).

343. Maheswaran, S., Englert, C., Bennett, P., Heinrich, G. & Haber, D. A. The WT1 gene product stabilizes p53 and inhibits p53-mediated apoptosis. *Genes Dev* **9**, 2143–2156 (1995).

344. Scharnhorst, V., Dekker, P., van der Eb, A. J. & Jochemsen, A. G. Physical interaction between Wilms tumor 1 and p73 proteins modulates their functions. *J Biol Chem* **275**, 10202–10211 (2000).

345. Wang, W., Lee, S. B., Palmer, R., Ellisen, L. W. & Haber, D. A. A functional interaction with CBP contributes to transcriptional activation by the Wilms tumor suppressor WT1. *J Biol Chem* **276**, 16810–16816 (2001).

346. Johnstone, R. W. *et al*. A novel repressor, par-4, modulates transcription and growth suppression functions of the Wilms' tumor suppressor WT1. *Mol Cell Biol* **16**, 6945–6956 (1996).

347. Cunningham, F. *et al*. Ensembl 2022. *Nucleic Acids Res* **50**, D988–D995 (2022).

348. Haber, D. A. *et al*. Alternative splicing and genomic structure of the Wilms tumor gene WT1. *Proc Natl Acad Sci U S A* **88**, 9618–9622 (1991).

349. Dallosso, A. R. *et al*. Genomic imprinting at the WT1 gene involves a novel coding transcript (AWT1) that shows deregulation in Wilms' tumours. *Hum Mol Genet* **13**, 405–415 (2004).

350. Dechsukhum, C., Ware, J. L., Ferreira-Gonzalez, A., Wilkinson, D. S. & Garrett, C. T. Detection of a novel truncated WT1 transcript in human neoplasia. *Mol Diagn* **5**, 117–128 (2000).

351. Hewitt, S. M. & Saunders, G. F. Differentially spliced exon 5 of the Wilms' tumor gene WT1 modifies gene function. *Anticancer Res* **16**, 621–626 (1996).

352. Natoli, T. A. *et al*. A mammal-specific exon of WT1 is not required for development or fertility. *Mol Cell Biol* **22**, 4433–4438 (2002).

353. Hastie, N. D. Life, sex, and WT1 isoforms--three amino acids can make all the difference. *Cell* **106**, 391–394 (2001).

354. Klamt, B. *et al*. Frasier syndrome is caused by defective alternative splicing of WT1 leading to an altered ratio of WT1 +/-KTS splice isoforms. *Hum Mol Genet* **7**, 709–714 (1998).

355. Bickmore, W. A. *et al*. Modulation of DNA binding specificity by alternative splicing of the Wilms tumor wt1 gene transcript. *Science* **257**, 235–237 (1992).

356. Laity, J. H., Dyson, H. J. & Wright, P. E. Molecular basis for modulation of biological function by alternate splicing of the Wilms' tumor suppressor protein. *Proc Natl Acad Sci U S A* **97**, 11932–11935 (2000).

357. Larsson, S. H. *et al*. Subnuclear localization of WT1 in splicing or transcription factor domains is regulated by alternative splicing. *Cell* **81**, 391–401 (1995).

358. Davies, R. C. *et al*. WT1 interacts with the splicing factor U2AF65 in an isoform-dependent manner and can be incorporated into spliceosomes. *Genes Dev* **12**, 3217–3225 (1998).

359. Bharathavikru, R. *et al*. Transcription factor Wilms' tumor 1 regulates developmental RNAs through 3' UTR interaction. *Genes Dev* **31**, 347–352 (2017).

360. Shabalina, S. A. & Koonin, E. V. Origins and evolution of eukaryotic RNA interference. *Trends Ecol Evol* **23**, 578–587 (2008).

361. Bartel, D. P. Metazoan MicroRNAs. *Cell* **173**, 20–51 (2018).

362. Meister, G. *et al*. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell* **15**, 185–197 (2004).

363. Yekta, S., Shih, I.-H. & Bartel, D. P. MicroRNA-directed cleavage of HOXB8 mRNA. *Science* **304**, 594–596 (2004).

364. Wightman, B., Bürglin, T. R., Gatto, J., Arasu, P. & Ruvkun, G. Negative regulatory sequences in the lin-14 3'-untranslated region are necessary to generate a temporal switch during Caenorhabditis elegans development. *Genes Dev* **5**, 1813–1824 (1991).

365. Moss, E. G., Lee, R. C. & Ambros, V. The cold shock domain protein LIN-28 controls developmental timing in C. elegans and is regulated by the lin-4 RNA. *Cell* **88**, 637–646 (1997).

366. Reinhart, B. J. *et al*. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403**, 901–906 (2000).

367. Lee, R. C., Feinbaum, R. L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843–854 (1993).

368. Pasquinelli, A. E. *et al*. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* **408**, 86–89 (2000).

369. Griffiths-Jones, S., Saini, H. K., van Dongen, S. & Enright, A. J. miRBase: tools for microRNA genomics. *Nucleic Acids Res* **36**, D154-158 (2008).

370. Rodriguez, A., Griffiths-Jones, S., Ashurst, J. L. & Bradley, A. Identification of mammalian microRNA host genes and transcription units. *Genome Res* **14**, 1902–1910 (2004).

371. Nguyen, T. A. *et al*. Functional Anatomy of the Human Microprocessor. *Cell* **161**, 1374–1387 (2015).

372. Lee, Y. *et al*. The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415–419 (2003).

373. Yi, R., Qin, Y., Macara, I. G. & Cullen, B. R. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev* **17**, 3011–3016 (2003).

374. Bohnsack, M. T., Czaplinski, K. & Gorlich, D. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA* **10**, 185–191 (2004).

375. Ha, M. & Kim, V. N. Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol* **15**, 509–524 (2014).

376. Zhang, H., Kolb, F. A., Jaskiewicz, L., Westhof, E. & Filipowicz, W. Single processing center models for human Dicer and bacterial RNase III. *Cell* **118**, 57–68 (2004).

377. Iwasaki, S. *et al*. Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. *Mol Cell* **39**, 292–299 (2010).

378. Frank, F., Sonenberg, N. & Nagar, B. Structural basis for 5’-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature* **465**, 818–822 (2010).

379. Suzuki, H. I. *et al*. Small-RNA asymmetry is directly driven by mammalian Argonautes. *Nat Struct Mol Biol* **22**, 512–521 (2015).

380. Mourelatos, Z. *et al*. miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev* **16**, 720–728 (2002).

381. Tabara, H. *et al*. The rde-1 gene, RNA interference, and transposon silencing in C. elegans. *Cell* **99**, 123–132 (1999).

382. Kawamata, T. & Tomari, Y. Making RISC. *Trends Biochem Sci* **35**, 368–376 (2010).

383. Liu, J. *et al*. Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**, 1437–1441 (2004).

384. Niaz, S. The AGO proteins: an overview. *Biol Chem* **399**, 525–547 (2018).

385. Rivas, F. V. *et al*. Purified Argonaute2 and an siRNA form recombinant human RISC. *Nat Struct Mol Biol* **12**, 340–349 (2005).

386. Lima, W. F. *et al*. Binding and Cleavage Specificities of Human Argonaute2. *Journal of Biological Chemistry* **284**, 26017–26028 (2009).

387. Song, J.-J., Smith, S. K., Hannon, G. J. & Joshua-Tor, L. Crystal Structure of Argonaute and Its Implications for RISC Slicer Activity. *Science* **305**, 1434–1437 (2004).

388. Hammond, S. M., Bernstein, E., Beach, D. & Hannon, G. J. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature* **404**, 293–296 (2000).

389. Agrawal, N. *et al*. RNA Interference: Biology, Mechanism, and Applications. *Microbiology and Molecular Biology Reviews* **67**, 657–685 (2003).

390. Elbashir, S. M. *et al*. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494–498 (2001).

391. Bobbin, M. L. & Rossi, J. J. RNA Interference (RNAi)-Based Therapeutics: Delivering on the Promise? *Annu Rev Pharmacol Toxicol* **56**, 103–122 (2016).

392. Tuschl, T. Expanding small RNA interference. *Nat Biotechnol* **20**, 446–448 (2002).

393. Lim, L. P. *et al*. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769–773 (2005).

394. Behm-Ansmant, I. *et al*. mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes Dev* **20**, 1885–1898 (2006).

395. Iwakawa, H.-O. & Tomari, Y. Life of RISC: Formation, action, and degradation of RNA-induced silencing complex. *Mol Cell* **82**, 30–43 (2022).

396. Fabian, M. R. *et al*. Mammalian miRNA RISC recruits CAF1 and PABP to affect PABP-dependent deadenylation. *Mol Cell* **35**, 868–880 (2009).

397. Jinek, M., Fabian, M. R., Coyle, S. M., Sonenberg, N. & Doudna, J. A. Structural insights into the human GW182-PABC interaction in microRNA-mediated deadenylation. *Nat Struct Mol Biol* **17**, 238–240 (2010).

398. Fabian, M. R. *et al*. miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4-NOT. *Nat Struct Mol Biol* **18**, 1211–1217 (2011).

399. Chen, C.-Y. A. & Shyu, A.-B. Mechanisms of deadenylation-dependent decay. *Wiley Interdiscip Rev RNA* **2**, 167–183 (2011).

400. Jonas, S. & Izaurralde, E. Towards a molecular understanding of microRNA-mediated gene silencing. *Nat Rev Genet* **16**, 421–433 (2015).

401. Rehwinkel, J., Behm-Ansmant, I., Gatfield, D. & Izaurralde, E. A crucial role for GW182 and the DCP1:DCP2 decapping complex in miRNA-mediated gene silencing. *RNA* **11**, 1640–1647 (2005).

402. Duchaine, T. F. & Fabian, M. R. Mechanistic Insights into MicroRNA-Mediated Gene Silencing. *Cold Spring Harb Perspect Biol* **11**, a032771 (2019).

403. Räsch, F., Weber, R., Izaurralde, E. & Igreja, C. 4E-T-bound mRNAs are stored in a silenced and deadenylated form. *Genes Dev* **34**, 847–860 (2020).

404. Kim, D. *et al*. General rules for functional microRNA targeting. *Nat Genet* **48**, 1517–1526 (2016).

405. Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215–233 (2009).

406. Agarwal, V., Bell, G. W., Nam, J.-W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *eLife* **4**, e05005 (2015).

407. Friedman, R. C., Farh, K. K.-H., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19**, 92–105 (2009).

408. Jan, C. H., Friedman, R. C., Ruby, J. G. & Bartel, D. P. Formation, regulation and evolution of Caenorhabditis elegans 3'UTRs. *Nature* **469**, 97–101 (2011).

409. Brennecke, J., Stark, A., Russell, R. B. & Cohen, S. M. Principles of microRNA-target recognition. *PLoS Biol* **3**, e85 (2005).

410. Ecsedi, M., Rausch, M. & Großhans, H. The let-7 microRNA directs vulval development through a single target. *Dev Cell* **32**, 335–344 (2015).

411. Weill, N. *et al*. MiRBooking simulates the stoichiometric mode of action of microRNAs. *Nucleic Acids Res* **43**, 6730–6738 (2015).

412. McGeary, S. E. *et al*. The biochemical basis of microRNA targeting efficacy. *Science* **366**, eaav1741 (2019).

413. Grimson, A. *et al*. MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing. *Molecular Cell* **27**, 91–105 (2007).

414. Wee, L. M., Flores-Jasso, C. F., Salomon, W. E. & Zamore, P. D. Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties. *Cell* **151**, 1055–1067 (2012).

415. Salomon, W. E., Jolly, S. M., Moore, M. J., Zamore, P. D. & Serebrov, V. Single-Molecule Imaging Reveals that Argonaute Reshapes the Binding Properties of Its Nucleic Acid Guides. *Cell* **162**, 84–95 (2015).

416. Moore, M. J. *et al*. miRNA-target chimeras reveal miRNA 3'-end pairing as a major determinant of Argonaute target specificity. *Nat Commun* **6**, 8864 (2015).

417. Broughton, J. P., Lovci, M. T., Huang, J. L., Yeo, G. W. & Pasquinelli, A. E. Pairing Beyond the Seed Supports MicroRNA Targeting Specificity. *Mol Cell* **64**, 320–333 (2016).

418. McGeary, S. E., Bisaria, N., Pham, T. M., Wang, P. Y. & Bartel, D. P. MicroRNA 3'-compensatory pairing occurs through two binding modes, with affinity shaped by nucleotide identity and position. *Elife* **11**, e69803 (2022).

419. Yan, Y. *et al*. The sequence features that define efficient and specific hAGO2-dependent miRNA silencing guides. *Nucleic Acids Res* **46**, 8181–8196 (2018).

420. Duan, Y., Veksler-Lublinsky, I. & Ambros, V. Critical contribution of 3' non-seed base pairing to the in vivo function of the evolutionarily conserved let-7a microRNA. *Cell Rep* **39**, 110745 (2022).

421. Sheu-Gruttadauria, J., Xiao, Y., Gebert, L. F. & MacRae, I. J. Beyond the seed: structural basis for supplementary microRNA targeting by human Argonaute2. *EMBO J* **38**, e101153 (2019).

422. Wightman, B., Ha, I. & Ruvkun, G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* **75**, 855–862 (1993).

423. Bernstein, E. *et al*. Dicer is essential for mouse development. *Nat Genet* **35**, 215–217 (2003).

424. Wang, Y., Medvid, R., Melton, C., Jaenisch, R. & Blelloch, R. DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nat Genet* **39**, 380–385 (2007).

425. Park, C. Y., Choi, Y. S. & McManus, M. T. Analysis of microRNA knockouts in mice. *Hum Mol Genet* **19**, R169-175 (2010).

426. Wienholds, E. *et al*. MicroRNA expression in zebrafish embryonic development. *Science* **309**, 310–311 (2005).

427. Kloosterman, W. P., Wienholds, E., de Bruijn, E., Kauppinen, S. & Plasterk, R. H. A. In situ detection of miRNAs in animal embryos using LNA-modified oligonucleotide probes. *Nat Methods* **3**, 27–29 (2006).

428. Ason, B. *et al*. Differences in vertebrate microRNA expression. *Proceedings of the National Academy of Sciences* **103**, 14385–14389 (2006).

429. Aboobaker, A. A., Tomancak, P., Patel, N., Rubin, G. M. & Lai, E. C. Drosophila microRNAs exhibit diverse spatial expression patterns during embryonic development. *Proc Natl Acad Sci U S A* **102**, 18017–18022 (2005).

430. Hammond, S. M. An overview of microRNAs. *Adv Drug Deliv Rev* **87**, 3–14 (2015).

431. van Rooij, E. *et al*. Control of stress-dependent cardiac growth and gene expression by a microRNA. *Science* **316**, 575–579 (2007).

432. van Rooij, E. *et al*. A family of microRNAs encoded by myosin genes governs myosin expression and muscle performance. *Dev Cell* **17**, 662–673 (2009).

433. Heidersbach, A. *et al*. microRNA-1 regulates sarcomere formation and suppresses smooth muscle gene expression in the mammalian heart. *Elife* **2**, e01323 (2013).

434. Wei, Y. *et al*. Multifaceted roles of miR-1s in repressing the fetal gene program in the heart. *Cell Res* **24**, 278–292 (2014).

435. Shibata, M., Nakao, H., Kiyonari, H., Abe, T. & Aizawa, S. MicroRNA-9 regulates neurogenesis in mouse telencephalon by targeting multiple transcription factors. *J Neurosci* **31**, 3407–3422 (2011).

436. Lu, W. *et al*. The microRNA miR-22 inhibits the histone deacetylase HDAC4 to promote T(H)17 cell-dependent emphysema. *Nat Immunol* **16**, 1185–1194 (2015).

437. Gurha, P. *et al*. Targeted deletion of microRNA-22 promotes stress-induced cardiac dilation and contractile dysfunction. *Circulation* **125**, 2751–2761 (2012).

438. Zhang, L. *et al*. MicroRNA-31 negatively regulates peripherally derived regulatory T-cell generation by repressing retinoic acid-inducible protein 3. *Nat Commun* **6**, 7639 (2015).

439. Moffett, H. F. *et al*. The microRNA miR-31 inhibits CD8+ T cell function in chronic viral infection. *Nat Immunol* **18**, 791–799 (2017).

440. Liu, Z. *et al*. Conditional knockout of microRNA-31 promotes the development of colitis associated cancer. *Biochemical and Biophysical Research Communications* **490**, 62–68 (2017).

441. Mok, Y. *et al*. MiR-210 is induced by Oct-2, regulates B cells, and inhibits autoantibody production. *J Immunol* **191**, 3037–3048 (2013).

442. Wang, H. *et al*. Negative regulation of Hif1a expression and TH17 differentiation by the hypoxia-regulated microRNA miR-210. *Nat Immunol* **15**, 393–401 (2014).

443. White, K. *et al*. Genetic and hypoxic alterations of the microRNA-210-ISCU1/2 axis promote iron-sulfur deficiency and pulmonary hypertension. *EMBO Mol Med* **7**, 695–713 (2015).

444. Weigelt, C. M. *et al*. Loss of miR-210 leads to progressive retinal degeneration in Drosophila melanogaster. *Life Sci Alliance* **2**, e201800149 (2019).

445. Riccardi, S. *et al*. MiR-210 promotes sensory hair cell formation in the organ of corti. *BMC Genomics* **17**, 309 (2016).

446. Bian, X. *et al*. MicroRNA-210 regulates placental adaptation to maternal hypoxic stress during pregnancy†. *Biol Reprod* **104**, 418–429 (2021).

447. Hasuwa, H., Ueda, J., Ikawa, M. & Okabe, M. miR-200b and miR-429 function in mouse ovulation and are essential for female fertility. *Science* **341**, 71–73 (2013).

448. Chen, X. *et al*. Dual regulation of HMGB1 by combined JNK1/2-ATF2 axis with miR-200 family in nonalcoholic steatohepatitis in mice. *FASEB J* **32**, 2722–2734 (2018).

449. Saliminejad, K., Khorram Khorshid, H. R., Soleymani Fard, S. & Ghaffari, S. H. An overview of microRNAs: Biology, functions, therapeutics, and analysis methods. *J Cell Physiol* **234**, 5451–5465 (2019).

450. Vidigal, J. A. & Ventura, A. The biological functions of miRNAs: lessons from in vivo studies. *Trends Cell Biol* **25**, 137–147 (2015).

451. Kulshreshtha, R., Davuluri, R. V., Calin, G. A. & Ivan, M. A microRNA component of the hypoxic response. *Cell Death Differ* **15**, 667–671 (2008).

452. Shen, G., Li, X., Jia, Y., Piazza, G. A. & Xi, Y. Hypoxia-regulated microRNAs in human cancer. *Acta Pharmacol Sin* **34**, 336–341 (2013).

453. Chan, S. Y. *et al*. MicroRNA-210 controls mitochondrial metabolism during hypoxia by repressing the iron-sulfur cluster assembly proteins ISCU1/2. *Cell Metab* **10**, 273–284 (2009).

454. Yang, W. *et al*. Downregulation of miR-210 expression inhibits proliferation, induces apoptosis and enhances radiosensitivity in hypoxic human hepatoma cells in vitro. *Exp Cell Res* **318**, 944–954 (2012).

455. Wang, F. *et al*. miR-210 suppresses BNIP3 to protect against the apoptosis of neural progenitor cells. *Stem Cell Res* **11**, 657–667 (2013).

456. Zhang, Z. *et al*. MicroRNA miR-210 modulates cellular response to hypoxia through the MYC antagonist MNT. *Cell Cycle* **8**, 2756–2768 (2009).

457. Zhao, Q. *et al*. HIF-1α Induces Multidrug Resistance in Gastric Cancer Cells by Inducing MiR-27a. *PLoS One* **10**, e0132746 (2015).

458. Urbich, C. *et al*. MicroRNA-27a/b controls endothelial cell repulsion and angiogenesis by targeting semaphorin 6A. *Blood* **119**, 1607–1616 (2012).

459. Mazar, J. *et al*. MicroRNA 211 Functions as a Metabolic Switch in Human Melanoma Cells. *Mol Cell Biol* **36**, 1090–1108 (2016).

460. Romano, G. & Kwong, L. N. miRNAs, Melanoma and Microenvironment: An Intricate Network. *Int J Mol Sci* **18**, 2354 (2017).

461. Nallamshetty, S., Chan, S. Y. & Loscalzo, J. Hypoxia: a master regulator of microRNA biogenesis and activity. *Free Radic Biol Med* **64**, 20–30 (2013).

462. Fernández-Hernando, C., Ramírez, C. M., Goedeke, L. & Suárez, Y. MicroRNAs in metabolic disease. *Arterioscler Thromb Vasc Biol* **33**, 178–185 (2013).

463. Esau, C. *et al*. miR-122 regulation of lipid metabolism revealed by in vivo antisense targeting. *Cell Metab* **3**, 87–98 (2006).

464. Tsai, W.-C. *et al*. MicroRNA-122 plays a critical role in liver homeostasis and hepatocarcinogenesis. *J Clin Invest* **122**, 2884–2897 (2012).

465. Hsu, S.-H. *et al*. Essential metabolic, anti-inflammatory, and anti-tumorigenic functions of miR-122 in liver. *J Clin Invest* **122**, 2871–2883 (2012).

466. Krützfeldt, J. *et al*. Silencing of microRNAs in vivo with ‘antagomirs’. *Nature* **438**, 685–689 (2005).

467. Guo, J. *et al*. Reduced miR-200b and miR-200c expression contributes to abnormal hepatic lipid accumulation by stimulating JUN expression and activating the transcription of srebp1. *Oncotarget* **7**, 36207–36219 (2016).

468. Dou, L. *et al*. miR-200s contribute to interleukin-6 (IL-6)-induced insulin resistance in hepatocytes. *J Biol Chem* **288**, 22596–22606 (2013).

