The generation of a human cell line to elucidate the role of Fumarate Hydratase loss in cancer

University of Cambridge

Christina Schmidt

Newnham College

February 2021

This dissertation is submitted for the degree of Doctor of Philosophy.

Declaration

This dissertation is submitted for the degree of Doctor of Philosophy at the University of Cambridge. The research described herein was conducted under the supervision of Dr. Christian Frezza in the Medical Research Council Cancer Unit at the Hutchison/MRC Research Centre, University of Cambridge, between October 2017 and February 2021. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution. It does not exceed the prescribed word limit of 60000 words.

February 2021

Christina Schmidt

Abstract

THE reprogramming of cellular metabolism is an established hallmark of cancer, which enables cancer cells to survive, proliferate, and metastasize even under harsh environmental conditions. These cancer-associated metabolic changes can affect several pathways one of which is mitochondrial metabolism. The suppression of mitochondrial metabolism has been associated with poor clinical outcomes and mitochondrial dysfunction has been associated with some hereditary and sporadic forms of cancer that arise from mutations in mitochondrial genes. Understanding the mechanisms responsible for cellular transformation and subsequent tumour formation in these hereditary, metabolically-impaired tumours could link dysregulated mitochondrial function and tumour formation.

Hereditary mutations and subsequent loss of the mitochondrial TCA cycle enzyme fumarate hydratase (FH) leads to Hereditary Leiomyomatosis and Renal Cell Cancer (HLRCC), an aggressive form of renal cancer associated with poor clinical outcome. The loss of FH triggers the accumulation of fumarate, which induces a multi-layer cellular reprogramming that contributes to tumorigenesis. Yet, it is unclear how FH loss influences the whole gene expression landscape and if the gene expression is regulated on the level of DNA-methylation, transcription or translation. In this thesis, I generated the first FH-deficient human renal epithelial cell lines using CRISPR/Cas9-based genome editing, and applied proteomics, metabolomics, and transcriptomics approaches to investigate how the loss of FH alters these cellular layers. First, I confirmed that this model faithfully recapitulates the biochemical and phenotypic markers of FH-deficiency as previously reported. Next, I developed a novel multi-omics tool, SiRCle (**Si**gnature **Re**gulatory **Cl**ust**e**rs) to disentangle this interconnected network of signalling cascades. Using SiRCle, I extracted clusters of increased/decreased gene expression that are regulated at the level of DNA methylation, transcription, and/or translation, and identified which clusters drive which phenotype of FH loss. By mapping the transcription factors that drive the genes of each cluster, I identified unique drivers that could be responsible for the cellular rewiring after FH loss.

It is now clear that the tumour microenvironment affects the phenotype of cancer cells, and hence that metabolic rewiring becomes essential for tumour cells to strive even under harsh environmental conditions. Yet, its effect on FH-deficient cells' behaviour is currently unknown. In this part of the thesis, I used a Tumour Roll for Analysis of Cellular Environment and

Response (TRACER), a 3D scaffold that develops oxygen and nutrient gradients similar to those observed in tumours. Using TRACER, I show that the main metabolic signature of FH loss, which is driven by the high levels of fumarate, is not influenced by the nutrient and oxygen gradients generated in this 3D model. Consequently, FH loss is a stronger driver of the metabolic signature than environmental cues. Moreover, by applying linear modelling to the metabolic profile of the cells over the different layers, I identify specific layer-dependent metabolic signatures in FH-deficient cells that are not observed in 2D culture. These results imply that *in vivo* FH loss could undergo previously unacknowledged compensatory metabolic changes, which underlines the important role of the microenvironment in dictating the phenotype of cancer cells.