469. Tran, M., Lee, S.-M., Shin, D.-J. & Wang, L. Loss of miR-141/200c ameliorates hepatic steatosis and inflammation by reprogramming multiple signaling pathways in NASH. *JCI Insight* **2**, 96094 (2017).

470. Calin, G. A. *et al*. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A* **99**, 15524–15529 (2002).

471. Li, Y. & Kowdley, K. V. MicroRNAs in common human diseases. *Genomics Proteomics Bioinformatics* **10**, 246–253 (2012).

472. Cimmino, A. *et al*. miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proc Natl Acad Sci U S A* **102**, 13944–13949 (2005).

473. Gregory, P. A. *et al*. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat Cell Biol* **10**, 593–601 (2008).

474. Christoffersen, N. R., Silahtaroglu, A., Orom, U. A., Kauppinen, S. & Lund, A. H. miR-200b mediates post-transcriptional repression of ZFHX1B. *RNA* **13**, 1172–1178 (2007).

475. Korpal, M., Lee, E. S., Hu, G. & Kang, Y. The miR-200 family inhibits epithelial-mesenchymal transition and cancer cell migration by direct targeting of E-cadherin transcriptional repressors ZEB1 and ZEB2. *J Biol Chem* **283**, 14910–14914 (2008).

476. Burk, U. *et al*. A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. *EMBO Rep* **9**, 582–589 (2008).

477. Wu, S.-G., Chang, T.-H., Liu, Y.-N. & Shih, J.-Y. MicroRNA in Lung Cancer Metastasis. *Cancers (Basel)* **11**, E265 (2019).

478. Howe, E. N., Cochrane, D. R. & Richer, J. K. The miR-200 and miR-221/222 microRNA families: opposing effects on epithelial identity. *J Mammary Gland Biol Neoplasia* **17**, 65–77 (2012).

479. Hui, A. B. Y. *et al*. Robust global micro-RNA profiling with formalin-fixed paraffin-embedded breast cancer tissues. *Lab Invest* **89**, 597–606 (2009).

480. Zheng, C., Yinghao, S. & Li, J. MiR-221 expression affects invasion potential of human prostate carcinoma cell lines by targeting DVL2. *Med Oncol* **29**, 815–822 (2012).

481. Gottardo, F. *et al*. Micro-RNA profiling in kidney and bladder cancers. *Urol Oncol* **25**, 387–392 (2007).

482. Sun, K. *et al*. MicroRNA-221 inhibits CDKN1C/p57 expression in human colorectal carcinoma. *Acta Pharmacol Sin* **32**, 375–384 (2011).

483. Chun-Zhi, Z. *et al*. MicroRNA-221 and microRNA-222 regulate gastric carcinoma cell proliferation and radioresistance by targeting PTEN. *BMC Cancer* **10**, 367 (2010).

484. Felicetti, F. *et al*. The promyelocytic leukemia zinc finger-microRNA-221/-222 pathway controls melanoma progression through multiple oncogenic mechanisms. *Cancer Res* **68**, 2745–2754 (2008).

485. Liang, Y.-K. *et al*. MiR-221/222 promote epithelial-mesenchymal transition by targeting Notch3 in breast cancer cell lines. *NPJ Breast Cancer* **4**, 20 (2018).

486. Guttilla, I. K. *et al*. Prolonged mammosphere culture of MCF-7 cells induces an EMT and repression of the estrogen receptor by microRNAs. *Breast Cancer Res Treat* **132**, 75–85 (2012).

487. Gramantieri, L. *et al*. MicroRNA-221 targets Bmf in hepatocellular carcinoma and correlates with tumor multifocality. *Clin Cancer Res* **15**, 5073–5081 (2009).

488. Cochrane, D. R. *et al*. MicroRNAs link estrogen receptor alpha status and Dicer levels in breast cancer. *Horm Cancer* **1**, 306–319 (2010).

489. Kristen, A. V. *et al*. Patisiran, an RNAi therapeutic for the treatment of hereditary transthyretin-mediated amyloidosis. *Neurodegener Dis Manag* **9**, 5–23 (2019).

490. Weng, Y., Xiao, H., Zhang, J., Liang, X.-J. & Huang, Y. RNAi therapeutic and its innovative biotechnological evolution. *Biotechnol Adv* **37**, 801–825 (2019).

491. Coelho, T. *et al*. A phase II, open-label, extension study of long-term patisiran treatment in patients with hereditary transthyretin-mediated (hATTR) amyloidosis. *Orphanet J Rare Dis* **15**, 179 (2020).

492. Swiecicki, P. L. *et al*. Hereditary ATTR amyloidosis: a single-institution experience with 266 patients. *Amyloid* **22**, 123–131 (2015).

493. Zhao, J. *et al*. MicroRNA-7: a promising new target in cancer therapy. *Cancer Cell Int* **15**, 103 (2015).

494. Gajda, E., Grzanka, M., Godlewska, M. & Gawel, D. The Role of miRNA-7 in the Biology of Cancer and Modulation of Drug Resistance. *Pharmaceuticals (Basel)* **14**, 149 (2021).

495. Pogribny, I. P. *et al*. Alterations of microRNAs and their targets are associated with acquired resistance of MCF-7 breast cancer cells to cisplatin. *Int J Cancer* **127**, 1785–1794 (2010).

496. Hosseinahli, N., Aghapour, M., Duijf, P. H. G. & Baradaran, B. Treating cancer with microRNA replacement therapy: A literature review. *J Cell Physiol* **233**, 5574–5588 (2018).

497. Ebert, M. S. & Sharp, P. A. MicroRNA sponges: progress and possibilities. *RNA* **16**, 2043–2050 (2010).

498. Ebert, M. S., Neilson, J. R. & Sharp, P. A. MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nat Methods* **4**, 721–726 (2007).

499. Franco-Zorrilla, J. M. *et al*. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat Genet* **39**, 1033–1037 (2007).

500. Carè, A. *et al*. MicroRNA-133 controls cardiac hypertrophy. *Nat Med* **13**, 613–618 (2007).

501. Gentner, B. *et al*. Stable knockdown of microRNA in vivo by lentiviral vectors. *Nat Methods* **6**, 63–66 (2009).

502. Memczak, S. *et al*. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**, 333–338 (2013).

503. Zhong, Q., Huang, J., Wei, J. & Wu, R. Circular RNA CDR1as sponges miR-7-5p to enhance E2F3 stability and promote the growth of nasopharyngeal carcinoma. *Cancer Cell Int* **19**, 252 (2019).

504. Zou, Y. *et al*. The Role of Circular RNA CDR1as/ciRS-7 in Regulating Tumor Microenvironment: A Pan-Cancer Analysis. *Biomolecules* **9**, E429 (2019).

505. Zhou, W.-Y. *et al*. Circular RNA: metabolism, functions and interactions with proteins. *Mol Cancer* **19**, 172 (2020).

506. Cheng, C. J. *et al*. MicroRNA silencing for cancer therapy targeted to the tumour microenvironment. *Nature* **518**, 107–110 (2015).

507. Wyatt, L. C. *et al*. Peptides of pHLIP family for targeted intracellular and extracellular delivery of cargo molecules to tumors. *Proc Natl Acad Sci U S A* **115**, E2811–E2818 (2018).

508. Gibson, D. G. Enzymatic assembly of overlapping DNA fragments. *Methods Enzymol* **498**, 349–361 (2011).

509. Gibson, D. G., Smith, H. O., Hutchison, C. A., Venter, J. C. & Merryman, C. Chemical synthesis of the mouse mitochondrial genome. *Nat Methods* **7**, 901–903 (2010).

510. Safran, M. *et al*. Mouse model for noninvasive imaging of HIF prolyl hydroxylase activity: assessment of an oral agent that stimulates erythropoietin production. *Proc Natl Acad Sci U S A* **103**, 105–110 (2006).

511. Porcaro, G., Amosso, E., Scarpella, R. & Carini, F. Doxycycline fluorescence-guided Er:YAG laser ablation combined with Nd:YAG/diode laser biostimulation for treating bisphosphonate-related osteonecrosis of the jaw. *Oral Surg Oral Med Oral Pathol Oral Radiol* **119**, e6–e12 (2015).

512. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online].

513. Fornasini, P. *The Uncertainty in Physical Measurements: An Introduction to Data Analysis in the Physics Laboratory*. (Springer Science & Business Media, 2008).

514. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

515. Shannon, P. *et al*. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498–2504 (2003).

516. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44–57 (2009).

517. Sherman, B. T. *et al*. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res* gkac194 (2022) doi:10.1093/nar/gkac194.

518. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLOS ONE* **6**, e21800 (2011).

519. Robinson, J. T. *et al*. Integrative genomics viewer. *Nat Biotechnol* **29**, 24–26 (2011).

520. Robinson, J. T., Thorvaldsdóttir, H., Wenger, A. M., Zehir, A. & Mesirov, J. P. Variant Review with the Integrative Genomics Viewer. *Cancer Research* **77**, e31–e34 (2017).

521. Garrido-Martín, D., Palumbo, E., Guigó, R. & Breschi, A. ggsashimi: Sashimi plot revised for browser- and annotation-independent splicing visualization. *PLOS Computational Biology* **14**, e1006360 (2018).

522. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).

523. Audemard, E. O. *et al*. Targeted variant detection using unaligned RNA-Seq reads. *Life Science Alliance* **2**, (2019).

524. Salamov, A. A., Nishikawa, T. & Swindells, M. B. Assessing protein coding region integrity in cDNA sequencing projects. *Bioinformatics* **14**, 384–390 (1998).

525. Lacazette, E. A laboratory practical illustrating the use of the ChIP-qPCR method in a robust model: Estrogen receptor alpha immunoprecipitation using Mcf-7 culture cells. *Biochemistry and Molecular Biology Education* **45**, 152–160 (2017).

526. Kann, M. *et al*. Genome-Wide Analysis of Wilms' Tumor 1-Controlled Gene Expression in Podocytes Reveals Key Regulatory Mechanisms. *JASN* **26**, 2097–2104 (2015).

527. He, Q. *et al*. Regulation of HIF-1{alpha} activity in adipose tissue by obesity-associated factors: adipogenesis, insulin, and hypoxia. *Am J Physiol Endocrinol Metab* **300**, E877-885 (2011).

528. Ye, J. *et al*. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**, 134 (2012).

529. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

530. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

531. Li, H. *et al*. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

532. Zhang, Y. *et al*. Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* **9**, R137 (2008).

533. Heinz, S. *et al*. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576–589 (2010).

534. Kurtenbach, S. & Harbour, J. W. SparK: A Publication-quality NGS Visualization Tool. 845529 Preprint at https://doi.org/10.1101/845529 (2019).

535. Semenza, G. L. A compendium of proteins that interact with HIF-1α. *Exp Cell Res* **356**, 128–135 (2017).

536. Keith, B., Johnson, R. S. & Simon, M. C. HIF1α and HIF2α: sibling rivalry in hypoxic tumour growth and progression. *Nat Rev Cancer* **12**, 9–22 (2011).

537. Schlabach, M. R., Hu, J. K., Li, M. & Elledge, S. J. Synthetic design of strong promoters. *Proc Natl Acad Sci U S A* **107**, 2538–2543 (2010).

538. Kim, H.-H. *et al*. Stabilization of hypoxia-inducible factor-1alpha is involved in the hypoxic stimuli-induced expression of vascular endothelial growth factor in osteoblastic cells. *Cytokine* **17**, 14–27 (2002).

539. Bae, S. *et al*. The hypoxia-mimetic agent cobalt chloride induces cell cycle arrest and alters gene expression in U266 multiple myeloma cells. *Int J Mol Med* **30**, 1180–1186 (2012).

540. Liu, Y., Cox, S. R., Morita, T. & Kourembanas, S. Hypoxia Regulates Vascular Endothelial Growth Factor Gene Expression in Endothelial Cells. *Circulation Research* **77**, 638–643 (1995).

541. Berra, E., Roux, D., Richard, D. E. & Pouysségur, J. Hypoxia-inducible factor-1α (HIF-1α) escapes O2-driven proteasomal degradation irrespective of its subcellular localization: nucleus or cytoplasm. *EMBO Rep* **2**, 615–620 (2001).

542. Bagnall, J. *et al*. Tight Control of Hypoxia-inducible Factor-α Transient Dynamics Is Essential for Cell Survival in Hypoxia. *Journal of Biological Chemistry* **289**, (2014).

543. Moroz, E. *et al*. Real-Time Imaging of HIF-1α Stabilization and Degradation. *PLOS ONE* **4**, e5077 (2009).

544. Khmelinskii, A. *et al*. Incomplete proteasomal degradation of green fluorescent proteins in the context of tandem fluorescent protein timers. *MBoC* **27**, 360–370 (2016).

545. Ma, L., Yan, H., Zhao, H. & Sun, J. Grainyhead-like 2 in development and cancer. *Tumour Biol*. **39**, 1010428317698375 (2017).

546. Shih, J.-Y. & Yang, P.-C. The EMT regulator slug and lung carcinogenesis. *Carcinogenesis* **32**, 1299–1304 (2011).

547. Mingyuan, X. *et al*. Hypoxia-inducible factor-1α activates transforming growth factor-β1/Smad signaling and increases collagen deposition in dermal fibroblasts. *Oncotarget* **9**, 3188–3197 (2017).

548. Kim, B.-G., Malek, E., Choi, S. H., Ignatz-Hoover, J. J. & Driscoll, J. J. Novel therapies emerging in oncology to target the TGF-β pathway. *Journal of Hematology & Oncology* **14**, 55 (2021).

549. Denko, N. C. Hypoxia, HIF1 and glucose metabolism in the solid tumour. *Nat Rev Cancer* **8**, 705–713 (2008).

550. Pelletier, J. *et al*. Glycogen Synthesis is Induced in Hypoxia by the Hypoxia-Inducible Factor and Promotes Cancer Cell Survival. *Frontiers in Oncology* **2**, (2012).

551. Izraely, S. *et al*. The melanoma brain metastatic microenvironment: aldolase C partakes in shaping the malignant phenotype of melanoma cells – a case of inter-tumor heterogeneity. *Molecular Oncology* **15**, 1376–1390 (2021).

552. Mark, H. F. Biochemistry, 2nd ed., Lubert Stryer, Freeman, San Francisco, 1981, 949 pp. *Journal of Polymer Science: Polymer Letters Edition* **19**, 150–151 (1981).

553. Garel, M. C. *et al*. Human bisphosphoglycerate mutase. Expression in Escherichia coli and use of site-directed mutagenesis in the evaluation of the role of the carboxyl-terminal region in the enzymatic mechanism. *J Biol Chem* **264**, 18966–18972 (1989).

554. Cho, J., King, J. S., Qian, X., Harwood, A. J. & Shears, S. B. Dephosphorylation of 2,3-bisphosphoglycerate by MIPP expands the regulatory capacity of the Rapoport–Luebering glycolytic shunt. *Proceedings of the National Academy of Sciences* **105**, 5998–6003 (2008).

555. Ellingson, W. J., Chesser, D. G. & Winder, W. W. Effects of 3-phosphoglycerate and other metabolites on the activation of AMP-activated protein kinase by LKB1-STRAD-MO25. *Am J Physiol Endocrinol Metab* **292**, E400-407 (2007).

556. Lieu, E. L., Nguyen, T., Rhyne, S. & Kim, J. Amino acids in cancer. *Exp Mol Med* **52**, 15–30 (2020).

557. van Geldermalsen, M. *et al*. ASCT2/SLC1A5 controls glutamine uptake and tumour growth in triple-negative basal-like breast cancer. *Oncogene* **35**, 3201–3208 (2016).

558. Kandasamy, P., Gyimesi, G., Kanai, Y. & Hediger, M. A. Amino acid transporters revisited: New views in health and disease. *Trends Biochem Sci* **43**, 752–789 (2018).

559. Yanagida, O. *et al*. Human L-type amino acid transporter 1 (LAT1): characterization of function and expression in tumor cell lines. *Biochim Biophys Acta* **1514**, 291–302 (2001).

560. Torrents, D. *et al*. Identification and characterization of a membrane protein (y+L amino acid transporter-1) that associates with 4F2hc to encode the amino acid transport activity y+L. A candidate gene for lysinuric protein intolerance. *J Biol Chem* **273**, 32437–32445 (1998).

561. Peng, H., Wang, Y. & Luo, W. Multifaceted role of branched-chain amino acid metabolism in cancer. *Oncogene* **39**, 6747–6756 (2020).

562. Zhang, B. *et al*. Regulation of branched-chain amino acid metabolism by hypoxia-inducible factor in glioblastoma. *Cell Mol Life Sci* **78**, 195–206 (2021).

563. Guo, W., Wang, H. & Li, C. Signal pathways of melanoma and targeted therapy. *Sig Transduct Target Ther* **6**, 1–39 (2021).

564. Zhang, B. *et al*. BCAT1 knockdown-mediated suppression of melanoma cell proliferation and migration is associated with reduced oxidative phosphorylation. *Am J Cancer Res* **11**, 2670–2683 (2021).

565. Liu, Y. *et al*. Branched-Chain Amino Acid Negatively Regulates KLF15 Expression via PI3K-AKT Pathway. *Front Physiol* **8**, 853 (2017).

566. Gray, S. *et al*. Regulation of Gluconeogenesis by Krüppel-like Factor 15. *Cell Metabolism* **5**, 305–312 (2007).

567. DeBerardinis, R. J. *et al*. Beyond aerobic glycolysis: Transformed cells can engage in glutamine metabolism that exceeds the requirement for protein and nucleotide synthesis. *Proceedings of the National Academy of Sciences* **104**, 19345–19350 (2007).

568. Morotti, M. *et al*. Hypoxia-induced switch in SNAT2/SLC38A2 regulation generates endocrine resistance in breast cancer. *Proceedings of the National Academy of Sciences* **116**, 12452–12461 (2019).

569. Yoo, H. C., Yu, Y. C., Sung, Y. & Han, J. M. Glutamine reliance in cell metabolism. *Exp Mol Med* **52**, 1496–1516 (2020).

570. Yoo, H. C. *et al*. A Variant of SLC1A5 Is a Mitochondrial Glutamine Transporter for Metabolic Reprogramming in Cancer Cells. *Cell Metabolism* **31**, 267-283.e12 (2020).

571. Lim, J. K. M. *et al*. Cystine/glutamate antiporter xCT (SLC7A11) facilitates oncogenic RAS transformation by preserving intracellular redox balance. *Proceedings of the National Academy of Sciences* **116**, 9433–9442 (2019).

572. Katt, W. P., Lukey, M. J. & Cerione, R. A. A tale of two glutaminases: homologous enzymes with distinct roles in tumorigenesis. *Future Med Chem* **9**, 223–243 (2017).

573. Scalise, M., Pochini, L., Galluccio, M., Console, L. & Indiveri, C. Glutamine Transport and Mitochondrial Metabolism in Cancer Cell Growth. *Front Oncol* **7**, 306 (2017).

574. Fiermonte, G. *et al*. Identification of the mitochondrial glutamate transporter. Bacterial expression, reconstitution, functional characterization, and tissue distribution of two human isoforms. *J Biol Chem* **277**, 19289–19294 (2002).

575. Molinari, F. *et al*. Impaired mitochondrial glutamate transport in autosomal recessive neonatal myoclonic epilepsy. *Am J Hum Genet* **76**, 334–339 (2005).

576. Mucaj, V., Shay, J. E. S. & Simon, M. C. Effects of hypoxia and HIFs on cancer metabolism. *Int J Hematol* **95**, 464–470 (2012).

577. Bender, T. & Martinou, J.-C. The mitochondrial pyruvate carrier in health and disease: To carry or not to carry? *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1863**, 2436–2442 (2016).

578. Taylor, E. B. Functional Properties of the Mitochondrial Carrier System. *Trends Cell Biol* **27**, 633–644 (2017).

579. Bender, T., Pena, G. & Martinou, J.-C. Regulation of mitochondrial pyruvate uptake by alternative pyruvate carrier complexes. *EMBO J* **34**, 911–924 (2015).

580. Compan, V. *et al*. Monitoring Mitochondrial Pyruvate Carrier Activity in Real Time Using a BRET-Based Biosensor: Investigation of the Warburg Effect. *Molecular Cell* **59**, 491–501 (2015).

581. Liang, Y.-J. & Jiang, J.-G. Characterization of malic enzyme and the regulation of its activity and metabolic engineering on lipid production. *RSC Adv.* **5**, 45558–45570 (2015).

582. Gray, L. R., Tompkins, S. C. & Taylor, E. B. Regulation of pyruvate metabolism and human disease. *Cell Mol Life Sci* **71**, 2577–2604 (2014).

583. Ju, H.-Q., Lin, J.-F., Tian, T., Xie, D. & Xu, R.-H. NADPH homeostasis in cancer: functions, mechanisms and therapeutic implications. *Sig Transduct Target Ther* **5**, 1–12 (2020).

584. Formenti, F. *et al*. Regulation of human metabolism by hypoxia-inducible factor. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 12722–12727 (2010).

585. Lu, C.-W., Lin, S.-C., Chen, K.-F., Lai, Y.-Y. & Tsai, S.-J. Induction of Pyruvate Dehydrogenase Kinase-3 by Hypoxia-inducible Factor-1 Promotes Metabolic Switch and Drug Resistance. *J Biol Chem* **283**, 28106–28114 (2008).

586. Kiesel, V. A. *et al*. Pyruvate carboxylase and cancer progression. *Cancer Metab* **9**, 20 (2021).

587. Yang, M., Soga, T., Pollard, P. & Adam, J. The emerging role of fumarate as an oncometabolite. *Frontiers in Oncology* **2**, (2012).

588. Metallo, C. M. *et al*. Reductive glutamine metabolism by IDH1 mediates lipogenesis under hypoxia. *Nature* **481**, 380–384 (2011).

589. Arnold, P. K. *et al*. A non-canonical tricarboxylic acid cycle underlies cellular identity. *Nature* **603**, 477–481 (2022).

590. Al-Maskari, M. *et al*. Site-1 protease function is essential for the generation of antibody secreting cells and reprogramming for secretory activity. *Sci Rep* **8**, 14338 (2018).

591. Shimano, H. & Sato, R. SREBP-regulated lipid metabolism: convergent physiology — divergent pathophysiology. *Nat Rev Endocrinol* **13**, 710–730 (2017).

592. Hwang, S. *et al*. Hypoxia-inducible factor 1α activates insulin-induced gene 2 (Insig-2) transcription for degradation of 3-hydroxy-3-methylglutaryl (HMG)-CoA reductase in the liver. *J Biol Chem* **292**, 9382–9393 (2017).

593. DeBose-Boyd, R. A. Feedback regulation of cholesterol synthesis: sterol-accelerated ubiquitination and degradation of HMG CoA reductase. *Cell Res* **18**, 609–621 (2008).

594. Yakubovskaya, V. I. The effect of fasting on the cholesterol metabolism. *Bull Exp Biol Med* **50**, 835–836 (1961).

595. Kondo, A. *et al*. Extracellular Acidic pH Activates the Sterol Regulatory Element-Binding Protein 2 to Promote Tumor Progression. *Cell Reports* **18**, 2228–2242 (2017).

596. Kambe, T., Tsuji, T., Hashimoto, A. & Itsumura, N. The Physiological, Biochemical, and Molecular Roles of Zinc Transporters in Zinc Homeostasis and Metabolism. *Physiol Rev* **95**, 749–784 (2015).

597. Kambe, T. Molecular Architecture and Function of ZnT Transporters. in *Current Topics in Membranes* vol. 69 199–220 (Elsevier, 2012).

598. Nishito, Y. & Kambe, T. Zinc transporter 1 (ZNT1) expression on the cell surface is elaborately controlled by cellular zinc levels. *J Biol Chem* **294**, 15686–15697 (2019).

599. Kagara, N., Tanaka, N., Noguchi, S. & Hirano, T. Zinc and its transporter ZIP10 are involved in invasive behavior of breast cancer cells. *Cancer Science* **98**, 692–697 (2007).

600. Jeong, J. *et al.* Promotion of vesicular zinc efflux by ZIP13 and its implications for spondylocheiro dysplastic Ehlers-Danlos syndrome. *Proc Natl Acad Sci U S A* **109**, E3530-3538 (2012).

601. Cheng, X. *et al.* Zinc transporter SLC39A13/ZIP13 facilitates the metastasis of human ovarian cancer cells via activating Src/FAK signaling pathway. *J Exp Clin Cancer Res* **40**, 199 (2021).

602. Bandara, K. V., Michael, M. Z. & Gleadle, J. M. MicroRNA Biogenesis in Hypoxia. *Microrna* **6**, 80–96 (2017).

603. Winter, J., Jung, S., Keller, S., Gregory, R. I. & Diederichs, S. Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat Cell Biol* **11**, 228–234 (2009).

604. Okuda, H. *et al.* miR-7 Suppresses Brain Metastasis of Breast Cancer Stem-Like Cells By Modulating KLF4. *Cancer Research* **73**, 1434–1444 (2013).

605. Zhang, H. *et al.* MiR-7, Inhibited Indirectly by LincRNA HOTAIR, Directly Inhibits SETDB1 and Reverses the EMT of Breast Cancer Stem Cells by Downregulating the STAT3 Pathway. *Stem Cells* **32**, 2858–2868 (2014).

606. Jin, H.-F. *et al.* Down-Regulation of miR-7 in Gastric Cancer Is Associated With Elevated LDH-A Expression and Chemoresistance to Cisplatin. *Front Cell Dev Biol* **8**, 555937 (2020).

607. Seong, M., Lee, J. & Kang, H. Hypoxia-induced regulation of mTOR signaling by miR-7 targeting REDD1. *J Cell Biochem* **120**, 4523–4532 (2019).

608. Li, M., Pan, M., You, C. & Dou, J. The Therapeutic Potential of miR-7 in Cancers. *Mini Rev Med Chem* **19**, 1707–1716 (2019).

609. Climent, M., Viggiani, G., Chen, Y.-W., Coulis, G. & Castaldi, A. MicroRNA and ROS Crosstalk in Cardiac and Pulmonary Diseases. *Int J Mol Sci* **21**, E4370 (2020).

610. Todeschini, P. *et al*. Comprehensive Profiling of Hypoxia-Related miRNAs Identifies miR-23a-3p Overexpression as a Marker of Platinum Resistance and Poor Prognosis in High-Grade Serous Ovarian Cancer. *Cancers (Basel)* **13**, 3358 (2021).

611. A, T., M, N., M, L., F, H.-V. & F, S.-O. Hypoxic Melanoma Cells Deliver microRNAs to Dendritic Cells and Cytotoxic T Lymphocytes through Connexin-43 Channels. *International journal of molecular sciences* **21**, (2020).

612. Nadiminty, N. *et al*. MicroRNA let-7c is downregulated in prostate cancer and suppresses prostate cancer growth. *PLoS One* **7**, e32832 (2012).

613. Peng, C.-Y. *et al*. Let-7c restores radiosensitivity and chemosensitivity and impairs stemness in oral cancer cells through inhibiting interleukin-8. *J Oral Pathol Med* **47**, 590–597 (2018).

614. Yao, Y. *et al*. [miR-let-7c-5p inhibits invasion and migration of bladder cancer cells by targeting HMGA2]. *Nan Fang Yi Ke Da Xue Xue Bao* **41**, 1022–1029 (2021).

615. Manaster, I. *et al*. MiRNA-mediated control of HLA-G expression and function. *PLoS One* **7**, e33395 (2012).

616. Song, B. *et al*. Long non-coding RNA HOTAIR promotes HLA-G expression via inhibiting miR-152 in gastric cancer cells. *Biochem Biophys Res Commun* **464**, 807–813 (2015).

617. Ma, J. *et al*. MiR-152 functions as a tumor suppressor in glioblastoma stem cells by targeting Krüppel-like factor 4. *Cancer Lett* **355**, 85–95 (2014).

618. Sahranavardfard, P. *et al*. MicroRNA-203 reinforces stemness properties in melanoma and augments tumorigenesis in vivo. *J Cell Physiol* **234**, 20193–20205 (2019).

619. Yang, X., Chen, Y. & Chen, L. The Versatile Role of microRNA-30a in Human Cancer. *Cell Physiol Biochem* **41**, 1616–1632 (2017).

620. Li, Y. *et al*. MiR-30a-5p confers cisplatin resistance by regulating IGF1R expression in melanoma cells. *BMC Cancer* **18**, 404 (2018).

621. Rissland, O. S., Hong, S.-J. & Bartel, D. P. MicroRNA Destabilization Enables Dynamic Regulation of the miR-16 Family in Response to Cell-Cycle Changes. *Molecular Cell* **43**, 993–1004 (2011).

622. Wang, Q. *et al*. microRNA-322/424 promotes liver fibrosis by regulating angiogenesis through targeting CUL2/HIF-1α pathway. *Life Sci* **266**, 118819 (2021).

623. Wang, F. *et al*. H19X-encoded miR-424(322)/-503 cluster: emerging roles in cell differentiation, proliferation, plasticity and metabolism. *Cell Mol Life Sci* **76**, 903–920 (2019).

624. Drasin, D. J. *et al*. TWIST1-Induced miR-424 Reversibly Drives Mesenchymal Programming while Inhibiting Tumor Initiation. *Cancer Research* **75**, 1908–1921 (2015).

625. Li, D., Liu, K., Li, Z., Wang, J. & Wang, X. miR-19a and miR-424 target TGFBR3 to promote epithelial-to-mesenchymal transition and migration of tongue squamous cell carcinoma cells. *Cell Adhesion & Migration* **12**, 236–246 (2018).

626. Rojas, F. *et al*. The Oncogenic Response to MiR-335 Is Associated with Cell Surface Expression of Membrane-Type 1 Matrix Metalloproteinase (MT1-MMP) Activity. *PLoS One* **10**, e0132026 (2015).

627. Mazar, J. *et al*. The Regulation of miRNA-211 Expression and Its Role in Melanoma Cell Invasiveness. *PLOS ONE* **5**, e13779 (2010).

628. Laurila, E. M. & Kallioniemi, A. The diverse role of miR-31 in regulating cancer associated phenotypes. *Genes, Chromosomes and Cancer* **52**, 1103–1113 (2013).

629. Nakagawa, Y. *et al*. Induced miR-31 by 5-fluorouracil exposure contributes to the resistance in colorectal tumors. *Cancer Sci* **110**, 2540–2548 (2019).

630. Peng, H. *et al*. MiR-31-5p promotes the cell growth, migration and invasion of colorectal cancer cells by targeting NUMB. *Biomed Pharmacother* **109**, 208–216 (2019).

631. Kadonosono, T. *et al*. Detection of the onset of ischemia and carcinogenesis by hypoxia-inducible transcription factor-based in vivo bioluminescence imaging. *PLoS One* **6**, e26640 (2011).

632. Lehmann, S. *et al*. Longitudinal and multimodal in vivo imaging of tumor hypoxia and its downstream molecular events. *Proceedings of the National Academy of Sciences* **106**, 14004–14009 (2009).

633. Georgakopoulos-Soares, I., Chartoumpekis, D. V., Kyriazopoulou, V. & Zaravinos, A. EMT Factors and Metabolic Pathways in Cancer. *Frontiers in Oncology* **10**, (2020).

634. Wu, M. *et al*. Role of exosomal microRNA-125b-5p in conferring the metastatic phenotype among pancreatic cancer cells with different potential of metastasis. *Life Sci* **255**, 117857 (2020).

635. Mazar, J. *et al*. Epigenetic regulation of microRNA genes and the role of miR-34b in cell invasion and motility in human melanoma. *PLoS One* **6**, e24922 (2011).

636. Zhuang, X.-F. *et al*. miR-34b inhibits the migration/invasion and promotes apoptosis of non-small-cell lung cancer cells by YAF2. *Eur Rev Med Pharmacol Sci* **23**, 2038–2046 (2019).

637. Shi, D., Xing, A., Gao, C. & Gao, P. [Expression of microRNA-100 in human gastric cancer]. *Zhonghua Bing Li Xue Za Zhi* **42**, 15–19 (2013).

638. Chen, D. *et al*. miR-100 induces epithelial-mesenchymal transition but suppresses tumorigenesis, migration and invasion. *PLoS Genet* **10**, e1004177 (2014).

639. Yang, S. *et al*. KITLG is a novel target of miR-34c that is associated with the inhibition of growth and invasion in colorectal cancer cells. *J Cell Mol Med* **18**, 2092–2102 (2014).

640. Song, J. *et al*. miR-34c-3p inhibits cell proliferation, migration and invasion of hepatocellular carcinoma by targeting MARCKS. *Int J Clin Exp Pathol* **8**, 12728–12737 (2015).

641. Wang, Y., Wang, X., Tang, J., Su, X. & Miao, Y. The study of mechanism of miR-34c-5p targeting FLOT2 to regulate proliferation, migration and invasion of osteosarcoma cells. *Artif Cells Nanomed Biotechnol* **47**, 3559–3568 (2019).

642. Xu, Z. *et al*. NOTCH1 regulates the proliferation and migration of bladder cancer cells by cooperating with long non-coding RNA HCG18 and microRNA-34c-5p. *J Cell Biochem* **120**, 6596–6604 (2019).

643. Wei, H. *et al*. miR-34c-5p targets Notch1 and suppresses the metastasis and invasion of cervical cancer. *Mol Med Rep* **23**, 120 (2021).

644. Li, C. *et al*. miRNA-199a-3p in plasma as a potential diagnostic biomarker for gastric cancer. *Ann Surg Oncol* **20 Suppl 3**, S397-405 (2013).

645. Nonaka, R. *et al*. Circulating miR-199a-3p as a novel serum biomarker for colorectal cancer. *Oncol Rep* **32**, 2354–2358 (2014).

646. Zhao, X. *et al*. SRF expedites metastasis and modulates the epithelial to mesenchymal transition by regulating miR-199a-5p expression in human gastric cancer. *Cell Death Differ* **21**, 1900–1913 (2014).

647. Wang, Y. *et al*. miR-199a-5p inhibits proliferation and induces apoptosis in hemangioma cells through targeting HIF1A. *Int J Immunopathol Pharmacol* **31**, 394632017749357 (2018).

648. Keremu, A. *et al*. Overcoming cisplatin resistance in osteosarcoma through the miR-199a-modulated inhibition of HIF-1α. *Biosci Rep* **39**, BSR20170080 (2019).

649. Qu, D., Yang, Y. & Huang, X. miR-199a-5p promotes proliferation and metastasis and epithelial-mesenchymal transition through targeting PIAS3 in cervical carcinoma. *J Cell Biochem* **120**, 13562–13572 (2019).

650. Wei, D. *et al*. MicroRNA-199a-5p suppresses migration and invasion in oral squamous cell carcinoma through inhibiting the EMT-related transcription factor SOX4. *Int J Mol Med* **44**, 185–195 (2019).

651. Li, J. *et al*. miR-221 Promotes Epithelial-Mesenchymal Transition through Targeting PTEN and Forms a Positive Feedback Loop with β-catenin/c-Jun Signaling Pathway in Extra-Hepatic Cholangiocarcinoma. *PLoS One* **10**, e0141168 (2015).

652. Felicetti, F. *et al*. Role of PLZF in melanoma progression. *Oncogene* **23**, 4567–4576 (2004).

653. Yang, L., Fan, Y., Zhang, X., Gao, L. & Ma, J. Role of miRNA-21/PTEN on the high glucose-induced EMT in human mesothelial peritoneal cells. *Am J Transl Res* **10**, 2590–2599 (2018).

654. Arisan, E. D. *et al*. MiR-21 Is Required for the Epithelial–Mesenchymal Transition in MDA-MB-231 Breast Cancer Cells. *Int J Mol Sci* **22**, 1557 (2021).

655. Kong, X., Liu, F. & Gao, J. MiR-155 promotes epithelial-mesenchymal transition in hepatocellular carcinoma cells through the activation of PI3K/SGK3/β-catenin signaling pathways. *Oncotarget* **7**, 66051 (2016).

656. Liu, X., Li, Y., Li, Z. & Hou, T. miR-155 promotes proliferation and epithelial-mesenchymal transition of MCF-7 cells. *Exp Ther Med* **21**, 218 (2021).

657. Avalle, L. *et al*. MicroRNAs-143 and -145 induce epithelial to mesenchymal transition and modulate the expression of junction proteins. *Cell Death Differ* **24**, 1750–1760 (2017).

658. An, N.-N., Shawn, J., Peng, J.-P., Wu, M.-D. & Huang, L.-G. Up-regulation of miR-190b promoted growth, invasion, migration and inhibited apoptosis of Wilms' tumor cells by repressing the PTEN expression. *Eur Rev Med Pharmacol Sci* **22**, 961–969 (2018).

659. Zhao, Q. *et al*. miR-190b promotes colorectal cancer progression through targeting forkhead box protein P2. *Exp Ther Med* **19**, 79–84 (2020).

660. Dou, R. *et al*. EMT-cancer cells-derived exosomal miR-27b-3p promotes circulating tumour cells-mediated metastasis by modulating vascular permeability in colorectal cancer. *Clin Transl Med* **11**, e595 (2021).

661. Liu, W. *et al*. miR-27a promotes proliferation, migration, and invasion of colorectal cancer by targeting FAM172A and acts as a diagnostic and prognostic biomarker. *Oncol Rep* **37**, 3554–3564 (2017).

662. Zhang, Z., Liu, S., Shi, R. & Zhao, G. miR-27 promotes human gastric cancer cell metastasis by inducing epithelial-to-mesenchymal transition. *Cancer Genet* **204**, 486–491 (2011).

663. Jiang, G., Shi, W., Fang, H. & Zhang, X. miR-27a promotes human breast cancer cell migration by inducing EMT in a FBXW7-dependent manner. *Mol Med Rep* **18**, 5417–5426 (2018).

664. Diao, H., Xu, X., Zhao, B. & Yang, G. miR-135a-5p inhibits tumor invasion by targeting ANGPT2 in gallbladder cancer. *Mol Med Rep* **24**, 528 (2021).

665. Duan, S. *et al*. MicroRNA-135a-3p is downregulated and serves as a tumour suppressor in ovarian cancer by targeting CCR2. *Biomed Pharmacother* **107**, 712–720 (2018).

666. Wang, J. *et al*. Effects of miR-135a-5p and miR-141 on proliferation, invasion and apoptosis of colorectal cancer SW620 cells. *Oncol Lett* **20**, 914–920 (2020).

667. Cai, L. *et al*. MicroRNA miR-330-3p suppresses the progression of ovarian cancer by targeting RIPK4. *Bioengineered* **12**, 440–449 (2021).

668. Mohammadi, A. *et al*. Restoration of miR-330 expression suppresses lung cancer cell viability, proliferation, and migration. *J Cell Physiol* **236**, 273–283 (2021).

669. Jin, Z., Jia, B., Tan, L. & Liu, Y. miR-330-3p suppresses liver cancer cell migration by targeting MAP2K1. *Oncol Lett* **18**, 314–320 (2019).

670. Mansoori, B. *et al*. miR-330 suppresses EMT and induces apoptosis by downregulating HMGA2 in human colorectal cancer. *J Cell Physiol* **235**, 920–931 (2020).

671. Park, D. H. *et al*. MicroRNA-146a inhibits epithelial mesenchymal transition in non-small cell lung cancer by targeting insulin receptor substrate 2. *Int J Oncol* **47**, 1545–1553 (2015).

672. Afshar-Khamseh, R., Javeri, A. & Taha, M. F. MiR-146a suppresses the expression of CXCR4 and alters survival, proliferation and migration rate in colorectal cancer cells. *Tissue Cell* **73**, 101654 (2021).

673. Qi, Y., Hou, Y. & Qi, L. miR-30d-5p represses the proliferation, migration, and invasion of lung squamous cell carcinoma via targeting DBF4. *J Environ Sci Health C Toxicol Carcinog* **39**, 251–268 (2021).

674. Han, M. *et al*. microRNA-30d mediated breast cancer invasion, migration, and EMT by targeting KLF11 and activating STAT3 pathway. *J Cell Biochem* **119**, 8138–8145 (2018).

675. Cui, K. & Bian, X. The microRNA cluster miR-30b/-30d prevents tumor cell switch from an epithelial to a mesenchymal-like phenotype in GBC. *Mol Ther Methods Clin Dev* **20**, 716–725 (2021).

676. Ye, Z., Zhao, L., Li, J., Chen, W. & Li, X. miR-30d Blocked Transforming Growth Factor β1-Induced Epithelial-Mesenchymal Transition by Targeting Snail in Ovarian Cancer Cells. *Int J Gynecol Cancer* **25**, 1574–1581 (2015).

677. Ma, Y.-X., Zhang, H., Li, X.-H. & Liu, Y.-H. MiR-30e-5p inhibits proliferation and metastasis of nasopharyngeal carcinoma cells by target-ing USP22. *Eur Rev Med Pharmacol Sci* **22**, 6342–6349 (2018).

678. Thorndike, R. L. Who belongs in the family? *Psychometrika* **18**, 267–276 (1953).

679. Ketchen, D. J. & Shook, C. L. The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique. *Strategic Management Journal* **17**, 441–458 (1996).

680. Ma, C. *et al*. Rce1 suppresses invasion and metastasis of hepatocellular carcinoma via epithelial-mesenchymal transition induced by the TGF-β1/H-Ras signaling pathway. *J Cell Physiol* **235**, 2506–2520 (2020).

681. Siljamäki, E. *et al*. H-Ras activation and fibroblast-induced TGF-β signaling promote laminin-332 accumulation and invasion in cutaneous squamous cell carcinoma. *Matrix Biol* **87**, 26–47 (2020).

682. Nagaoka, K. *et al*. CPEB1 mediates epithelial-to-mesenchyme transition and breast cancer metastasis. *Oncogene* **35**, 2893–2901 (2016).

683. Chen, Z. & Liu, B. Sphk1 participates in malignant progression of breast cancer by regulating epithelial-mesenchymal transition and stem cell characteristics. *Tissue Cell* **65**, 101380 (2020).

684. Ohnishi, Y., Inoue, H., Furukawa, M., Kakudo, K. & Nozaki, M. Heparin-binding epidermal growth factor-like growth factor is a potent regulator of invasion activity in oral squamous cell carcinoma. *Oncol Rep* **27**, 954–958 (2012).

685. Jayachandran, A. *et al*. Thrombospondin 1 promotes an aggressive phenotype through epithelial-to-mesenchymal transition in human melanoma. *Oncotarget* **5**, 5782–5797 (2014).

686. Hou, C.-H., Lin, F.-L., Hou, S.-M. & Liu, J.-F. Cyr61 promotes epithelial-mesenchymal transition and tumor metastasis of osteosarcoma by Raf-1/MEK/ERK/Elk-1/TWIST-1 signaling pathway. *Mol Cancer* **13**, 236 (2014).

687. Sugiyama, H. WT1 (Wilms' tumor gene 1): biology and cancer immunotherapy. *Jpn J Clin Oncol* **40**, 377–387 (2010).

688. Wagner, K.-D., Wagner, N. & Schedl, A. The complex life of WT1. *J Cell Sci* **116**, 1653–1658 (2003).

689. Wang, G. L. & Semenza, G. L. Characterization of hypoxia-inducible factor 1 and regulation of DNA binding activity by hypoxia. *J Biol Chem* **268**, 21513–21518 (1993).

690. Depping, R., Schindler, S. G., Jacobi, C., Kirschner, K. M. & Scholz, H. Nuclear transport of Wilms' tumour protein Wt1 involves importins α and β. *Cell Physiol Biochem* **29**, 223–232 (2012).

691. Wingfield, P. N-Terminal Methionine Processing. *Current protocols in protein science* **88**, 6.14.1 (2017).

692. Drazic, A., Myklebust, L. M., Ree, R. & Arnesen, T. The world of protein acetylation. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1864**, 1372–1401 (2016).