Contents

List of Figures

Cancer Metabolism

Introduction

1.1. The role of metabolism in cancer

1.1.1. The metabolic determinants of tumorigenesis

During oncogenesis, cells acquire various molecular features that pave the way to malignant transformation. These features are known as the "Hallmarks of Cancer", initially defined as: proliferative signalling, insensitivity to growth suppressors, evasion of cell death, uncontrolled growth, replicative immortality, sustained angiogenesis, and tissue invasion and metastasis[1]. The "Hallmarks of Cancer" are acquired through multiple mechanistic strategies[1] and most of these features are thought to arise from mutations in several oncogenes and tumour suppressor genes, which are defined as "driver genes", that regulate the tumorigenic phenotype[1–3]. This view of cancer as a genetic disease dates back to 1914, when Theodor Boveri postulated that multiple chromosomal defects could result in cancer[4], which formed the basis for the somatic mutation theory[5]. It took until the early 70s when Knudson presented the "two-hit" hypothesis to explain the origin of cancer[6]. He predicted that hereditary cancers harbour an initial germline mutation, and that a second mutation acquired somatically (second hit) leads to tumour formation[6]. Yet, detailed analysis of those oncogenes and tumour suppressors suggests that many of them play a key role in metabolism[7,8], and the reprogramming of metabolism has been added to the "Hallmarks of Cancer"[2]. However, the role of metabolism in cancer is far from a recent discovery. Indeed, the rewiring of cellular metabolism in cancer was already explored more than 130 years ago by Ernst Freund. He noticed high sugar levels in the blood of cancer patients and proposed that reduction in sugar could impact tumour growth[10]. Almost 25 years later, in 1911, the German scientist Wassermann proposed that accelerated proliferation of cancer cells increases their oxygen consumption[11]. Further studies performed just a few years later by Eleanor Van Ness Van Alstyne and colleagues[12] and William Woglom[13] showed that increased carbohydrate intake resulted in accelerated rat sarcoma growth. Together, this seminal work supported the idea that the usage of nutrients and oxygen is different in tumours. A few years after these initial findings, Otto Warburg systematically investigated these observations. He observed that cancer cells ferment glucose to lactate, even in the presence of

oxygen, while normal cells fully oxidise glucose through the mitochondria[14], key "metabolic hubs" of the cell[15]. Based on these findings, Warburg concluded that accelerated glycolysis in cancer is caused by a mitochondrial dysfunction[16]. However, following experiments demonstrated that cancer cells maintain part of their oxidative capacity[17] and that the complete loss of mitochondrial function can be detrimental for cancer cells[18–20]. These counterintuitive findings opened a debate on the role of mitochondrial function in cancer, which is still ongoing[21] and will be discussed in more detail below.

A century after these seminal discoveries, scientists have chartered the metabolic intricacies of cancer and identified multiple "Hallmarks of Cancer Metabolism" (**Figure 1**), which include increased glucose and amino acid uptake, increased demand for nitrogen, and the usage of intermediates deriving from glycolysis for NADPH production[9]. The metabolic

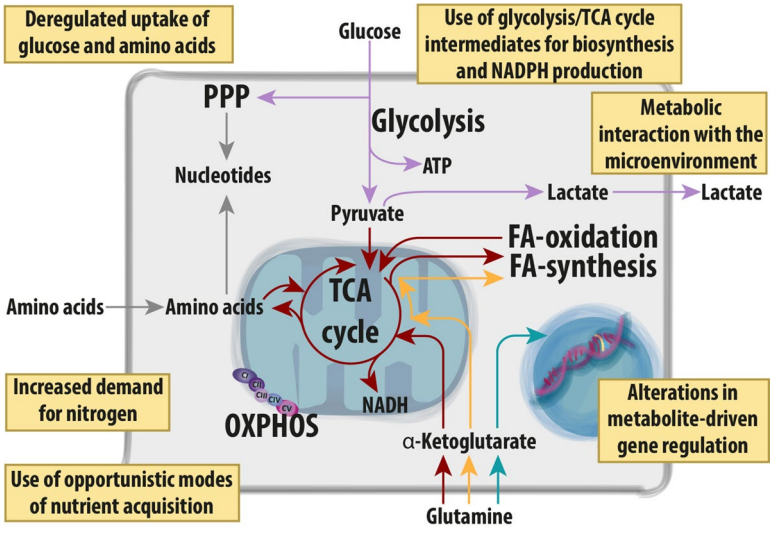


Figure 1: The Hallmarks of Cancer Metabolism.