693. Chen, M. & Cook, K. D. Oxidation Artifacts in the Electrospray Mass Spectrometry of Aβ Peptide. *Anal Chem* **79**, 2031–2036 (2007).

694. Hossain, A., Nixon, M., Kuo, M. T. & Saunders, G. F. N-terminally truncated WT1 protein with oncogenic properties overexpressed in leukemia. *J Biol Chem* **281**, 28122–28130 (2006).

695. Ullmark, T., Montano, G. & Gullberg, U. DNA and RNA binding by the Wilms' tumour gene 1 (WT1) protein +KTS and −KTS isoforms-From initial observations to recent global genomic analyses. *Eur J Haematol* **100**, 229–240 (2018).

696. Zapata-Benavides, P. *et al*. WT1 silencing by RNAi synergizes with chemotherapeutic agents and induces chemosensitization to doxorubicin and cisplatin in B16F10 murine melanoma cells. *Oncol Lett* **3**, 751–755 (2012).

697. Dumur, C. I. *et al*. Analytical validation of a real-time reverse transcription-polymerase chain reaction quantitation of different transcripts of the Wilms' tumor suppressor gene (WT1). *Anal Biochem* **309**, 127–136 (2002).

698. Simon, C. *et al*. A key role for EZH2 and associated genes in mouse and human adult T-cell acute leukemia. *Genes Dev* **26**, 651–656 (2012).

699. Maslak, P. G. *et al*. Phase 2 trial of a multivalent WT1 peptide vaccine (galinpepimut-S) in acute myeloid leukemia. *Blood Adv* **2**, 224–234 (2018).

700. Corish, P. & Tyler-Smith, C. Attenuation of green fluorescent protein half-life in mammalian cells. *Protein Engineering, Design and Selection* **12**, 1035–1040 (1999).

701. Yoshinari, K., Miyagishi, M. & Taira, K. Effects on RNAi of the tight structure, sequence and position of the targeted region. *Nucleic Acids Res* **32**, 691–699 (2004).

702. Bronkhorst, A. W., van Cleef, K. W. R., Venselaar, H. & van Rij, R. P. A dsRNA-binding protein of a complex invertebrate DNA virus suppresses the Drosophila RNAi response. *Nucleic Acids Res* **42**, 12237–12248 (2014).

703. Schwarz, D. S. *et al*. Asymmetry in the Assembly of the RNAi Enzyme Complex. *Cell* **115**, 199–208 (2003).

704. Khvorova, A., Reynolds, A. & Jayasena, S. D. Functional siRNAs and miRNAs Exhibit Strand Bias. *Cell* **115**, 209–216 (2003).

705. Alghoul, F., Eriani, G. & Martin, F. RNA Secondary Structure Study by Chemical Probing Methods Using DMS and CMCT. *Methods Mol Biol* **2300**, 241–250 (2021).

706. Burmistrz, M., Krakowski, K. & Krawczyk-Balska, A. RNA-Targeting CRISPR-Cas Systems and Their Applications. *Int J Mol Sci* **21**, E1122 (2020).

707. Martin, S. L., Li, J. & Weisz, J. A. Deletion analysis defines distinct functional domains for protein-protein and nucleic acid interactions in the ORF1 protein of mouse LINE-1. *J Mol Biol* **304**, 11–20 (2000).

708. Myers, R. M., Fischer, S. G., Lerman, L. S. & Maniatis, T. Nearly all single base substitutions in DNA fragments joined to a GC-clamp can be detected by denaturing gradient gel electrophoresis. *Nucleic Acids Research* **13**, 3131–3145 (1985).

709. Sheffield, V. C., Cox, D. R., Lerman, L. S. & Myers, R. M. Attachment of a 40-base-pair G + C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. *Proceedings of the National Academy of Sciences* **86**, 232–236 (1989).

710. Gao, Z., Herrera-Carrillo, E. & Berkhout, B. Delineation of the Exact Transcription Termination Signal for Type 3 Polymerase III. *Mol Ther Nucleic Acids* **10**, 36–44 (2017).

711. Boden, D. *et al*. Enhanced gene silencing of HIV-1 specific siRNA using microRNA designed hairpins. *Nucleic Acids Res* **32**, 1154–1158 (2004).

712. Westerhout, E. M., Ooms, M., Vink, M., Das, A. T. & Berkhout, B. HIV-1 can escape from RNA interference by evolving an alternative structure in its RNA genome. *Nucleic Acids Res* **33**, 796–804 (2005).

713. Haasnoot, J., Westerhout, E. M. & Berkhout, B. RNA interference against viruses: strike and counterstrike. *Nat Biotechnol* **25**, 1435–1443 (2007).

714. Aqil, M., Naqvi, A. R., Bano, A. S. & Jameel, S. The HIV-1 Nef Protein Binds Argonaute-2 and Functions as a Viral Suppressor of RNA Interference. *PLOS ONE* **8**, e74472 (2013).

715. IUPAC Codes. https://www.bioinformatics.org/sms/iupac.html.

716. ter Brake, O., Konstantinova, P., Ceylan, M. & Berkhout, B. Silencing of HIV-1 with RNA interference: a multiple shRNA approach. *Mol Ther* **14**, 883–892 (2006).

717. Coburn, G. A. & Cullen, B. R. Potent and specific inhibition of human immunodeficiency virus type 1 replication by RNA interference. *J Virol* **76**, 9225–9231 (2002).

718. Rossi, J. J. RNAi as a treatment for HIV-1 infection. *BioTechniques* **40**, S25–S29 (2006).

719. Vordermark, D., Shibata, T. & Brown, J. M. Green fluorescent protein is a suitable reporter of tumor hypoxia despite an oxygen requirement for chromophore formation. *Neoplasia* **3**, 527–534 (2001).

720. Schmitz, C. *et al*. Live reporting for hypoxia: Hypoxia sensor-modified mesenchymal stem cells as in vitro reporters. *Biotechnol Bioeng* **117**, 3265–3276 (2020).

721. Tsien, R. Y. The green fluorescent protein. *Annu Rev Biochem* **67**, 509–544 (1998).

722. Reid, B. G. & Flynn, G. C. Chromophore formation in green fluorescent protein. *Biochemistry* **36**, 6786–6791 (1997).

723. Kumagai, A. *et al*. A bilirubin-inducible fluorescent protein from eel muscle. *Cell* **153**, 1602–1611 (2013).

724. Asp, M., Bergenstråhle, J. & Lundeberg, J. Spatially Resolved Transcriptomes—Next Generation Tools for Tissue Exploration. *BioEssays* **42**, 1900221 (2020).

725. Vordermark, D. *et al*. Glucose requirement for hypoxic accumulation of hypoxia-inducible factor-1α (HIF-1α). *Cancer Letters* **230**, 122–133 (2005).

726. Osada-Oka, M., Hashiba, Y., Akiba, S., Imaoka, S. & Sato, T. Glucose is necessary for stabilization of hypoxia-inducible factor-1α under hypoxia: Contribution of the pentose phosphate pathway to this stabilization. *FEBS Letters* **584**, 3073–3079 (2010).

727. Karuppagounder, S. S. *et al*. In vitro ischemia suppresses hypoxic induction of hypoxia-inducible factor-1α by inhibition of synthesis and not enhanced degradation. *Journal of Neuroscience Research* **91**, 1066–1075 (2013).

728. Zheng, X. *et al*. Proteomic analysis for the assessment of different lots of fetal bovine serum as a raw material for cell culture. Part IV. Application of proteomics to the manufacture of biological drugs. *Biotechnol Prog* **22**, 1294–1300 (2006).

729. Freshney, R. I. *Culture of animal cells: a manual of basic technique and specialized applications*. (Wiley-Blackwell, 2010).

730. Hashemi, M., Ghavami, S., Eshraghi, M., Booy, E. P. & Los, M. Cytotoxic effects of intra and extracellular zinc chelation on human breast cancer cells. *Eur J Pharmacol* **557**, 9–19

(2007).

731. Lennon, A. J., Scott, N. R., Chapman, B. E. & Kuchel, P. W. Hemoglobin affinity for 2,3-bisphosphoglycerate in solutions and intact erythrocytes: studies using pulsed-field gradient nuclear magnetic resonance and Monte Carlo simulations. *Biophys J* **67**, 2096–2109 (1994).

732. Martinotti, S. & Ranzato, E. Scratch Wound Healing Assay. *Methods Mol Biol* **2109**, 225–229 (2020).

733. Tomasova, L., Guttenberg, Z., Hoffmann, B. & Merkel, R. Advanced 2D/3D cell migration assay for faster evaluation of chemotaxis of slow-moving cells. *PLoS One* **14**, e0219708 (2019).

734. Draoui, N. & Feron, O. Lactate shuttles at a glance: from physiological paradigms to anti-cancer treatments. *Dis Model Mech* **4**, 727–732 (2011).

735. Chang, Y.-N. *et al*. Hypoxia-regulated lncRNAs in cancer. *Gene* **575**, 1–8 (2016).

736. Warford, A., Howat, W. & McCafferty, J. Expression profiling by high-throughput immunohistochemistry. *Journal of Immunological Methods* **290**, 81–92 (2004).

737. Crosetto, N., Bienko, M. & van Oudenaarden, A. Spatially resolved transcriptomics and beyond. *Nat Rev Genet* **16**, 57–66 (2015).

738. Kim, D. *et al*. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* **14**, R36 (2013).

739. Ghosh, S. & Chan, C.-K. K. Analysis of RNA-Seq Data Using TopHat and Cufflinks. *Methods Mol Biol* **1374**, 339–361 (2016).

740. Deschamps-Francoeur, G., Simoneau, J. & Scott, M. S. Handling multi-mapped reads in RNA-seq. *Computational and Structural Biotechnology Journal* **18**, 1569–1576 (2020).

741. Baruzzo, G. *et al*. Simulation-based comprehensive benchmarking of RNA-seq aligners. *Nat Methods* **14**, 135–139 (2017).

742. Dobin, A. *et al*. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

743. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907–915 (2019).

744. Sarkar, D., Leung, E. Y., Baguley, B. C., Finlay, G. J. & Askarian-Amiri, M. E. Epigenetic regulation in human melanoma: past and future. *Epigenetics* **10**, 103–121 (2015).

745. Fullwood, M. J. & Ruan, Y. ChIP-based methods for the identification of long-range chromatin interactions. *J Cell Biochem* **107**, 30–39 (2009).

746. Fullwood, M. J., Wei, C.-L., Liu, E. T. & Ruan, Y. Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome Res.* **19**, 521–532 (2009).

747. Frohman, M. A. Rapid amplification of complementary DNA ends for generation of full-length complementary DNAs: thermal RACE. *Methods Enzymol* **218**, 340–356 (1993).

748. Takahashi, H., Lassmann, T., Murata, M. & Carninci, P. 5' end-centered expression profiling using cap-analysis gene expression and next-generation sequencing. *Nat Protoc* **7**, 542–561 (2012).

749. Gaberc-Porekar, V. & Menart, V. Potential for Using Histidine Tags in Purification of Proteins at Large Scale. *Chemical Engineering & Technology* **28**, 1306–1314 (2005).

750. Chant, A., Kraemer-Pecore, C. M., Watkin, R. & Kneale, G. G. Attachment of a histidine tag to the minimal zinc finger protein of the Aspergillus nidulans gene regulatory protein AreA causes a conformational change at the DNA-binding site. *Protein Expr Purif* **39**, 152–159 (2005).

751. Roux, K. J., Kim, D. I., Burke, B. & May, D. G. BioID: A Screen for Protein-Protein Interactions. *Curr Protoc Protein Sci* **91**, 19.23.1-19.23.15 (2018).

752. Sears, R. M., May, D. G. & Roux, K. J. BioID as a Tool for Protein-Proximity Labeling in Living Cells. *Methods Mol Biol* **2012**, 299–313 (2019).

753. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **44**, D7–D19 (2016).

754. Eng, C.-H. L. *et al*. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **568**, 235–239 (2019).

755. Marx, V. Method of the Year: spatially resolved transcriptomics. *Nat Methods* **18**, 9–14 (2021).

756. Xiao, T. & Zhou, W. The third generation sequencing: the advanced approach to genetic diseases. *Translational Pediatrics* **9**, 163 (2020).

757. Moldován, N. *et al*. Third-generation Sequencing Reveals Extensive Polycistronism and Transcriptional Overlapping in a Baculovirus. *Sci Rep* **8**, 8604 (2018).

758. Thomas, S. N. *et al*. Exosomal Proteome Profiling: A Potential Multi-Marker Cellular Phenotyping Tool to Characterize Hypoxia-Induced Radiation Resistance in Breast Cancer. *Proteomes* **1**, 87–108 (2013).

759. Wilson, W. R. & Hay, M. P. Targeting hypoxia in cancer therapy. *Nat Rev Cancer* **11**, 393–410 (2011).

760. Taylor, B. L. & Zhulin, I. B. PAS Domains: Internal Sensors of Oxygen, Redox Potential, and Light. *Microbiology and Molecular Biology Reviews* **63**, 479–506 (1999).

Supplemental Material

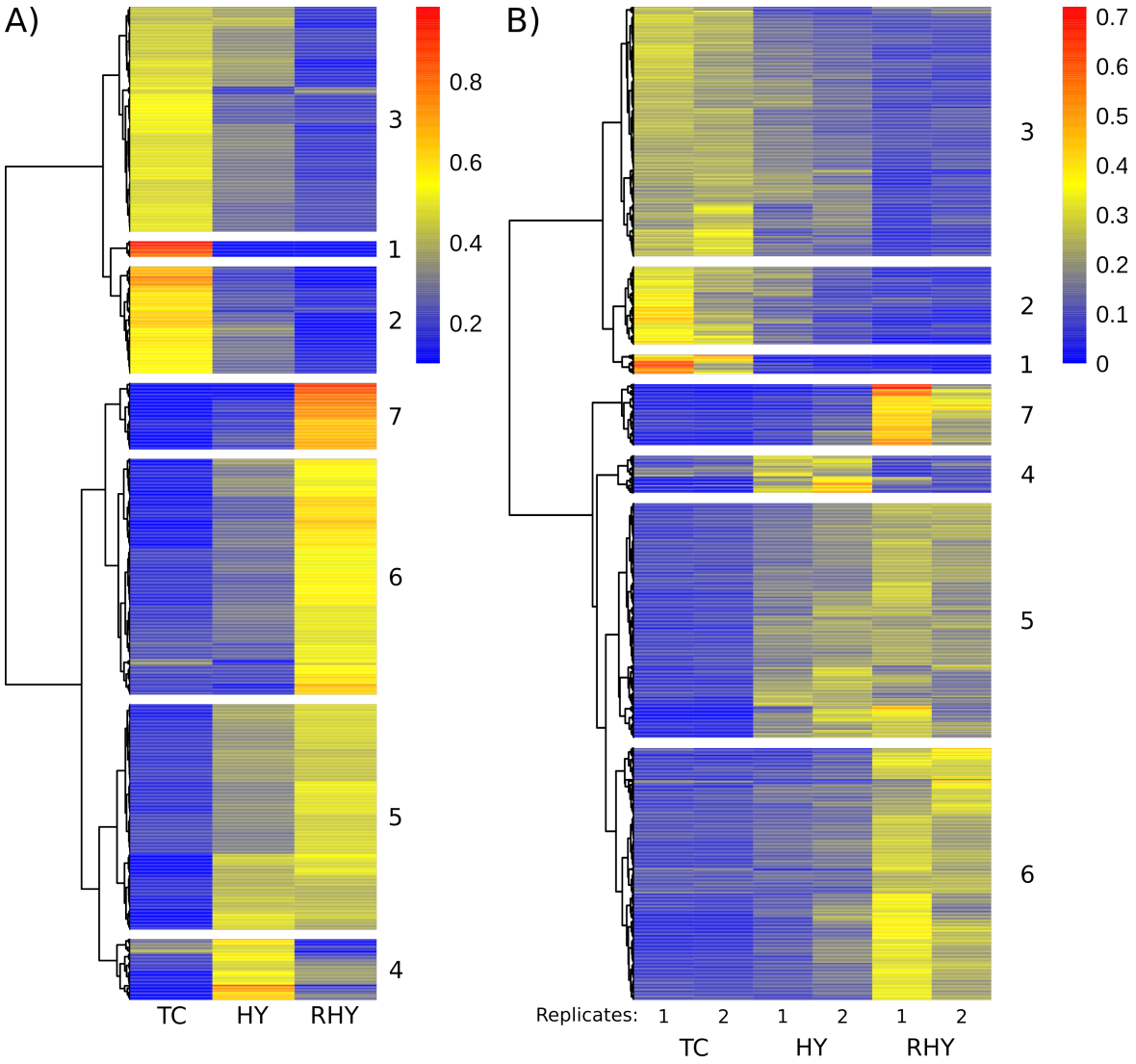

Figure 58: Supplemental; MHY DEG heatmap clustering justification and replicate variability

A) MHY DEG heatmap as shown in Chapter 3 with original dendrogram. Note that cluster numbers are out of order as presented in Chapter 3. Clusters were re-ordered without breaking the dendrogram's linkage for aestetic pruposes. The same methodology was used as in Chapter 3. **B)** MHY DEG heatmap rendered using DESeq2 normalized readcounts for replicates rather than average values between replicates. The same methodology was used as in Chapter 3.

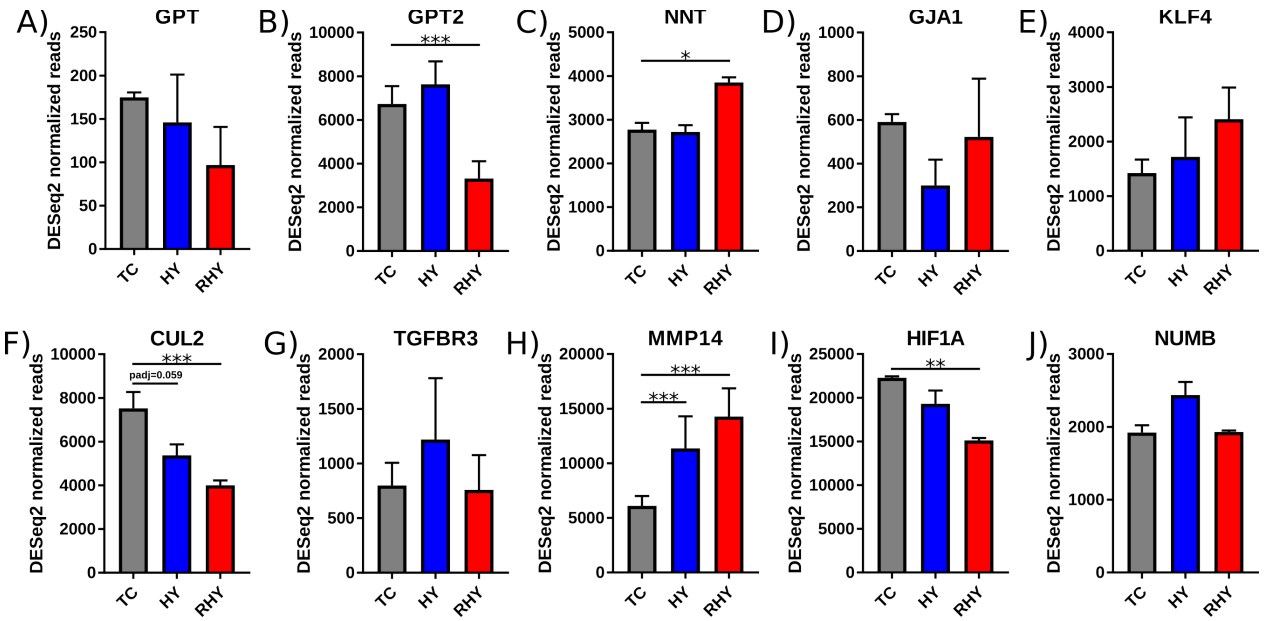

Figure 59: Supplemental; MHY literature-linked gene expression

A-J) Gene expression histograms of genes mentioned in Chapter 3. Expression is DESeq2 normalized reads, error bars are SD. * denotes relative significance as calculated by DESeq2 Benjamini-Hochberg adjusted p-value (padj). *: padj < 0.05, **: padj < 0.01, ***: padj < 0.001.

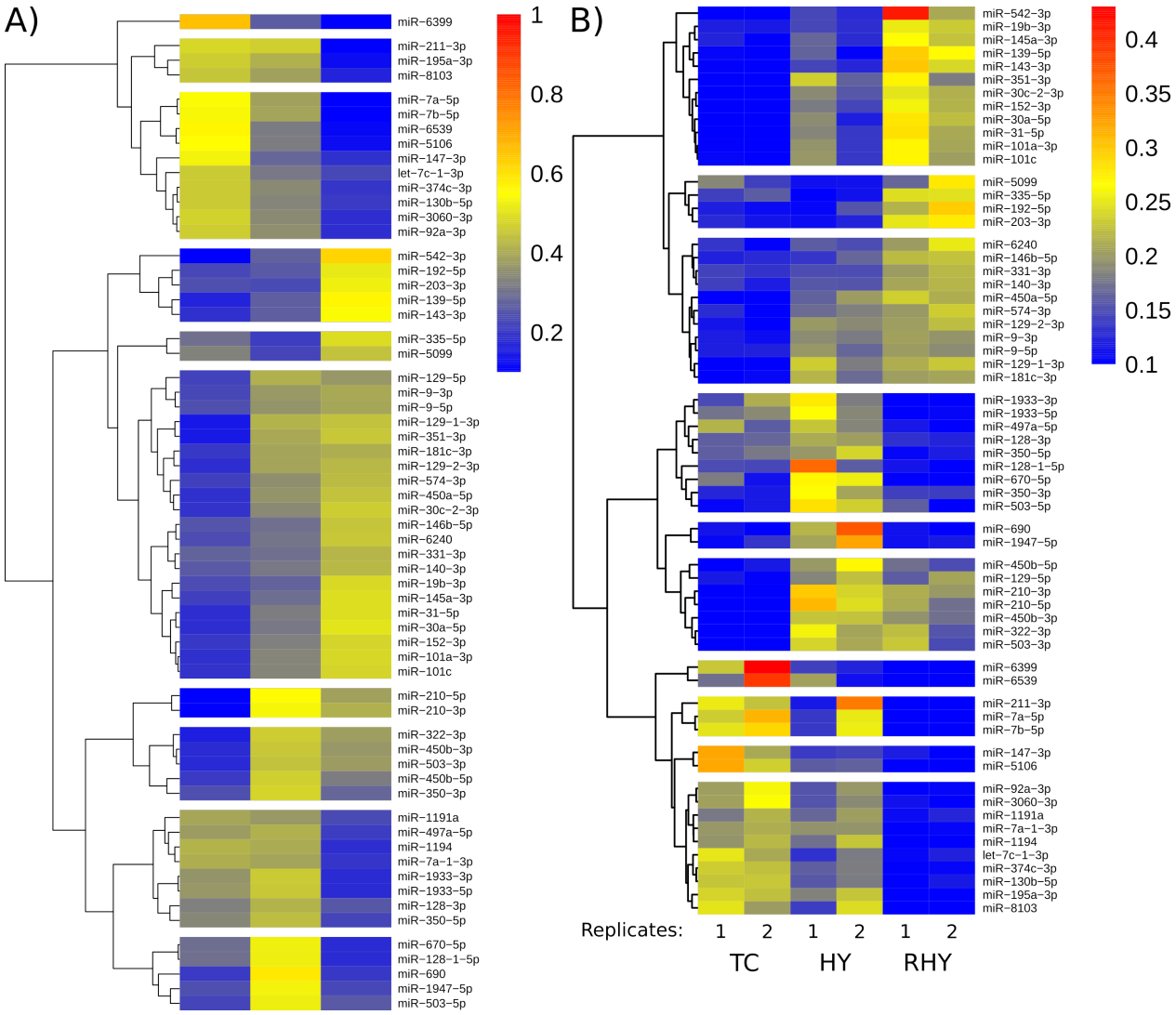

Figure 60: Supplemental; MHY DEmiR heatmap clustering justification and replicate variability

A) MHY DEmiR heatmap as shown in Chapter 3 with original dendrogram. **B)** MHY DEG heatmap rendered using DESeq2 normalized readcounts for replicates rather than average values between replicates. Colorbar set to automatic rather than forced to 0.1-1 .

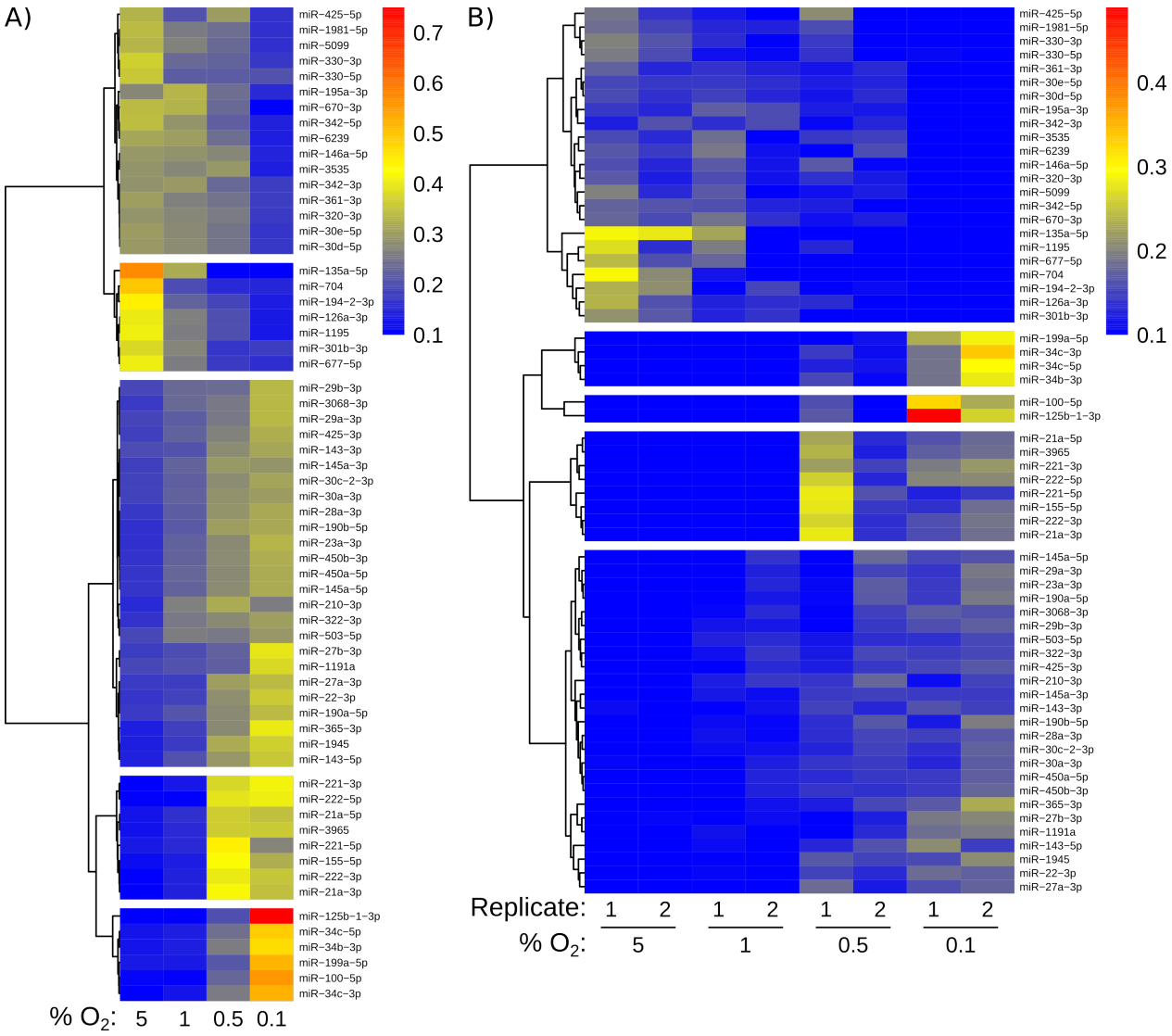

Figure 61: Supplemental; LTHY DEmiR heatmap with dendrogram and replicates

A) Original heatmap of **Figure 26D**. Clusters were re-aranged without breaking the dendogram's linkage for aestetic purposes. **B**) Heirarchically clustered heatmap of LTHY DEmiRs, usirng replicates rather than averaged values. Same normalization and clustering methods used as in **Figure 26D**.

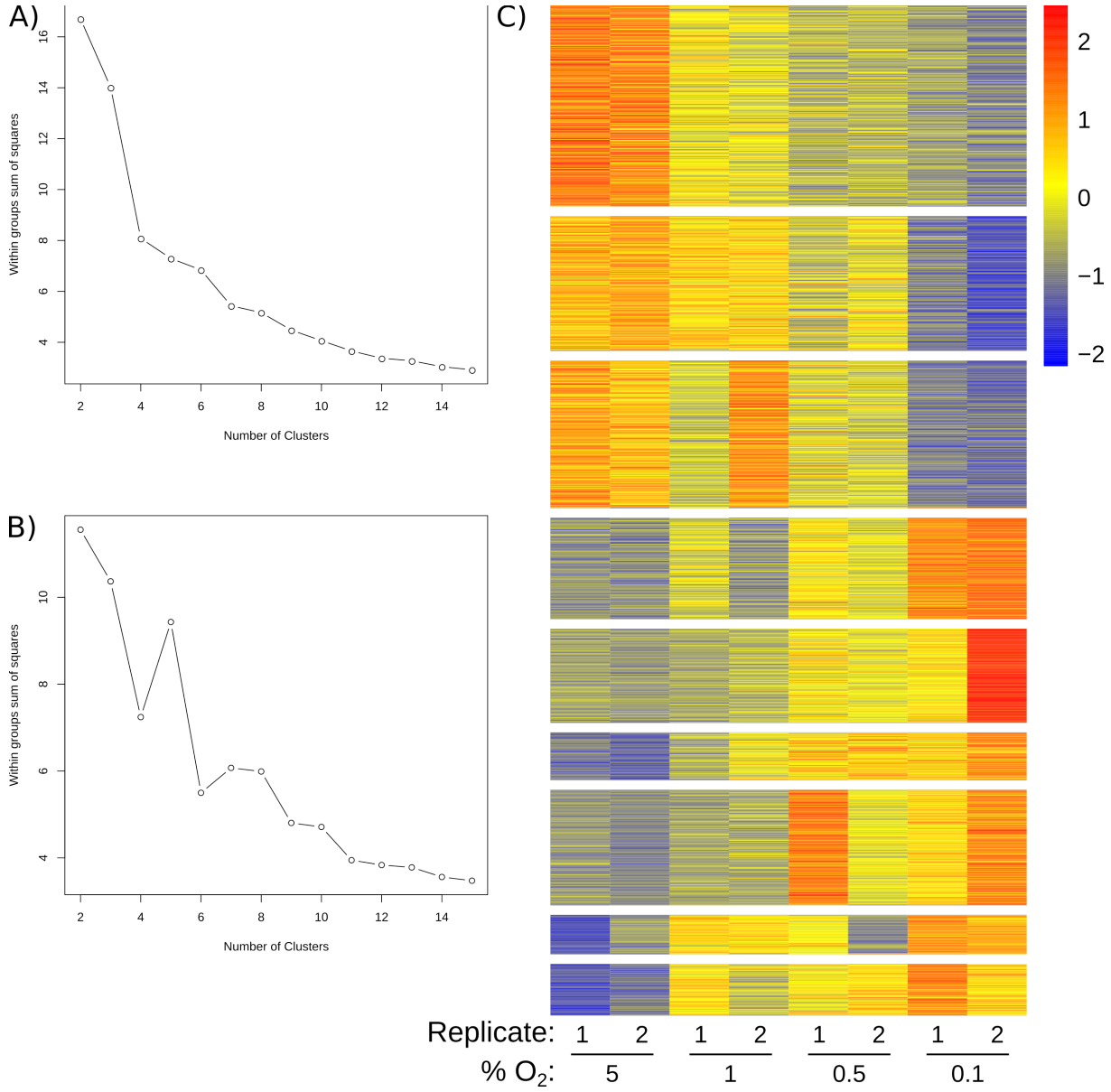

Figure 62: Supplemental; LTHY DEG kmeans clustering validation and replicate rendered

A) Cluster number validation test for Figure 27A. Intra-cluster variability is plotted against number of clusters. **B**) Cluster number validation test for replatice-based LTHY DEG data, same methdology used as in (A). **C**) kmeans clustered heatmap of LTHY DEGs using replicate data. Same methodology used as in **Figure 27A**.

Table 18: GFP NT3'UTR sa/shRNA constructs predicted folding

Site	Construct	Folding	MFE (kcal/mol)
Site 1	shRNA	CGGTTATCGCAATCACGACCC ||||||||||||||||||| GCCAATAGCGTTAGTGCTGGG	-60.843
	3nt bridge	CGGTTATCGCAATCACGACCC ||||||||||:::|||||||| GCCAATAGCGaatGTGCTGGG	-54.854
	4nt bridge	CCGGTTATCGCAATCACGACCC ||||||||||::| ||||||| GGCCAATAGCaat-GTGCTGGG	-52.832
	5nt bridge	CCCGGTTATCGCAATCACGACCC ||||||||||::| ||||||| GGGCCAATAGaat--GTGCTGGG	-54.579
Site 2	shRNA	CCCCGTAATTAAATTCGCGCG ||||||||||||||||||| GGGGCATTAATTTAAGCGCGC	-59.898
	3nt bridge	CCCCGTAATTAAATTCGCGCG ||||||||||:::|||||||| GGGGCATTAAaaaAAGCGCGC	-54.181
	4nt bridge	ACCCCGTAATTAAATTCGCGCG |||||||||| |::||||||| TGGGGCATTA-taaAAGCGCGC	-53.093
	5nt bridge	GACCCCGTAATTAAATTCGCGCG ||||||||||:| ||||||| CTGGGGCATTtaa--AAGCGCGC	-53.708
Site 3	shRNA	CTACTCACACGTTAGCCAGTC ||||||||||||||||||| GATGAGTGTGCAATCGGTCAG	-60.082
	3nt bridge	CTACTCACACGTTAGCCAGTC ||||||||||:::|||||||| GATGAGTGTGattTCGGTCAG	-51.167
	4nt bridge	GCTACTCACACGTTAGCCAGTC |||||||||| :::||||||| CGATGAGTGT-attTCGGTCAG	-52.202
	5nt bridge	CGCTACTCACACGTTAGCCAGTC |||||||| |:::||||||| GCGATGAGT--GattTCGGTCAG	-52.553

Source code for WT1 DNA binding motif searching script:

```
import sys
from Bio import Entrez, SeqIO, SeqUtils, motifs, pairwise2
from Bio.Align import substitution_matrices
```

```
import os
import pickle
from urllib.error import HTTPError
import time

Entrez.email = 'jordan.quenneville@umontreal.ca'

mmu_chrom2_acc = {'1':"CM000994.3",
        '2':'CM000995.3',
        '3':'CM000996.3',
        '4':'CM000997.3',
        '5':'CM000998.3',
        '6':'CM000999.3',
        '7':'CM001000.3',
        '8':'CM001001.3',
        '9':'CM001002.3',
        '10':'CM001003.3',
        '11':'CM001004.3',
        '12':'CM001005.3',
        '13':'CM001006.3',
        '14':'CM001007.3',
        '15':'CM001008.3',
        '16':'CM001009.3',
        '17':'CM001010.3',
        '18':'CM001011.3',
        '19':'CM001012.3',
        'X':'CM001013.3',
        'Y':'CM001014.3'
        }


def fetch_motifs():
        #returns list of biopython motif objects
        motif_fi_l = []
        for dirpath, dirnames, files in os.walk(os.path.abspath(os.curdir+'/Motifs/')):
                for fi in files:
                        if fi[-7:] == '.matrix': #actually tab format
                                motif_fi_l.append(fi)
                                path = dirpath
        motif_dic = {} #{motif_name : motif object}
        for fi in motif_fi_l:
                with open(path+'/'+fi, 'r') as f:
                        temp = motifs.read(f, 'pfm-four-columns')
                        motif_dic[fi.replace('.matrix', '')] = temp
        return motif_dic


def fetch_seq(chrom, start, end):
        #expecting string casts of all 3, and 'chr' to be removed
        print(chrom, start, end)
        time.sleep(0.5)
        if not chrom in mmu_chrom2_acc.keys():
                print('region not part of standard chromosomes, skipping')
                return 'NA'
        try:
                handle = Entrez.efetch(db="nucleotide",
                                                id=mmu_chrom2_acc[chrom],
                                                rettype="fasta",
                                                strand=1,
                                                seq_start=int(start),
                                                seq_stop=int(end))
                record = SeqIO.read(handle, "fasta")
                handle.close()
        except HTTPError:
                print('HTTPError encountered, likely hit request limit')
                return 'NA'
        return record.seq
```

```
motif_dic = fetch_motifs()

peak_fi_l = []
for dirpath, dirnames, files in os.walk(os.path.abspath(os.curdir+'/annotated_tables/')):
	for fi in files:
		if fi[-4:] == '.csv': #actually tab format
			peak_fi_l.append(fi)

header = []
peak_dic = {} #{file:{peakID:[chrom, start, end, ...]}}
for fi in peak_fi_l:
	peak_dic[fi] = {}
	with open(dirpath+'/'+fi, 'r') as f:
		header_switch = 0
		for i, line in enumerate(f):
			line = line.strip().split('\t')
			if i == 0:
				for x, e in enumerate(line):
					if e == 'Chr':
						chr_index = x
					elif e == 'Start':
						start_index = x
					elif e == 'End':
						end_index = x
					elif e == 'Peak Score':
						score_index = x
					elif 'PeakID' in e:
						id_index = x
					elif e == 'Gene Name':
						name_index = x
					elif e == 'Annotation':
						ann_index = x
			else:
				peak_dic[fi][line[id_index]] = [line[chr_index].replace('chr', ''),
line[start_index], line[end_index],

	line[score_index], line[name_index], line[ann_index]]

if not 'pickle' in sys.argv:
	for fi in peak_dic.keys():
		print(fi)
		for i, k in enumerate(peak_dic[fi].keys()):
			seq = fetch_seq(peak_dic[fi][k][0], peak_dic[fi][k][1], peak_dic[fi][k][2])
			peak_dic[fi][k] = peak_dic[fi][k] + [seq]
	f = open('peak_dic_pickle', 'wb')
	pickle.dump(peak_dic, f)
	f.close()
else:
	peak_dic = pickle.load(open('peak_dic_pickle', 'rb'))

blosum62 = substitution_matrices.load("BLOSUM62")
for motif_id in motif_dic.keys():
	for fi in peak_dic.keys():
		for k in peak_dic[fi].keys():
			seq = peak_dic[fi][k][-1]
			seq_OGS = peak_dic[fi][k][4]
			if seq == 'NA':
				continue
			else:
				motif = motif_dic[motif_id]
				search = SeqUtils.nt_search(str(seq), motif.degenerate_consensus)
				if len(search) > 1:
					print(seq_OGS, motif_id, search) #; sys.exit()
				search_rc = SeqUtils.nt_search(str(seq.reverse_complement),
motif.degenerate_consensus)
				continue
```

```
                      '''
                      else:
                              motif = motif_dic[motif_id]
                              alignments = pairwise2.align.localds(seq,
str(motif.degenerate_consensus), blosum62, -10, -1)
                              print(seq_OGS, motif_id)
                              print(pairwise2.format_alignment(*alignments[0]))
                              #sys.exit()


'''

'''

sys.exit()


# junk code for testing queries
handle = Entrez.efetch(db="nucleotide",
                                        id="307603377",
                                        rettype="fasta",
                                        strand=1,
                                        seq_start=4000100,
                                        seq_stop=4000200)
record = SeqIO.read(handle, "fasta")
handle.close()
print(record.seq)

handle = Entrez.efetch(db="nucleotide",
                                        id="CM000994.3",
                                        rettype="fasta",
                                        strand=1,
                                        seq_start=4000100,
                                        seq_stop=4000200)
record = SeqIO.read(handle, "fasta")
handle.close()
print(record.seq)
```

~~~~~~

Source code for saRNA seed evaluation script:

```
#!/usr/bin/env python
# coding: utf-8
# written in python2.7
import sys
import itertools
import numpy

nt1 = ''
nt2 = ''
bulges = [0,3,4,5,6,7,8] # list of bridge lengths to test.

def calc_GC(seq): #expected to be pre-parsed
        counter = 0
        if len(seq) == 0:
                return -1.0
        for c in seq:
                if c in 'GC':
                        #print 'pass'
                        counter +=1
        return float(counter)/float(len(seq))*100.0

def make_revComp(seq):
        comp = ''
        for c in seq:
                if c == 'G':
```
~~~~~~

```
                        comp=comp+'C'
                elif c == 'C':
                        comp=comp+'G'
                elif c == 'A':
                        comp=comp+'U'
                elif c == 'U':
                        comp=comp+'A'
                else:
                        print 'wtf', seq ;sys.exit()
        return comp[::-1]

with open('NT_seq1', 'r') as f:
        for line in f:
                nt1 = line.strip().upper()

with open('NT_seq2', 'r') as f:
        for line in f:
                nt2 = line.strip().upper()

nt1_rna = nt1.replace('T', 'U')
nt2_rna = nt2.replace('T', 'U')

seqs = [nt1_rna, nt2_rna]

with open('output.tab', 'w') as w:
        base_header = ['seqName', 'seedIndex', 'seedSeq']
        header = base_header+['bulge_'+str(x)+'_GC' for x in bulges]
        header = header+['stdDev', 'mean','seed_RC']
        w.write('\t'.join(header))
        w.write('\n')
        for x,seq in enumerate(seqs):
                start = 0
                end = 21
                bulge_data = {}
                var_data =[]
                for i, c in enumerate(seq):
                        if 1 + end > len(seq):
                                break
                        else:
                                seed = seq[start+13:end-1]
                                tail = seq[start:end-8]
                                for bulge_size in bulges:
                                        try:
                                                bulge_data[str(start+13)+':'+str(end-
1)].append(calc_GC(seq[start-bulge_size:end-11]))
                                        except KeyError:
                                                bulge_data[str(start+13)+':'+str(end-1)]    =    [seed,
calc_GC(seq[start-bulge_size:end-11])]

                                bulge_data[str(start+13)+':'+str(end-
1)].append(numpy.std(bulge_data[str(start+13)+':'+str(end-1)][1:]))
                                bulge_data[str(start+13)+':'+str(end-
1)].append(numpy.mean(bulge_data[str(start+13)+':'+str(end-1)][1:-1]))
                                var_data.append(bulge_data[str(start+13)+':'+str(end-1)][-2])
                                start += 1
                                end += 1
                print len(bulge_data.keys())
                for seed_range in bulge_data.keys():
                        line  =  ['seq'+str(x),  seed_range.replace(':',  '::')]+[str(y)  for  y  in
bulge_data[seed_range]]+[make_revComp(bulge_data[seed_range][0])]
                        w.write('\t'.join(line))
                        w.write('\n')

# D    C   B   A   seed   no one cares about U
#GGAU CCG ACU CGU AGUAUAA U
# tail is 13nt long
```

~~~~~~

SynmiR design algorithm 2; ration-design pseudo-code:

```
parse design parameters:
       target(s) sequence file location
       minimal/maximal bridge sequence length
       target sequence extension length used for folding
       number of target sites to design on when in single-target mode
       maximum allowed polyU stretch in synmiR designs
       Bregion binding sequences score threshold
       multiprocessing poolsize
       synmiR Bregion start, end, and synmiR total length

read in target sequence(s)
parse target sequence(s) using seedScan.