The Hallmarks of Cancer Metabolism (yellow boxes) support tumorigenesis and can be divided into three layers: 1. Oncogene-directed nutrient uptake, which includes the usage of opportunistic modes of nutrient acquisition and deregulated uptake of glucose and amino acids, 2. Reprogramming of intracellular metabolism, which includes the increased demand for nitrogen and the use of glycolysis/TCA cycle intermediates for biosynthesis and NADPH production, and 3. Metabolite-directed changes in cell behaviour and function including metabolic interactions with the microenvironment and alterations in metabolite-driven gene regulation. Together, all these metabolic alterations allow the cell to create new biomass to sustain proliferation even under nutrient deprivation and changing microenvironments. Glycolysis (lilac arrows) provides the precursors for amino acid, nucleotide and fatty acid (FA) synthesis, and generates energy in form of ATP. The TCA cycle (red arrows) generates ATP via oxidative phosphorylation (OXPHOS) and is fuelled by pyruvate, glutamine (amino acid, red/yellow arrow), aspartate (amino acid, grey arrow) and by fatty acid oxidation. Glutamine is important to regenerate NADH in the oxidative carboxylation (red arrows) and to fuel the fatty acid synthesis via reductive carboxylation (yellow arrows). Moreover, aKG which derives from glutamine can modulate histone and DNA methylation and regulate gene expression (turquoise arrows). Cells also synthesise nitrogen-containing molecules including nucleotides and non-essential amino acids, which are important for cell proliferation and requires nitrogen. Different tumour origin, stage and availability of nutrients and oxygen shapes the distinct metabolic features in cancer. These metabolic features can be modulated via different metabolic routes (arrows).

rewiring in cancer is important to produce energy, antioxidant power and intermediates for the biosynthesis of macromolecules[22]. Metabolic changes can be associated with the reprogramming of intracellular metabolism, oncogene-directed nutrient uptake, and metabolite-directed changes in cell-function[9]. The latter can be driven by the microenvironment and/or *via* alterations in metabolite-driven gene regulation[9]. Given that these "Hallmarks of Cancer Metabolism" have been proposed to drive tumorigenesis[9], they could replace the need of additional somatic mutations for full-blown transformation. An example of this concept is that of tumours driven by mutations in key mitochondrial enzymes such as fumarate hydratase (FH), succinate dehydrogenase (SDH) or isocitrate dehydrogenase (IDH). In these tumours, additional downstream oncogenic mutations thought to induce the "Hallmarks of Cancer"[1,3] are instead substituted by transforming metabolic changes, as described in the "Hallmarks of Cancer Metabolism"[9]. This aspect will be discussed more in details below.

Recent evidence showed that metabolic rewiring also plays a crucial role during tumour progression and metastasis[23–25]. For instance, metastasis formation appears to require metabolic factors that allow cells to acquire cell-autonomous properties for increased invasiveness and/or to overcome the challenging alterations of the microenvironment[24]. This changing microenvironment during tumour growth and progression includes gradients of nutrient and oxygen[23], which establish within the solid tumour cell populations with distinct metabolic configurations depending on the distance from blood vessels or stromal cells, and leads to tumour heterogeneity[26]. Furthermore, the tumour interstitial fluid deriving from different regions of the same tumour show distinct metabolic profiles[27]. Consequently, the metabolic reprogramming during tumour progression require both "metabolic flexibility" and "metabolic plasticity"[23], whereby metabolic flexibility ensures the ability to use different nutrients, whilst metabolic plasticity is the ability to process the same nutrient *via* different metabolic pathways[23].

Together, these lines of evidence show that understanding if changes in metabolism are a mere consequence of transformation or could have a driving role is crucial. To date, researchers hypothesise that further understanding the mechanisms responsible for cellular transformation, subsequent tumour formation and progression in hereditary, metabolically impaired tumours could provide the link between dysregulated mitochondrial function and tumour formation.