seedScan:
       read target sequence(s), build seed:targeting_masterDict
       seed:targeting_masterDict structure:
       {seed_seq : {
                      target_name: [
                                    [[“Bregion_seq”, bridgeIndex, “Bridge_seq”, seedIndex],
                                     [“Bregion_seq”, bridgeIndex, “Bridge_seq”, seedIndex],
                                     ...
                                    ],
                                    [[site2 Bregion data]],
                                    [[site3 Bregion data]],
                                   ...
                                   ],
                    target2_name: [...]
                   },
        seed2_seq: {...},
        ...
       }

       example with HIV data (single target sequence example):
          {"GTCCCAA": {
                      "pNL_HIV_seq": [
                      [["AGGGC", 7, "TAATTTG", 23],
                       ["GGGCT", 6, "AATTTG", 23],
                       ["GGCTA", 5, "ATTTG", 23],
                       ["GCTAA", 4, "TTTG", 23]]
                      ]
                      },
            "TCCCAAA": ...,
               ...
         }
```
~~~~~~

Select seeds from seed:targeting_masterDict which match design parameters:

if multi-sequence targeting, seed must be present on all targets

if single-sequence targeting, seed must be present the exact number of times specified in design parameters

seeds which pass these parameters are passed into a new dictionary: targeting_dict

Evaluate all Bregion binding site combinations for a given seed in parallel using the following scoring function

Scoring function:

Score starts at 1

for each nt in Bregion binding site consensus sequence, multiply score using the following formula:

new_score = current_score * (consensus_nt_weight + positional_weight)

Where the consensus_nt_weight is determined using the following table:

consensus is all G or C	= 1.5
consensus is all A or U	= 1.25
consensus is all GA or CU (G:U bonding)	= 1.1
consensus is other two-nucleotide combinations	= 0.5
consensus is other three-nucleotide combinations	= 0.25
consensus is all four nucleotides	= 0.1

Where positional_weight is calculated using the following formula:

0.5 * (consensus_position / consensus_length)

return final score

Retain top scoring Bregion binding site combinations that pass the score threshold as defined in the design parameters

For each of these passing seed:Bregion_binding_sites combinations, design the Abox, Bregion, and Dregion of the synmiR

Abox design function:

evaluate the bridge sequences for a given seed:Bregion_binding_site combination

if bridge sequences are 3nts or less, maximize binding

if bridge sequences are 3nts or more, minimize binding

prioritize the binding preferences of short bridges

evaluate binding ability using mcff using:

5' and 3' GC clamp

the bridge sequence

all possible Abox sequences

Bregion design function:

deconvolute the Bregion into all possible nucleotide sequences

continue designing on all

Dregion design function:

Given the set of target Dregion binding site sequences

Design a Dregion sequence which maximizes binding to the weakest Bregion binding site

Try to get at least 5nt of contiguous tail-sequence binding

Otherwise, maximize binding to Dregion binding sites

```
Final design QC and folding:
        remove designs which introduce polyU stretches > specified amount
        using mcff, fold all design on local target sequences
                including local target sequence extension
        parse dot-bracket structure into double-stranded structure.

Output all synmiR data, sequences, and foldings to file
```