1.1.2. The role of mitochondria in cancer

Mitochondria are important metabolic hubs of a cell: They not only produce energy in the form of ATP, but they also coordinate several metabolic pathways required for survival and proliferation[15]. In a recent study from our laboratory, aimed at identifying the metabolic

landscape of cancer, we found that the suppression of mitochondrial genes is as a key metabolic feature of cancer progression and metastasis[28]. These results appeared consistent with the initial Warburg hypothesis and were further corroborated by other computational studies[29–31]. However, these studies are based on the transcriptomic profile of tumours (TCGA database) and do not take into account tumour heterogeneity or the protein landscape, which are important to understand the actual phenotype of cancer, as will be discussed in details below. In this context, it is also important to note that the complete loss of mitochondrial function can be detrimental for cancer cells and that not all tumours exhibit mitochondrial dysfunction[18,19]. Mitochondrial dysfunction can be caused by mutations of nuclear-encoded mitochondrial genes or mitochondrial DNA or be secondary to the activation of specific oncogenic cascades and environmental cues such as changes in nutrient and/or oxygen availability[6,21,32]. A key finding that puts mitochondrial dysfunction in the driving seat of cancer is that hereditary and sporadic forms of cancer can arise from mutations in mitochondrial genes, including *FH*, *SDH* and *IDH*[32]. These mutations of TCA cycle enzymes subsequently lead to the accumulation of metabolites that have additional signalling functions, such as fumarate (FH-deficiency), succinate (SDH-deficiency) and 2-Hydroxyglutarate (2HG, IDH-mutation)[33]. The function of these metabolites in tumorigenesis is comparable to that of oncogenes and hence they are often termed as "oncometabolites"[33]. This signalling function of intermediates and derivates of the TCA cycle has not only been associated with intracellular oncogenic signalling, but also with alteration of immune cell effector function[34]. For example, fumarate has been shown to enhance cytokine production in monocytes upon re-stimulation with LPS in trained immunity[34,35].

To sum up, mitochondria play a crucial role in cancer, and cancer formation appears to have both genetic and metabolic roots. It is clear that dysfunctional mitochondrial metabolism plays a role in the process of tumorigenesis and progression, and there is strong evidence that mitochondrial dysfunction could even drive cellular transformation in some circumstances. However, the mechanisms underpinning transformation in these metabolically impaired tumours are still not fully understood. Unravelling these mechanisms could provide the link between dysregulated mitochondrial function and transformation. In this thesis, I will focus on the role of FH loss in cancerous transformation.

1.2. Fumarate Hydratase mutations in cancer

The *FH* gene encodes for both a cytosolic and a mitochondrial enzyme, which differ only in their peptide sequence at the N-terminus[36]. As part of the mitochondrial TCA cycle, FH catalyses the reversible hydration of fumarate to malate[15] (**Figure 2A**). Hereditary germline

mutations of one allele of FH and the loss of the wild type allele predispose to Hereditary Leiomyomatosis and Renal Cell Cancer (HLRCC), a cancer syndrome characterised by benign skin and uterine leiomyomas, and papillary type II renal cell carcinoma, an aggressive form of renal cell carcinoma (RCC)[37,38](**Figure 2B**). So far, no correlation has been found between the site of mutation and clinical outcome in patients[39], thus indicating that loss of FH activity and not neomorphic functions of the mutant protein is responsible for cellular transformation. Noteworthy, sporadic FH loss[40,41,42,43] or its transcriptional downregulation[44,45,46] have been observed in many other tumour types, which shows a key role of *FH* loss in human cancers (**Figure 2B**).

The complete loss of the *FH* gene induces a multi-layer cellular reprogramming that includes metabolic rewiring essential for cells to survive the severe mitochondrial dysfunction and molecular alterations, which together can promote tumorigenesis[47]. In detail, we have postulated that FH driven tumorigenesis occurs *via* a cascade of specific steps that contribute to cellular transformation[47]. The first step upon the loss of FH is a compensatory metabolic adaptation, which is required to survive the truncation of the TCA cycle. The second step is the activation of oncogenic signalling cascades *via* transcriptional, epigenetic and post-translational reprogramming in part dependent on fumarate accumulation (**Figure 3**).

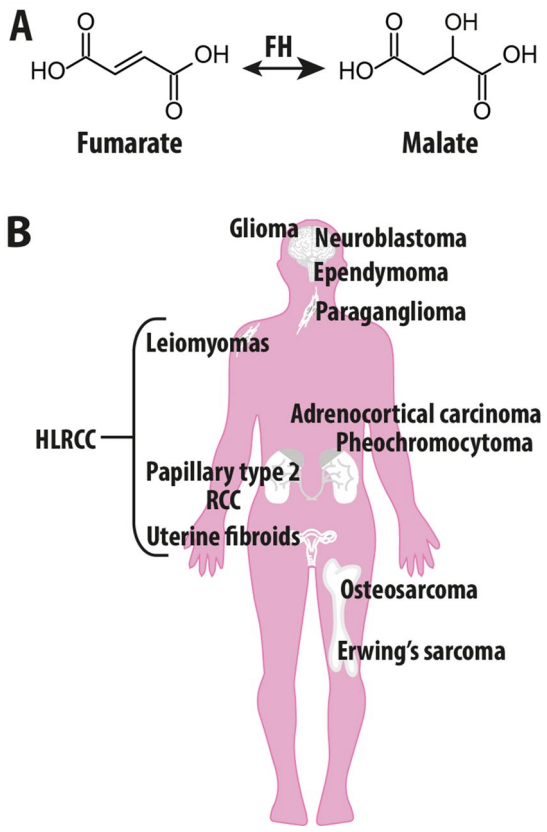


Figure 2: Fumarate Hydratase in cancer.

(**A**) Schematic of the chemical reaction catalysed by FH.

(**B**) Tissues where sporadic or hereditary loss of FH leads to cancer. The sporadic loss of FH has been reported in Hereditary Leiomyomatosis and Renal Cell Cancer (HLRCC) including skin leiomyomas, uterine fibroids, and papillary type II RCC[37,38]. Sporadic loss of FH has also been reported in other tumour types such as pheochromocytomas, paragangliomas[40,41], adrenocortical carcinoma[37], neuroblastomas[42,43], glioma, ependymoma, osteosarcoma, and Ewing's sarcoma[37]. Transcriptional downregulation of FH was found in sporadic clear cell carcinomas[44] and in colorectal cancer[45], and additional evidence suggests the involvement of FH mutations in breast, bladder, and testicular cancers[46].

1.2.1. Compensatory metabolic adaptation in FH-deficient cells

The loss of FH leads to a truncated mitochondrial TCA cycle and subsequent mitochondrial dysfunction. Since mitochondrial metabolism is involved in fatty acid oxidation, carbohydrate metabolism and amino acid synthesis, all processes essential for cell growth and survival[48], cells need to engage in a series of metabolic adaptations to overcome possible defects in these processes (**Figure 3** and **Figure 4**)[47]. Based on previous findings from the lab, the initial response to the TCA cycle impairment is the compensatory switch towards glycolysis, whereby glucose is diverted into lactate production[49] and the pentose phosphate pathway[50], and glutamine oxidation is activated[49]. The latter becomes the main source of carbons for the TCA cycle[49]. Moreover, glutamine oxidation also allows the regeneration of NADH, crucial for oxidative phosphorylation (OXPHOS) and the maintenance of mitochondrial membrane potential[49]. In these conditions, fumarate starts to accumulate up to millimolar levels[51] and in order to release the glutamine-derived carbons from the TCA cycle without further exacerbating the accumulation of fumarate, FH-deficient cells engage in the haem biosynthesis and degradation pathway, which becomes essential for their survival[49]. In patient-derived FH-