* * *

SynmiR design algorithm 2; rational-design source code:

```
#!/usr/bin/env python
# coding: utf-8

# correct call: python3 miRDesign_jq_v8.py tars=pNL-HIV.fa Bregion_start_offset=1
max_tar_bulge=12 local_tar_seq_len_extension=10 single_tar_sites_allowed=3 maxPolyU=3
BscoreThresh=0.85 poolsize=2 regions=12,16,21

import sys
import os
import subprocess
from Bio import SeqIO
from Bio.Seq import Seq
import statistics as stats
import regex as re
import numpy as np
import itertools
import pandas as pd
import time
import multiprocessing as mp
from statistics import mean
# keep getting numpy VisibleDeprecationWarning despite not using numpy. Include this code to
suppress warnings
import warnings
warnings.filterwarnings("ignore", category=np.VisibleDeprecationWarning)


print('time at start:', time.asctime())


def mcffCommandLine( strand1, strand2, mcffVersion = '35', explore = '10' ):
        return 'mcff' + mcffVersion + ' -s ' + strand1 + ' -sd ' + strand2 + ' -vv -e ' + explore
+ ' -alt 2> /dev/null'
def mcffCommandLine_v2( strand1, strand2, mcffVersion = '35', explore = '10' ):
        return '/home/jordanquenneville/MC-Flashfold-v35/src/mcff'  + ' -s ' + strand1 + ' -sd ' +
strand2 + ' -vv -e ' + explore + ' -alt 2> /dev/null' #+' -tables /home/jordanquenneville/MC-
Flashfold/tables'
def mcffCommandLine_Abox( strand1, strand2, mask ,mcffVersion = '35', explore = '10' ):
        return '/home/jordanquenneville/MC-Flashfold-v35/src/mcff'  + ' -s ' + strand1 + ' -sd ' +
strand2 + ' -vv -e ' + explore + ' -alt -m \''+mask+'\' 2> /dev/null'

def findLongestRepeat(string, char):
        string = string.upper()
        char = char.upper()
        repeats = []
        curr_rep = 0
        for c in string:
                if c == char:
```

```
                        curr_rep +=1
                else:
                        repeats.append(curr_rep)
                        curr_rep = 0
        repeats.append(curr_rep)
        if len(repeats) > 0:
                return max(repeats)
        else:
                return 0

# remove final designs with poly U > maxPolyU passed variable
def remove_polyU_polyA(df, colname):
        # takes main df (or any df with a 'fullDesign' column)
        # returns df with designs that have minimal polyU in them
        # gets maxPolyU from initial script call, and is defaulted to 3
        # 3 is the max allowed polyU stretch based on
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5725217/

        indeces_to_drop = []
        for index, row in df.iterrows():
                longestPolyU = findLongestRepeat(row[colname].upper().replace('T','U'), 'U')
#;print(longestPolyU)
                longestPolyA = findLongestRepeat(row[colname], 'A') #;print(longestPolyA)
                print(row[colname], longestPolyU, longestPolyA)
                if longestPolyU > maxPolyU or longestPolyA > maxPolyU:
                        indeces_to_drop.append(index)
        print(indeces_to_drop)
        print('# of searches',len(df.index), '# indeces dropped',len(indeces_to_drop))
        if not len(indeces_to_drop) == 0:
                df2 = df.drop(indeces_to_drop)
                return df2
        else:
                return df

def hybridiation_consensus_generator_v3_Bartel2022(t_seqs):
        # based on Bartel 2022 paper DOI: 10.7554/eLife.69803
        # penalize non-canonical binding, including GU
        # add GC clamp to seed side of region

        # convert input to lower case
        t_seqs = [x.lower() for x in t_seqs]

        # weights to score based on prioritizing gc
        n2_weight = 0.5
        n3_weight = 0.25
        n4_weight = 0.1
        gc_clamp_score = 1.25

        # initialize scoring variables
        score = 1.0
        consensus = ''
        num_poss = 1
        num_seqs = len(t_seqs)
        seq_len = len(t_seqs[0]) # assumes all sequences are the same length

        score_weights={
        'a':[1.25, 'u'],
        'u':[1.25, 'a'],
        'g':[1.5, 'c'],
        'c':[1.5, 'g'],
        'ag':[0.95, 'u'],
        'cu':[0.95, 'g'],
        'cg':[n2_weight, 's'],
        'au':[n2_weight, 'w'],
        'ac':[n2_weight, 'k'],
        'gu':[n2_weight, 'm'],
        'cgu':[n3_weight, 'v'],
```

```
        'agu':[n3_weight, 'h'],
        'acu':[n3_weight, 'd'],
        'acg':[n3_weight, 'b'],
        'acgu':[n4_weight,'n']
                        }
        prev_gc_clamp_switch = 0
        for i, char_set in enumerate(zip(*t_seqs)):
                # positional weight calculated here; prioritizes end binding (seed proximal)
                posi_weight = 0.5 + float(i)/seq_len #; print(posi_weight)

                # sanitize input
                char_set = ''.join(char_set)
                if 't' in char_set:
                        char_set = char_set.replace('t', 'u')
                char_set = ''.join(sorted(list(set(char_set)))) # sort char_set alphabetically
                if char_set in score_weights.keys(): # compute scores for all solvable values
                        consensus += score_weights[char_set][1]
                        score *= score_weights[char_set][0] + posi_weight
                else:
                        print('char_set not in dict', i, char_set, t_seqs) ;sys.exit()
                # implement GC clamp scoring
                if i >= seq_len-2 and set(char_set).issubset(set('gc')) and len(char_set) == 1:
                        score *= gc_clamp_score
        return score, num_poss, consensus

def hybridiation_consensus_generator_v2_multi(t_seqs):
        # given a tuple of variable number of RNA seqs
        # where all RNA seqs are the same length
        # aligns all seqs, returns score and miR-tail consensus 3'-> 5'
        # if all seqs at position are G/A, add U to consensus
        # if all seqs at position are C/U, add G to consensus
        # those 2 conditions account for GU pairing and add a 0.75 to score
        # uses standard nucleotide annotation for dual and triple nuclotide hybridization
        # dual nucleotide possibility = score + 0.5
        # triple nucleotide possibility =  score + 0.25
        # n possibility = score + 0
        #       new properties in v2:
        #               higher score to gc binding
        #               higher weight to 'end' (seed proximal) of region

        #       results of testing:
        #                ran with increased weights to G/C on test_data_v2:
        #                        increased scores, didn't change top ranking seed
        #                        but brought other GC rich seeds to the top of the list
        #                        ex: syn-913 was 10th ranked originally; is now second

        # convert input to lower case
        t_seqs = [x.lower() for x in t_seqs]

        # weights to score based on prioritizing gc
        gc_weight = 1.5
        au_weight = 1.25
        gu_weight = 1.10
        n2_weight = 0.5
        n3_weight = 0.25
        n4_weight = 0.1

        score = 1.0
        consensus = ''
        num_poss = 1
        num_seqs = len(t_seqs)
        seq_len = len(t_seqs[0]) # assumes all sequences are the same length
        for i, char_set in enumerate(zip(*t_seqs)):
                # positional weight calculated here; prioritizes end binding (seed proximal)
                posi_weight = 0.5 + float(i)/seq_len #; print(posi_weight)
                if 't' in char_set:
                        char_set = tuple(''.join(char_set).replace('t', 'u'))
```

```
                # sanitize input
            if ''.join(char_set) == 'a'*num_seqs:
                consensus += 'u'
                score *= au_weight +posi_weight
            elif ''.join(char_set) == 'g'*num_seqs:
                consensus += 'c'
                score *= gc_weight +posi_weight
            elif ''.join(char_set) == 'c'*num_seqs:
                consensus += 'g'
                score *= gc_weight +posi_weight
            elif ''.join(char_set) == 'u'*num_seqs:
                consensus +=  'a'
                score *= au_weight +posi_weight
            elif set(char_set).issubset(set('ga')):
                consensus+= 'u'
                score *= gu_weight +posi_weight
                # ie, complements with u thru GU pairing
            elif set(char_set).issubset(set('cu')):
                consensus += 'g'
                score *= gu_weight +posi_weight
                # ie, complements with g thru GU pairing
            elif set(char_set).issubset(set('gc')):
                consensus+='s'
                num_poss = num_poss*2
                score *= n2_weight +posi_weight
            elif set(char_set).issubset(set('au')):
                consensus+='w'
                num_poss = num_poss*2
                score *= n2_weight +posi_weight
            elif set(char_set).issubset(set('ac')):
                consensus+='k'
                num_poss = num_poss*2
                score *= n2_weight +posi_weight
            elif set(char_set).issubset(set('gu')):
                consensus+='m'
                num_poss = num_poss*2
                score *= n2_weight +posi_weight
            elif set(char_set).issubset(set('gcu')):
                consensus+='v'
                num_poss = num_poss*3
                score *= n3_weight +posi_weight
            elif set(char_set).issubset(set('agu')):
                consensus+='h'
                num_poss = num_poss*3
                score *= n3_weight +posi_weight
            elif set(char_set).issubset(set('acu')):
                consensus+='d'
                num_poss = num_poss*3
                score *= n3_weight +posi_weight
            elif set(char_set).issubset(set('agc')):
                consensus+='b'
                num_poss = num_poss*3
                score *= n3_weight +posi_weight
            else:
                consensus += 'n'
                score *= n4_weight +posi_weight
        return score, num_poss, consensus

# sliding window of mRNAs to find common seeds
# each seed is saved with it's potential tail seq for further validation
# depreaciated function, not used anymore
def seedScan_v1(target_l, tail_len = 13):
        # uses 7mer seed type (nt 2-7), relies on further analyses to determine seed category
        # seed is last 7nt of sliding window
        seedScan_output = {}
        window_size = 7+ tail_len +1 # +1 to account for not searching for nt 8 of seed
        for target in target_l:
```

```
                tar_seq = str(target['seq'])
                for i in range(len(tar_seq) - window_size +1):
                        window = tar_seq[i:i+window_size]
                        #print(window, len(window))
                        seed = window[-7:]
                        tail = window[:-7]
                        #print(seedScan_output)
                        try:
                                seedScan_output[seed][target['name']] = seedScan_output[seed]
[target['name']] + [tail]
                        except KeyError:
                                if seed in seedScan_output.keys():
                                        seedScan_output[seed][target['name']] = [tail]


#print('foo----------------------------------------------------------------------')#;sys.exit()
                                else:
                                        seedScan_output[seed] = {target['name']:[tail]}
        return seedScan_output

# current seed scan function
def seedScan_Bartel(target_l, tail_len = 5, bulge_size = 5, start_offset = 1, Bregion_len = 5):
        print('starting seedScan_v2', tail_len, bulge_size, start_offset, Bregion_len)
        # uses 7mer seed type (nt 2-7), relies on further analyses to determine seed category
        # seed is last 7nt of sliding window
        # used Bartel 2007 binding rules: 1-5nt target bulge, pairing with a 6nt window nt 12-17
miR side
        # https://www.sciencedirect.com/science/article/pii/S1097276507004078
        # data struct:
        # {seed_seq1:{target1:[[target1_seq1, bulge_pos, Abox_seq], [target1_seq2,
bulge_pos], ...],
        #                      target2:...},
        #  seed_seq2:...}
        # tail_len = Dregion binding length
        # Bregion_len = Bregion binding length; section that will be optimized for
        # bulge_size = maximum target side bugle
        # start_offset = minimum target side bulge
        # defaults derived from Bartel 2007
        # target_l = a list of target sequences, in DNA annotation
        seedScan_output = {}
        window_size = 7+ tail_len+ Bregion_len +bulge_size # +1 to account for not searching for
nt 8 of seed
        for target in target_l:
                tar_seq = str(target['seq']).upper()
                for i in range(len(tar_seq) - window_size +1):
                        window = tar_seq[i:i+window_size]
                        seed = window[-7:]
                        seed_index = i+window_size
                        tail = []
                        window_range_start = tail_len
                        window_range_end = tail_len + bulge_size - start_offset +1 # +1 for python
range function
                        for x in range( window_range_start, window_range_end):
                                bulge_index = len(window) - x- len(seed) - Bregion_len
                                tail_seq = window[x:x+Bregion_len]
                                tar_AboxSeq = window[x+Bregion_len:-7]
                                tail.append((tail_seq, bulge_index, tar_AboxSeq, seed_index))
                        try:
                                seedScan_output[seed][target['name']] = seedScan_output[seed]
[target['name']] + [tail]
                        except KeyError:
                                if seed in seedScan_output.keys():
                                        seedScan_output[seed][target['name']] = [tail]
                                else:
                                        seedScan_output[seed] = {target['name']:[tail]}
        print('finished seedScan_v2')
        return seedScan_output
```

```
def design_A_region(Aboxes, Abox_len=3, Abox_max_len4binding = 4):
	# design optimal sequence between seed and tail
	# runs mcff on end of seed to start offset, on all possible nt combos
	# use mask and fake flanking seq to force proper structure
	# something like this: -s GGGcccccCCC -zd GGGcccCCC -alt -vv -m '(((xxxxx(((  )))xxx)))'
	# to prodive a minimal score? Not needed I think. Just take worst mfe (highest).
	# returns Abox in 3->5 direction to match consensus output and make manual hybridization
easier
	print('generating Abox')
	Abox = ''
	data = []
	target_seqs = []
	for Abox_d in Aboxes:
		target_seqs.append(list(Abox_d.values())[0].upper().replace('T','U'))
	for Abox_seq in itertools.product('ACGU', repeat=Abox_len):
		Abox_seq = ''.join(Abox_seq)
		mfes = []
		dic = {'Abox_seq':Abox_seq}
		for Abox_d in Aboxes:
			target = list(Abox_d.keys())[0]
			tar_bulge_seq = list(Abox_d.values())[0].upper().replace('T','U')
			dic[target+'_ROI'] = tar_bulge_seq
			mask = '((('+'x'*len(tar_bulge_seq)+'(((  )))'+'x'*Abox_len+')))'
			cmd = mcffCommandLine_Abox('GGG'+tar_bulge_seq+'CCC', 'GGG'+Abox_seq+'CCC',
mask) #;print(cmd);sys.exit()
			mcff_output = str(subprocess.check_output(cmd ,shell=True)) ;print(cmd,'\
n'+mcff_output)
			mfe = float(mcff_output.split('mfe(')[1].split(')')[0])
			dic[target+'_mfe'] = mfe
			mfes.append(mfe)
		dic['average_mfe'] = stats.mean(mfes) ; dic['sd_mfe'] = stats.stdev(mfes)
		data.append(dic)
	Abox_df = pd.DataFrame(data)
	# design for full non-binding Abox
	if Abox_max_len4binding <= min([len(list(x.values())[0]) for x in Aboxes]):
		print('designing for full non-binding')
		Abox = Abox_df.iloc[Abox_df['average_mfe'].argmax()]['Abox_seq']
		mfe = Abox_df.iloc[Abox_df['average_mfe'].argmax()]['average_mfe']
		sd = Abox_df.iloc[Abox_df['average_mfe'].argmax()]['sd_mfe']
	else:
		samples_to_bind_to = []
		samples_to_not_bind = []
		for Abox_d in Aboxes:
			if len(list(Abox_d.values())[0]) <= Abox_max_len4binding:
				samples_to_bind_to.append(list(Abox_d.keys())[0])
			else:
				samples_to_not_bind.append(list(Abox_d.keys())[0])
		print(samples_to_bind_to, '\n', samples_to_not_bind)
		# design for mixture of binding and non-binding
		if len(samples_to_not_bind) > 0:
			print('designing for mixture of binding/non-binding')
			df2 = Abox_df ; df2['average_mfe_binders'] = df2[[x+'_mfe' for x in
samples_to_bind_to]].mean(axis=1)
			df2['average_mfe_nonBinders'] = df2[[x+'_mfe' for x in
samples_to_not_bind]].mean(axis=1)
			df3 = df2.loc[(df2.average_mfe_binders ==
df2.iloc[df2['average_mfe_binders'].argmin()].average_mfe_binders)]
			df3 = df3.loc[(df3.average_mfe_nonBinders ==
df3.iloc[df3['average_mfe_nonBinders'].argmax()].average_mfe_nonBinders)]
			df3 = df3.reset_index().iloc[0] # if there are more than one solution,
return top row
			Abox = df3['Abox_seq']
			mfe = df3['average_mfe']
			sd = df3['sd_mfe']
		# design for full binding
		else:
```

```
                        print('designing for full binding')
                        Abox = Abox_df.iloc[Abox_df['average_mfe'].argmin()]['Abox_seq']
                        mfe = Abox_df.iloc[Abox_df['average_mfe'].argmin()]['average_mfe']
                        sd = Abox_df.iloc[Abox_df['average_mfe'].argmin()]['sd_mfe']
        print('Abox design done', Abox, mfe, sd) #;sys.exit()
        return Abox[::-1], mfe, sd # returning Abox in 3->5 direction

def deconvolute_consensus(consensus):
        # given a consensus string, deconvolutes it into a list of seqs
        replacement_d = {
        's':['g','c'],
        'w':['a','u'],
        'm':['a','c'],
        'k':['g','u'],
        'r':['a','g'],
        'y':['c','u'],
        'b':['c','g','u'],
        'd':['a','g','u'],
        'h':['a','c','u'],
        'v':['a','c','g'],
        'n':['a','g','c','u']
        }
        consensus_l = list(consensus)
        temp = []
        for c in consensus_l:
                if c in replacement_d.keys():
                        temp.append(replacement_d[c])
                else:
                        temp.append(c)
        temp2 = []
        for t in itertools.product(*temp):
                temp2.append(''.join(t))
        return temp2

def gen_bindingProfile(seq1, seq2):
        # given 2 strings of nts of equal length, return binding profile
        # binding annotations:
        # |: canonical pairing
        # !: GU paring
        # .: no pairing. Tried using space, found it harder to determine motif
        pairing_d = {
                'gc': '|',
                'cg': '|',
                'au': '|',
                'ua': '|',
                'gu': '!',
                'ug': '!',
        }
        # sanitize inputs
        seq1 = seq1.lower().replace('t','u')
        seq2 = seq2.lower().replace('t','u')
        # begin pairing
        output = ''
        for char_set in zip(seq1, seq2):
                char_set = ''.join(char_set)
                if char_set in pairing_d.keys():
                        output += pairing_d[char_set]
                else:
                        output += '.'
        return output

def designDregion(Bbox_seq, l_BbindingProfiles, l_DregionSeqs, DregionSeqPoss):
        # given a list of Bregion binding profiles and
        # a list of Dregion target sequences (5'->3' target orientation)
        # optimize Dregions by extending and stabilizing the weaker Bregions
        print('possible DregionSeqs', DregionSeqPoss)
        Dregion_df = pd.DataFrame(columns=['miR_Dseq',
```

```
                                                                'complete_binding_profiles',
                                                                '3primeBindingSize',
                                                                'num_unpaired'])
        atleast5bp_endregion = {}
        for miR_Dseq in DregionSeqPoss:
                if max([findLongestRepeat(Bbox_seq[::-1]+miR_Dseq,
'A'),findLongestRepeat(Bbox_seq[::-1]+miR_Dseq, 'U')]) > maxPolyU:
                        continue
                DregionsBindingProfiles = []
                for tar_index, tar_DregionSeq  in enumerate(l_DregionSeqs):
                        DregionBindingProfile = gen_bindingProfile(tar_DregionSeq, miR_Dseq)
                        DregionsBindingProfiles.append(DregionBindingProfile)
                complete_binding_profiles = [i + j for i, j in zip(DregionsBindingProfiles,
l_BbindingProfiles)]

                # just use the complete binding profiles
                island_sizes = []
                numbers_unpaired = []
                for binding_profile in complete_binding_profiles:
                        max_island_size = -1
                        for i, island in enumerate(binding_profile.split('.')[::-1]): # reversing
so we look at the 3' first
                                if len(island) >max_island_size:
                                        max_island_size = len(island)
                        numbers_unpaired.append(binding_profile.count('.'))
                        island_sizes.append(max_island_size)
                Dregion_df.loc[miR_Dseq] = [miR_Dseq, complete_binding_profiles, island_sizes,
numbers_unpaired]
        if len(Dregion_df.index) == 0: # no Dregion sequences pass polyA/U filter
                # generate NxN Dseq possibilities without those with polyA/U
                nbyn = [x for x in deconvolute_consensus('nnnnn') if
max([findLongestRepeat(Bbox_seq[::-1]+x, 'A'),findLongestRepeat(Bbox_seq[::-1]+x, 'U')]) <=
maxPolyU]
                print('no originally identified DregionSeqs pass polyU/A filter; generating nbyn
for testing\nwill test:', len(nbyn), 'possible Dregion sequences')
                for miR_Dseq in nbyn:
                        DregionsBindingProfiles = []
                        for tar_index, tar_DregionSeq  in enumerate(l_DregionSeqs):
                                DregionBindingProfile = gen_bindingProfile(tar_DregionSeq, miR_Dseq)
                                DregionsBindingProfiles.append(DregionBindingProfile)
                        complete_binding_profiles = [i + j for i, j in zip(DregionsBindingProfiles,
l_BbindingProfiles)]

                        # just use the complete binding profiles
                        island_sizes = []
                        numbers_unpaired = []
                        for binding_profile in complete_binding_profiles:
                                max_island_size = -1
                                for i, island in enumerate(binding_profile.split('.')[::-1]): #
reversing so we look at the 3' first
                                        if len(island) >max_island_size:
                                                max_island_size = len(island)
                                numbers_unpaired.append(binding_profile.count('.'))
                                island_sizes.append(max_island_size)
                        Dregion_df.loc[miR_Dseq] = [miR_Dseq, complete_binding_profiles,
island_sizes, numbers_unpaired]

        passes_bindingSize = pd.DataFrame(columns=['miR_Dseq',
                                                                'complete_binding_profiles',
                                                                '3primeBindingSize',
                                                                'num_unpaired',
                                                                'sum_unpaired'])
        for i, row in Dregion_df.iterrows():
                if min(row['3primeBindingSize']) >= 5:
                        passes_bindingSize.loc[i] = list(row) + [sum(row['num_unpaired'])]
        best_design = passes_bindingSize[passes_bindingSize.sum_unpaired ==
passes_bindingSize.sum_unpaired.min()]
```

```
        if best_design.shape[0] > 1:
                print('multiple possible best designs')
                print(best_design) #;sys.exit()
                print(best_design.shape)
                print('returning all good designs')
                return list(best_design.miR_Dseq)
        elif best_design.shape[0] == 0:
                print('no Dregion can generate 5bp of contiguous binding')
                # figure out which target is the weakest, pick the Dseq with max binding to that
region
                print(Dregion_df)
                # need to come up with list of average 3primeBindingSize to determine true weakest
                av_3primeBindingSize_l = []
                for i in range(len(l_BbindingProfiles)):
                        av = mean([ x[i] for x in Dregion_df['3primeBindingSize']])
                        av_3primeBindingSize_l.append(av)
                weakest_tar_index = min(range(len(av_3primeBindingSize_l)),
key=av_3primeBindingSize_l.__getitem__)
                print('weakest target is', weakest_tar_index)
                min_unpaired = 9999
                best_design['unpaired_of_interest'] = 'NA'
                for i, row in Dregion_df.iterrows():
                        if row['num_unpaired'][weakest_tar_index] <= min_unpaired:
                                min_unpaired = row['num_unpaired'][weakest_tar_index]
                                best_design.loc[i] = list(row) + [sum(row['num_unpaired']),
min_unpaired]
                best_design = best_design[best_design.unpaired_of_interest ==
best_design.unpaired_of_interest.min()]
                if best_design.shape[0] > 1:
                        print(best_design)
                        print('multiple possible best designs, returning all')
                        return list(best_design.miR_Dseq)
        print('only one Dregion seq minimizes unbound nts', best_design.miR_Dseq[0],
best_design.complete_binding_profiles[0]) #;sys.exit()
        return str(best_design.miR_Dseq[0])

def optimizeTailBinding(df):
        print('optimizing tail/Dbox binding')
        # given a pd df, with a seed column and target_BregionSeq and _bulgeIndex
        # designs complete tail sequence, starting with Bregion, then Dregion

        # declare lists that will become columns in final df
        # Dregion_seqs = []
        # fullDesign = []
        # gnerating columns
        scores_df['Dregion_seq'] = 'dummy'
        scores_df['fullDesign'] = 'values'
        # re-collect target seqs
        target_d = {}
        for target in targets:
                target_d[target['name']] = str(target['seq'])

        # dyamically find B/D region sequence columns, bulgeIndex columns, and target names
        seq_cols = [col for col in scores_df.columns if '_BregionSeq' in col]
        bulgeIndex_cols = [col for col in scores_df.columns if '_bulgeIndex' in col]
        target_names = [x.replace('_bulgeIndex', '') for x in bulgeIndex_cols]
        DregionSeqs_colNames = [col for col in scores_df.columns if '_DregionSeq' in col]
        AboxSeq_cols = [col for col in scores_df.columns if '_AboxSeq' in col]


        # begin design process
        counter = 1
        for index, row in df.iterrows():

        print('\nnumber:',counter,'/index',index,'////////////////////////////////////////')
                counter +=1
                bestBbox = ''
```

```
                bestBboxBindingProfile = ''
                best_total_unpaired = 9999

                BregionsSeqs = list(row[seq_cols])
                BregionDeconvoluted = deconvolute_consensus(row['Bregion_consensus'])
                print(BregionsSeqs, BregionDeconvoluted)
                if len(BregionDeconvoluted) == 1:
                        print('no Bregion optimization needed')
                        bestBbox = BregionDeconvoluted[0]
                        bindingProfiles = []
                        for tar_Bbox in BregionsSeqs:
                                bindingProfiles.append(gen_bindingProfile(tar_Bbox, bestBbox))
                        bestBboxBindingProfile = bindingProfiles
                else:
                        print('optimizing Bbox')
                        for miR_Bbox in BregionDeconvoluted:
                                if findLongestRepeat(miR_Bbox, 'a') > maxPolyU or
findLongestRepeat(miR_Bbox, 'u') > maxPolyU:
                                        print('Bbox version has too many polyA/Us,', miR_Bbox,
'dropping')
                                        continue
                                bindingProfiles = []
                                for tar_Bbox in BregionsSeqs:
                                        bindingProfiles.append(gen_bindingProfile(tar_Bbox,
miR_Bbox))
                                total_unpaired = ''.join(bindingProfiles).count('.')
                                print(bindingProfiles, 'total unpaired:',total_unpaired, miR_Bbox)
                                if total_unpaired < best_total_unpaired:
                                        best_total_unpaired = total_unpaired
                                        bestBboxBindingProfile = bindingProfiles
                                        bestBbox = miR_Bbox
                print(bestBbox, bestBboxBindingProfile, '<--- best Bbox binding profile')
                # end of Bregion optimization
                # begin optimizing Dregion
                print('now optimizing Dregion')
                mir_seq = 'a'+str(Seq(row['seed']).reverse_complement()) + row['Abox_seq'][::-
1].lower()+bestBbox[::-1].upper()
                # perform mcff based tesing. Currently done, but not needed
                Dregion_tests_fileName = 'DregionTests_mcff_'+mir_seq+'.tsv'
                if os.path.exists(Dregion_tests_fileName):
                        print('removeing previous Dregion test analysis')
                        subprocess.call(['rm', Dregion_tests_fileName])
                # use mcff to find best Dregion
                for target in target_names:
                        # collect target sequence
                        bulgeIndex = str(row[target+'_bulgeIndex'])
                        tar_Abox_col = [col for col in AboxSeq_cols if target in col][0]
                        regex = row[target+'_DregionSeq']+row[target+'_BregionSeq']
+row[tar_Abox_col]+row['seed']+'.'
                        print('looking for', regex) #; sys.exit()
                        DregionDeconvoluted = deconvolute_consensus(row['Dregion_consensus'])
                # end of mcff method

                # start of extend Bregion binding method
                # description: check Bbox binding profiles, design Dregion to extend weaker Bboxes

                DregionSeqs = list(row[DregionSeqs_colNames])
                Dregion_design = designDregion(bestBbox, bestBboxBindingProfile, DregionSeqs,
DregionDeconvoluted) #;stopPoint=input()
                if type(Dregion_design) == type('string'): # ie Dregion design function returns a
single design
                        full_seq = str(mir_seq+Dregion_design[::-1].lower()) #; print(full_seq) #;
sys.exit()
                        scores_df.loc[index, 'Dregion_seq'] = Dregion_design[::-1].lower()
                        scores_df.loc[index, 'fullDesign'] = full_seq
                else: # ie if Dregion design function returns multiple designs
                        # expects lists of DregionSeqs
```

```
                        # update existing row, then add new rows to end of df
                        scores_df.loc[index, 'Dregion_seq'] = Dregion_design[0][::-1].lower()
                        scores_df.loc[index, 'fullDesign'] = str(mir_seq+Dregion_design[0][::-
1].lower())
                        for seq in Dregion_design[1:]:
                                full_seq = str(mir_seq+seq[::-1].lower())
                                temp_row = row
                                temp_row['Dregion_seq'] = seq[::-1].lower()
                                temp_row['fullDesign'] = full_seq
                                scores_df.loc[len(scores_df)] = list(temp_row)
        return scores_df

def determineSeedType(df):
        #       determine seed type
        # seed types:
        #        stringent:
        #                8mer: binding of miR nts 2-8 (7nt hybridizing) AND 3' nt of target is A
        #                7mer-A1: binding of miR nts 2-7 (6nt hybridizing) AND 3' nt of target is A
        #                7mer-m8: binding of miR nts 2-8 (7nt hybridizing)
        #                relative strengths: 8mer > 7mer-m8 > 7mer-A1
        #        non stringent:
        #                many kinds, not going to investigate.

        print('determining seed types')
        # given a pd df, with a seed column and target_BregionSeq and _bulgeIndex
        # re-collect target seqs
        target_d = {}
        for target in targets:
                target_d[target['name']] = str(target['seq'])

        # dyamically find sequence columns, bulgeIndex columns, and target names
        seq_cols = [col for col in df.columns if '_BregionSeq' in col]
        bulgeIndex_cols = [col for col in scores_df.columns if '_bulgeIndex' in col]
        target_names = [x.replace('_bulgeIndex', '') for x in bulgeIndex_cols]
        AboxSeq_cols = [col for col in scores_df.columns if '_AboxSeq' in col]
        seed_index_cols = [col for col in scores_df.columns if '_seed_index' in col]

        for target in target_names:
                temp = target+'_seedType'
                df[temp] = 'strawberry' # dummy value, I like strawberries
        for index, row, in df.iterrows():
                seed = str(row['seed'])
                seedTypes = []
                for target in target_names:
                        seq_col = [col for col in seq_cols if target in col][0]
                        bindingSite = str(row.loc[seq_col])
                        bulgeIndex_col = [col for col in bulgeIndex_cols if target in col][0]
                        bulgeIndex = str(row[bulgeIndex_col])
                        tar_Abox_col = [col for col in AboxSeq_cols if target in col][0]
                        tar_Abox_seq = str(row[tar_Abox_col]) #;print(tar_Abox_seq); sys.exit()
                        regex = bindingSite+tar_Abox_seq+seed #;print(regex) #; sys.exit()
                        seed_index_col = [col for col in seed_index_cols if target in col][0]
                        seed_index = row[seed_index_col]
                        m = re.findall(regex, target_d[target][seed_index-50:seed_index+1],
overlapped=True)
                        orf_pos1=target_d[target][seed_index].upper()
                        if orf_pos1 == 'A' and len(seed) == 7:
                                seedType = '8mer'
                        elif not orf_pos1 == 'A' and len(seed) == 7:
                                seedType = '7mer-m8'
                        elif orf_pos1 == 'A' and len(seed) == 6:
                                seedType = '7mer-A1'
                        else:
                                seedType = 'NA'
                        scores_df.at[index, target+'_seedType'] = seedType
        print('finished determining seed types')
        return df
```

```
# returns mir-based index
def build_hybridization_index(db, mcff_string):
	print('building hybridization index')
	print(db, mcff_string)
	tar_len = len(db.split('  ')[0])
	tar_seq = mcff_string[:tar_len].lower()
	mir_seq = mcff_string[tar_len+2:].lower()
	target_db = db.split('  ')[0]  ; mir_db = db.split('  ')[1]
	target_indeces = []
	mir_indeces = []
	for i, c in enumerate(target_db[::-1]):
		if c in '(<{': #keeping curly in case
			target_indeces.append(i)
	for i, c in enumerate(mir_db):
		if c in ')>}':
			mir_indeces.append(i)
	tar_index = {}
	rev_tar = tar_seq[::-1]
	for pair in zip(target_indeces, mir_indeces):
		tar_index[pair[0]] = (pair[1], rev_tar[pair[0]]+mir_seq[pair[1]])
	mir_index = {}
	for pair in zip(mir_indeces, target_indeces):
		mir_index[pair[0]] = (pair[1],mir_seq[pair[0]]+rev_tar[pair[1]])
	return tar_index, mir_index

# builds double stranded structure from mir sequence, target seq(3'->5'), and mir-based
hybridization index
def printDB(mirna, target, Dbps):
	# takes duplex binding information from Yaml record and prints it as string
	printer_annotation = {'gc': '|', 'cg':'|', 'au':'|', 'ua':'|', 'at':'|','ta':'|',
'gu':'!', 'ug':'!','gt':'!', 'tg':'!'}
	targetPrint = ''
	miRNAPrint = ''
	bindingPrint = ''
	prev_mir_index = 0
	prev_tar_index = 0
	switch = 0
	for k in Dbps.keys():
		if switch == 1 and abs(prev_mir_index -k) > 1 and abs(prev_tar_index -Dbps[k]) > 1:
			mir_loop_size = k - prev_mir_index -1
			tar_loop_size = Dbps[k] - prev_tar_index -1
			# minus 1 to account for bouned ends of loop
			targetPrint = '-'*(mir_loop_size - tar_loop_size)
+target[prev_tar_index+1:Dbps[k]][::-1] + targetPrint
			#adding on loop structure
			try:
				targetPrint = target[Dbps[k]] + targetPrint
			#adding next bound element after loop
			except IndexError:
				# caused by end of target string being used
				targetPrint = target[-1] + targetPrint
			miRNAPrint = '-'*(tar_loop_size - mir_loop_size)+mirna[prev_mir_index+1:k]
[::-1] + miRNAPrint
			miRNAPrint = mirna[k]+miRNAPrint
			pair = targetPrint[0].lower()+miRNAPrint[0].lower()
			bindingPrint = ' '*max([mir_loop_size, tar_loop_size]) + bindingPrint
			try:
				bindingPrint = printer_annotation[pair.lower()] + bindingPrint
			except KeyError: # ie non-canonical
				bindingPrint = ':'+ bindingPrint
		elif switch == 1 and abs(prev_mir_index - k) > 1: # if there's a gap on mir side
			targetPrint = '-'*abs(prev_mir_index - k+1)+targetPrint
			miRNAPrint = mirna[prev_mir_index+1:k+1][::-1]   + miRNAPrint
			targetPrint = target[Dbps[k]] + targetPrint
			bindingPrint = ' '*abs(prev_mir_index - k+1)+bindingPrint
			pair = target[Dbps[k]].lower()+mirna[k].lower()
```

```
                        try:
                                bindingPrint = printer_annotation[pair.lower()] + bindingPrint
                        except KeyError: # ie non-canonical
                                bindingPrint = ':'+ bindingPrint
                elif switch == 1 and abs(prev_tar_index - Dbps[k]) > 1: # if there's a gap on
target side
                        targetPrint = target[Dbps[k] : prev_tar_index] + targetPrint
                        miRNAPrint = '-'*abs(prev_tar_index - Dbps[k]+1)+miRNAPrint
                        bindingPrint = ' '*abs(prev_tar_index - Dbps[k]+1)+bindingPrint
                        temp = target[prev_tar_index+1:Dbps[k]+1][::-1]
                        targetPrint = temp + targetPrint
                        miRNAPrint = mirna[k]+miRNAPrint
                        pair = target[Dbps[k]].lower()+mirna[k].lower()
                        try:
                                bindingPrint = printer_annotation[pair.lower()] + bindingPrint
                        except KeyError: # ie non-canonical
                                bindingPrint = ':'+ bindingPrint
                else:
                        targetPrint = target[Dbps[k]] + targetPrint
                        miRNAPrint = mirna[k]+miRNAPrint
                        pair = target[Dbps[k]].lower()+mirna[k].lower()
                        try:
                                bindingPrint = printer_annotation[pair.lower()] + bindingPrint
                        except KeyError: # ie non-canonical
                                bindingPrint = ':'+ bindingPrint
                switch = 1
                prev_tar_index = Dbps[k]
                prev_mir_index = k
        # main construction complete
        # now checking for dangling bases on mir side
        if not len(mirna) == len(miRNAPrint.replace('-', '')):
                dangling_size = len(mirna) - len(miRNAPrint.replace('-', ''))
                targetPrint = '-'*dangling_size+targetPrint
                bindingPrint = ' '*dangling_size+bindingPrint
                miRNAPrint = mirna[-dangling_size:][::-1]+miRNAPrint
        print(targetPrint, '\t<-----targetPrint\n'+bindingPrint, '\n'+ miRNAPrint, '\t<-----
miRNAPrint')
        return targetPrint, bindingPrint, miRNAPrint

def pprint_to_f(scores_df, out='pprint.txt'):
        target_names = [x.replace('_bulgeIndex', '') for x in bulgeIndex_cols]
        with open(out, 'w') as f:
                for index, row in scores_df.iterrows():
                        for target in target_names:
                                f.write(row[target+'_pprint']+'\n')
                        f.write('################################################################\
n')
        print('written pprint to file', out)

def combinationsExplorer(variable_tuple):
        print('started combinationsExplorer', mp.current_process().name)
        print(variable_tuple) ;sys.exit()
        # tail_combo = variable_tuple[0]
        # bulge_combo = variable_tuple[1]
        # Abox_combo = variable_tuple[2]
        # seed_index_combo = variable_tuple[3]
        score, num_poss, consensus =
hybridiation_consensus_generator_v3_Bartel2022(variable_tuple[0])
        row = (seed, *variable_tuple[0], *variable_tuple[1], score, num_poss, consensus,
*variable_tuple[2], *variable_tuple[3])
        return score, row

def combinationsExplorer_v2(iteratorISlice):
        explored = []
        local_max_score = 0
        counter = 0
        for tail_combo, bulge_combo, Abox_combo, seed_index_combo in iteratorISlice:
```

```
                counter+=1
                score, num_poss, consensus =
hybridiation_consensus_generator_v3_Bartel2022(tail_combo)
                if max([findLongestRepeat(consensus, 'A'), findLongestRepeat(consensus, 'U')]) <=
maxPolyU and score >= local_max_score*BscoreThresh:
                        row = (seed, *tail_combo, *bulge_combo, score, num_poss, consensus,
*Abox_combo, *seed_index_combo)
                        if score > local_max_score:
                                local_max_score = score
                                if len(scores) ==0:
                                        explored.append(row)
                                else:
                                        indeces_to_drop = [i for i, x in enumerate(explored) if
x[score_index] < local_max_score*BscoreThresh]
                                        for i in sorted(indeces_to_drop, reverse=True):
                                                del explored[i]
                                        explored.append(row)

                        elif score >= local_max_score*BscoreThresh:
                                explored.append(row)
        return local_max_score, explored


# define variables for run
# sys.args can be used

if __name__ == '__main__':
        if 'tars' in [x.split('=')[0] for x in sys.argv]:
                for i,y in enumerate([x.split('=')[0] for x in sys.argv]):
                        if 'tars' in y:
                                target_file = sys.argv[i].split('=')[1]
                                print('target file is', target_file)
        else:
                target_file = 'pNL-HIV.fa'
                print('no target file specified, defaulting to', target_file)

        #target_file = 'pNL-HIV.fa'

        if 'Bregion_start_offset' in [x.split('=')[0] for x in sys.argv]:
                for i,y in enumerate([x.split('=')[0] for x in sys.argv]):
                        if 'Bregion_start_offset' in y:
                                start_offset = int(sys.argv[i].split('=')[1])
                                print('start offset is', start_offset)
        else:
                start_offset=2
                print('no Bregion binding site offset detected, defaulting to 1nt from seed')

        if 'max_tar_bulge' in [x.split('=')[0] for x in sys.argv]:
                for i,y in enumerate([x.split('=')[0] for x in sys.argv]):
                        if 'max_tar_bulge' in y:
                                bulge_size = int(sys.argv[i].split('=')[1])
                                print('bridge set to', bulge_size)
        else:
                bulge_size=7
                print('no max bridge specified, defaulting to',bulge_size,'nt from seed')

        if 'local_tar_seq_len_extension' in [x.split('=')[0] for x in sys.argv]:
                for i,y in enumerate([x.split('=')[0] for x in sys.argv]):
                        if 'local_tar_seq_len_extension' in y:
                                local_tar_seq_len_extension = int(sys.argv[i].split('=')[1])
                                print('local_tar_seq_len_extension is', local_tar_seq_len_extension)
        else:
                local_tar_seq_len_extension = 5
                print('no local target sequence length extension for folding specified, defaulting
to:', local_tar_seq_len_extension)

        if 'single_tar_sites_allowed' in [x.split('=')[0] for x in sys.argv]:
```

```
                for i,y in enumerate([x.split('=')[0] for x in sys.argv]):
                        if 'single_tar_sites_allowed' in y:
                                single_tar_sites_allowed = int(sys.argv[i].split('=')[1])
                                print('single target mode number of sites per design',
single_tar_sites_allowed)
        else:
                single_tar_sites_allowed=3
                print('no number of single target sites per design specified, defaulting to:',
single_tar_sites_allowed)

        if 'maxPolyU' in [x.split('=')[0] for x in sys.argv]:
                for i,y in enumerate([x.split('=')[0] for x in sys.argv]):
                        if 'maxPolyU' in y:
                                maxPolyU = int(sys.argv[i].split('=')[1])
                                print('maximum polyU stretch allowed:', single_tar_sites_allowed)
        else:
                maxPolyU=3
                print('no maximum polyU stretch detected, defaulting to', maxPolyU)

        if 'BscoreThresh' in [x.split('=')[0] for x in sys.argv]:
                for i,y in enumerate([x.split('=')[0] for x in sys.argv]):
                        if 'BscoreThresh' in y:
                                BscoreThresh = float(sys.argv[i].split('=')[1])
                                print('Bregion consensus score threshold:', BscoreThresh)
        else:
                BscoreThresh= 0.9
                print('no Bregion consensus score threshold detected, defaulting to', BscoreThresh)

        if 'poolsize' in [x.split('=')[0] for x in sys.argv]:
                for i,y in enumerate([x.split('=')[0] for x in sys.argv]):
                        if 'poolsize' in y:
                                pool_size = int(sys.argv[i].split('=')[1])
                                print('pool_size set to:', pool_size)
        else:
                pool_size = 10
                print('no pool size detected, defaulting to', pool_size)

        if 'regions' in [x.split('=')[0] for x in sys.argv]:
                for i,y in enumerate([x.split('=')[0] for x in sys.argv]):
                        if 'regions' in y:
                                Bregion_start, Bregion_end, synmiR_len = [int(i) for i in
sys.argv[i].split('=')[1].split(',')]
                                Bregion_len = Bregion_end-Bregion_start +1
                                print('Bregion_start, Bregion_end, synmiR_len:', Bregion_start,
Bregion_end, synmiR_len)
                                print('Bregion_len set to:', Bregion_len)
        else:
                Bregion_start = 12; Bregion_end= 16; Bregion_len = 5 ; synmiR_len = 21
                print('no regions detected, defaulting to:')
                print('Bregion_start', Bregion_start)
                print('Bregion_end:', Bregion_end)
                print('Bregion_len:', Bregion_len)
                print('synmiR_len:', synmiR_len)

        ###### set parameters here for testing #####

        # start_offset = 2
        # bulge_size = 12
        # local_tar_seq_len_extension = 10
        # single_tar_sites_allowed = 3

        ##########################################

        # build root for output files
        if len(target_file.split('.')) ==2:
                out_fis_root = target_file.split('.')[0]
        elif len(target_file.split('.')) ==1:
```

```
                out_fis_root = target_file
        else:
                out_fis_root = '.'.join(target_file.split('.')[:-1])
        out_fis_root += '_startOffset='+str(start_offset)+\
                '_bridgeLen='+str(bulge_size)+\
                '_foldingExtnesion='+str(local_tar_seq_len_extension)+\
                '_singleTarSites='+str(single_tar_sites_allowed)+\
                '_maxPolyU='+str(maxPolyU)+\
                '_BscoreThresh='+str(BscoreThresh)+\
                '_regions='+str(Bregion_start)+','+str(Bregion_end)+','+str(synmiR_len)
        print(out_fis_root)

        targets = []
        for record in SeqIO.parse(target_file, "fasta"):
                targets.append({'name':record.id, 'seq':record.seq})
        print(targets)

        seed_tail_dict_v2 = seedScan_Bartel(targets, start_offset=start_offset,
bulge_size=bulge_size, Bregion_len=Bregion_len)

        # serialize dictionary to examine
        # import json
        # with open(out_fis_root+'_seed_tail_dict_v2', 'w') as f:
        #       f.write(json.dumps(seed_tail_dict_v2))
        # print('dict_serialized')

        #### determine how many sites to examine ####
        counter = 0
        hits_all_tars = {}


        #multi-target branch
        if len(targets) > 1:
                for k in seed_tail_dict_v2.keys():
                        if len(seed_tail_dict_v2[k].keys()) == len(targets):
                                counter += 1
                                hits_all_tars[k] = seed_tail_dict_v2[k]
        #single target branch
        else:
                size_dict = {}
                for k in seed_tail_dict_v2.keys():
                        for i in seed_tail_dict_v2[k].keys():
                                try:
                                        size_dict[len(seed_tail_dict_v2[k][i])] =
size_dict[len(seed_tail_dict_v2[k][i])] +1
                                except KeyError:
                                        size_dict[len(seed_tail_dict_v2[k][i])] = 1
                                if len(seed_tail_dict_v2[k][i]) == single_tar_sites_allowed:
                                        hits_all_tars[k] = seed_tail_dict_v2[k]
                                        counter += 1
                for k in size_dict.keys():
                        print(k, size_dict[k])

                ### code for generating seed_seq : site_num lists ###

                # with open('pNL-HIV_rc_complete_seed_list.txt', 'w') as f:
                #       for k in seed_tail_dict_v2.keys():
                #               for i in seed_tail_dict_v2[k].keys():
                #                       f.write(k+': '+str(len(seed_tail_dict_v2[k][i]))+'
'+','.join(str(x[0][-1]) for x in seed_tail_dict_v2[k][i])+'\n')
                # with open('3_hitting_seeds.txt', 'w') as f:
                #       for k in seed_tail_dict_v2.keys():
                #               for i in seed_tail_dict_v2[k].keys():
                #                       if len(seed_tail_dict_v2[k][i]) == 3:
                #                               f.write(k+': '+str(len(seed_tail_dict_v2[k][i]))+'\
n')
                # sys.exit()
```

```
	print(str(counter)+'/'+str(len(seed_tail_dict_v2.keys()))+' seed are multi-targeting')
	# remove seed:tail dictionary now that it's been parsed to new variable
	seed_tail_dict_v2 = None
	if counter == 0:
		print('no multi-targeting sites identified') ; sys.exit()

	#### FINISHED DETERMINING HOW MANY SITES TO EXAMINE ####
	scores = []
	target_order = [x['name'] for x in targets]

	### BEGIN EXPLORING COMBINATIONS AND CAST DATAFRAME ###
	max_score = -1

	# multi-target branch
	if len(targets) > 1:
		print('multi target mode detected and excecuting')
		print('seeds to build for:', len(hits_all_tars.keys()))
		seed_combination_space_size = (bulge_size-start_offset+1)**single_tar_sites_allowed
		combination_space_size =len(hits_all_tars.keys())*(bulge_size-
start_offset+1)**single_tar_sites_allowed
		print('single seed combination space:', seed_combination_space_size)
		print('total combination space', combination_space_size)
		header = ['seed', *[x+'_BregionSeq' for x in target_order],
                                        *[x+'_bulgeIndex' for x in target_order],
                                        'Bregion_score', 'Bregion_num_possibilities',
'Bregion_consensus',
                                        *[x+'_AboxSeq' for x in target_order],
                                        *[x+'_seed_index' for x in target_order]]
		score_index = header.index('Bregion_score') ;print('score_index', score_index)
		counter =0 ;times = []
		for seed in hits_all_tars.keys():
			tails = []
			bulges = []
			Aboxes = []
			seed_indexes = []
			for tar in target_order:
				tails.append((x[0] for x in itertools.chain(*hits_all_tars[seed]
[tar])))
				bulges.append((x[1] for x in itertools.chain(*hits_all_tars[seed]
[tar])))
				Aboxes.append((x[2] for x in itertools.chain(*hits_all_tars[seed]
[tar])))
				seed_indexes.append((x[3] for x in
itertools.chain(*hits_all_tars[seed][tar])))
			tail_combos = itertools.product(*tails)
			bulge_combos = itertools.product(*bulges)
			Abox_combos = itertools.product(*Aboxes)
			seed_index_combos = itertools.product(*seed_indexes)
			masterIterator = zip(tail_combos, bulge_combos, Abox_combos,
seed_index_combos)

			start = 0
			iteratorChunks = []
			for i in range(int(seed_combination_space_size/pool_size),
seed_combination_space_size, int(seed_combination_space_size/pool_size)):
				iteratorChunks.append(itertools.islice(masterIterator, start, i))
				start = i
			iteratorChunks.append(itertools.islice(masterIterator, i,
seed_combination_space_size))
			t0 = time.time()
	### start of multiprocessing
			with mp.Pool(processes = pool_size) as pool:
				for score_found, explored in pool.imap(combinationsExplorer_v2,
iteratorChunks, chunksize=1):
                                        if score_found > max_score:
                                                max_score = score_found
```

```
                                        scores += explored
                                        indeces_to_drop = [i for i, x in enumerate(scores) if
x[score_index] < max_score*BscoreThresh]
                                        for i in sorted(indeces_to_drop, reverse=True):
                                                del scores[i]
                                elif score_found >= max_score*BscoreThresh:
                                        scores += explored
                        indeces_to_drop = [i for i, x in enumerate(scores) if x[score_index] <
max_score*BscoreThresh]
                        for i in sorted(indeces_to_drop, reverse=True):
                                del scores[i]
                        print(seed, time.time()-t0) ; times.append(time.time()-t0)  #; sys.exit()

        ### end of multiprocessing
                print('finished constructing scores variable')
                scores_df = pd.DataFrame(scores, columns = header)
                scores_df = scores_df.sort_values(by=['Bregion_score',
'Bregion_num_possibilities'], ascending=False)
        # single target branch
        else:
                print('single target mode detected and excecuting')
                print('seeds to build for:', len(hits_all_tars.keys()))
                seed_combination_space_size = (bulge_size-start_offset+1)**single_tar_sites_allowed
                combination_space_size =len(hits_all_tars.keys())*(bulge_size-
start_offset+1)**single_tar_sites_allowed
                print('single seed combination space:', seed_combination_space_size)
                print('total combination space', combination_space_size)
                site_order = [str(x+1) for x in list(range(single_tar_sites_allowed))]
                target_name = targets[0]['name']
                header=['seed', *[target_name+'_site'+x+'_BregionSeq' for x in site_order],

*[target_name+'_site'+x+'_bulgeIndex' for x in site_order],
                                                                        'Bregion_score',
'Bregion_num_possibilities', 'Bregion_consensus',

*[target_name+'_site'+x+'_AboxSeq' for x in site_order],

*[target_name+'_site'+x+'_seed_index' for x in site_order]]
                score_index = header.index('Bregion_score') ;print('score_index', score_index)
                counter =0 ;times = []
                for seed in hits_all_tars.keys():
                        if max([findLongestRepeat(seed, 'A'), findLongestRepeat(seed, 'U')]) >
maxPolyU:
                                print(seed, 'failed polyA/U filter, skipping')
                                continue

                        tails = []
                        bulges = []
                        Aboxes = []
                        seed_indexes = []

                        # generate needed generators
                        for tar in hits_all_tars[seed]:
                                for site in [int(x)-1 for x in site_order]: # convert x back to list
index
                                        tails.append((x[0] for x in hits_all_tars[seed][tar][site]))
                                        bulges.append((x[1] for x in hits_all_tars[seed][tar][site]))
                                        Aboxes.append((x[2] for x in hits_all_tars[seed][tar][site]))
                                        seed_indexes.append((x[3] for x in hits_all_tars[seed][tar]
[site]))
                        tail_combos = itertools.product(*tails)
                        bulge_combos = itertools.product(*bulges)
                        Abox_combos = itertools.product(*Aboxes)
                        seed_index_combos = itertools.product(*seed_indexes)
                        masterIterator = zip(tail_combos, bulge_combos, Abox_combos,
seed_index_combos)
```

```
                        start = 0
                        iteratorChunks = []
                        for i in range(int(seed_combination_space_size/pool_size),
seed_combination_space_size, int(seed_combination_space_size/pool_size)):
                                iteratorChunks.append(itertools.islice(masterIterator, start, i))
                                start = i
                        iteratorChunks.append(itertools.islice(masterIterator, i,
seed_combination_space_size))
                        t0 = time.time()
        # ### start of multiprocessing
                        with mp.Pool(processes = pool_size) as pool:
                                for score_found, explored in pool.imap(combinationsExplorer_v2,
iteratorChunks, chunksize=1):
                                        if score_found > max_score:
                                                max_score = score_found
                                                scores += explored
                                                indeces_to_drop = [i for i, x in enumerate(scores) if
x[score_index] < max_score*BscoreThresh]
                                                for i in sorted(indeces_to_drop, reverse=True):
                                                        del scores[i]
                                        elif score_found >= max_score*BscoreThresh:
                                                scores += explored
                        indeces_to_drop = [i for i, x in enumerate(scores) if x[score_index] <
max_score*BscoreThresh]
                        for i in sorted(indeces_to_drop, reverse=True):
                                del scores[i]
        ### end of mutliprocessing
                        print(seed, time.time()-t0) ; times.append(time.time()-t0)  #; sys.exit()
                print(sum(times)/len(times), '<-average time per seed') #;sys.exit()
                print('max_score determined to be', max_score);print('total combinations
explored:', counter) #;sys.exit()
                print('finished constructing scores variable', len(scores)) #; sys.exit()
                scores_df = pd.DataFrame(scores, columns = header)
                print('df casted')
                scores_df = scores_df.sort_values(by=['Bregion_score',
'Bregion_num_possibilities'], ascending=False)
                for new_name in [target_name+'_site'+x for x in site_order]:
                        targets.append({'name':new_name,'seq':targets[0]['seq']})

        scores = None # remove origianl variable; we'll be working with the df
        scores_df.to_csv(out_fis_root+'_scores_v3.csv', sep=',')
        print(scores_df)

        # add Dbox seq (nt 17-21) for each target to main df, and generate consensus

        seq_cols = [col for col in scores_df.columns if '_BregionSeq' in col]
        bulgeIndex_cols = [col for col in scores_df.columns if '_bulgeIndex' in col]
        target_names = [x.replace('_bulgeIndex', '') for x in bulgeIndex_cols]
        AboxSeq_cols = [col for col in scores_df.columns if '_AboxSeq' in col]

        Dregion_len = synmiR_len - Bregion_end # need to control for Dregion_len = 0 edge case
        print('designing for Dregion; length=', Dregion_len)
        scores_df['Dregion_consensus'] = 'nnnn'
        scores_df['Dregion_score'] = -1
        scores_df['Dregion_num_poss'] = -1
        for x in target_names:
                scores_df[x+'_DregionSeq'] = 'nnnnn'

        target_d = {}
        for target in targets:
                target_d[target['name']] = str(target['seq'])
        for index, row in scores_df.iterrows():
                seed = str(row['seed'])
                Dboxes = []
                for target in target_names:
                        seq_col = [col for col in seq_cols if target in col][0]
                        bindingSite = str(row.loc[seq_col])
```

```
                        bulgeIndex_col = [col for col in bulgeIndex_cols if target in col][0]
                        bulgeIndex = str(row[bulgeIndex_col])
                        tar_Abox_col = [col for col in AboxSeq_cols if target in col][0] #;
print(tar_Abox_col) ; sys.exit()
                        tar_Abox_seq = str(row.loc[tar_Abox_col])
                        regex = bindingSite+tar_Abox_seq+seed
                        local_tar_seq = target_d[target][row[target+'_seed_index'] -
len(regex):row[target+'_seed_index']].upper() #;sys.exit()

                        m = re.findall(regex, local_tar_seq, overlapped=True)
                        if len(m) > 1:
                                print('we got a problem: can\'t find original target sequence') ;
sys.exit()
                        for match in m:
                                Dbox = target_d[target][target_d[target].upper().index(match)-
Dregion_len:target_d[target].upper().index(match)]
                                Dboxes.append(Dbox)
                        scores_df.at[index, target+'_DregionSeq'] = Dbox
                score, num_poss, consensus = hybridiation_consensus_generator_v2_multi(Dboxes)
                scores_df.at[index, 'Dregion_consensus'] = consensus
                scores_df.at[index, 'Dregion_score'] = score
                scores_df.at[index, 'Dregion_num_poss'] = num_poss
        target_d = {}
        for target in targets:
                target_d[target['name']] = str(target['seq'])

        # dyamically find sequence columns, bulgeIndex columns, tar_ABoxSeq, and target names
        seq_cols = [col for col in scores_df.columns if '_BregionSeq' in col]
        bulgeIndex_cols = [col for col in scores_df.columns if '_bulgeIndex' in col]
        target_names = [x.replace('_bulgeIndex', '') for x in bulgeIndex_cols]
        AboxSeq_cols = [col for col in scores_df.columns if '_AboxSeq' in col]

        # add columns to df that will be filled in
        scores_df['Abox_seq'] = 'nnn'
        scores_df['Abox_mfe'] = -1.0
        scores_df['Abox_sd'] = -1.0

        counter = 0
        for index, row in scores_df.iterrows():
                seed = str(row['seed'])
                Aboxes = []
                for target in target_names:
                        tar_Abox_col = [col for col in AboxSeq_cols if target in col][0]
                        Abox = str(row.loc[tar_Abox_col]) #;print(Abox);sys.exit()
                        Aboxes.append({target:Abox})
                Abox, mfe, sd = design_A_region(Aboxes, Abox_len=Bregion_start-8-1)
                scores_df.at[index, 'Abox_seq'] = Abox
                scores_df.at[index, 'Abox_mfe'] = mfe
                scores_df.at[index, 'Abox_sd'] = sd
                counter += 1 ; print(counter)

        scores_df = optimizeTailBinding(scores_df)

        print('shape before polyU/A removal', scores_df.shape)
        scores_df = remove_polyU_polyA(scores_df, 'fullDesign')
        print('shape after polyU/A removal', scores_df.shape)

        print('reseting index')
        scores_df = scores_df.sort_values(by=['seed','Bregion_score'], ascending=False)
        scores_df = scores_df.reset_index(drop=True)

        scores_df = determineSeedType(scores_df)

        target_names = [x.replace('_bulgeIndex', '') for x in bulgeIndex_cols]
        for name in target_names:
                scores_df[name+'_localSeq'] = 'NNNNNNNNNN'
                scores_df[name+'_pprint'] = 'xxxxxxxxxx'
```

```
                scores_df[name+'_mfe'] = 1.0
        # dyamically find B/D region sequence columns, bulgeIndex columns, and target names
        Bseq_cols = [col for col in scores_df.columns if '_BregionSeq' in col]
        bulgeIndex_cols = [col for col in scores_df.columns if '_bulgeIndex' in col]
        DregionSeqs_colNames = [col for col in scores_df.columns if '_DregionSeq' in col]
        AboxSeq_cols = [col for col in scores_df.columns if '_AboxSeq' in col]
        for index, row in scores_df.iterrows():
                for target in target_names:
                        seed_index = row[target+'_seed_index']
                        local_tarSeq_len = 1+len(row['seed'])+row[target+'_bulgeIndex']
+len(row[target+'_BregionSeq'])+len(row[target+'_DregionSeq'])+local_tar_seq_len_extension
                        if seed_index > local_tarSeq_len and seed_index+1 < len(target_d[target]):
                                local_tarSeq = target_d[target][seed_index-
local_tarSeq_len:seed_index+1]
                        elif seed_index < local_tarSeq_len:
                                local_tarSeq = target_d[target][:seed_index]
                        else:
                                local_tarSeq = target_d[target][-local_tarSeq_len:]
                        local_tarSeq = local_tarSeq.lower().replace('t','u')
                        cmd = mcffCommandLine_v2(local_tarSeq, row['fullDesign'], explore='2')#
reducing explore since we only want mfe
                        mcff_output = subprocess.check_output(cmd ,shell=True).decode('utf-
8').strip()
                        mfe = float(mcff_output.split('mfe(')[1].split(')')[0])
                        mcff_l = mcff_output.split('\n')[2:]
                        mcff_fold_string = mcff_l[0]
                        dbs =  [x.split(' -')[0] for x in mcff_l[2:]] #;print(dbs)
                        energies = [float('-'+x.split(' -')[1]) for x in mcff_l[2:]]
                        mcff_df = pd.DataFrame()
                        mcff_df['dotBracket'] = dbs ; mcff_df['dG'] = energies ; mcff_df =
mcff_df.sort_values(by=['dG'], ascending=True)
                        mfe_db = mcff_df.iloc[0,0]
                        tar_index, mir_index = build_hybridization_index(mfe_db, mcff_fold_string)
                        mir_index_v2 = {}
                        for k in mir_index.keys():
                                mir_index_v2[k] = mir_index[k][0]
                        tar_pprint, binding_profile, mir_pprint =
printDB(row['fullDesign'].replace('T','U').replace('t','u'), local_tarSeq[::-1], mir_index_v2)
                        full_print = tar_pprint+'   '+target+'  '+str(mfe)+'
'+str(row[target+'_seed_index'])+'\n'+binding_profile+'\n'+ mir_pprint + '   synmiR-'+str(index)
                        scores_df.loc[index, target+'_pprint'] = full_print
                        scores_df.loc[index,target+'_localSeq'] = tar_pprint.replace('-','')
                        scores_df.loc[index, target+'_mfe'] = mfe
        pprint_to_f(scores_df, out=out_fis_root+'_pprint.txt')
        scores_df.to_csv(out_fis_root+ '_finalDesigns.tsv', sep='\t')
        print('time at end:', time.asctime())
```