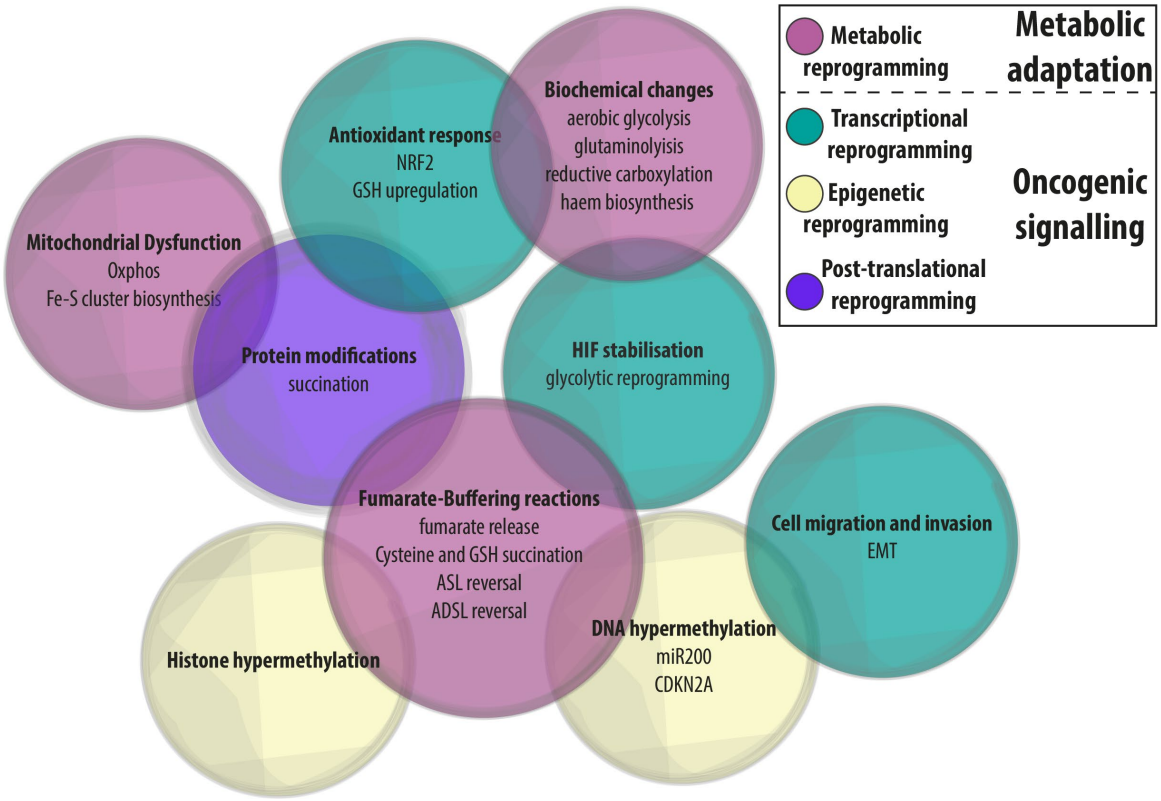


Figure 3: Survival of FH loss and tumorigenesis.

Schematics representing the different aspects of FH-deficient cellular reprogramming. The overlap of the circles represents potential connections.

deficient cells, namely UOK262 cells, glutamine-derived carbons are also shunted into lipid biosynthesis *via* reductive carboxylation[52], a biochemical pathway where glutamine leads to the synthesis of lipogenic acetyl coenzyme A (acetyl-CoA). Yet, reductive carboxylation has not been observed in Fh1-deficient mouse cells[49] or fibroblasts[53]. Lastly, fumarate accumulation also alters a set of enzymatic reactions in which it is directly involved as a substrate or a product such as the urea cycle, purine nucleotide cycle (PNC), and SDH. As a consequence, the accumulation of fumarate in FH-deficient cells leads to the reversal of argininosuccinate lyase (ASL) and the accumulation of argininosuccinate[51,54], the reversal of adenylosuccinate lyase (ADSL) and subsequent accumulation of adenylosuccinate (unpublished data from the lab) and inhibition of SDH, which reduces mitochondrial respiration[55].

1.2.2. The oncogenic signalling elicited by FH loss

Besides dysregulating metabolic pathways as described above, FH-deficient cells engage in fumarate-buffering strategies that include the release of fumarate to the microenvironment[56] and the reversal of fumarate-producing biochemical pathways. The two main events I want to discuss in this context are: 1) The competitive inhibition of enzymes that require structurally similar metabolites to fumarate such as α-ketoglutarate (aKG) or succinate as substrate or product[57] and 2) The non-enzymatic chemical reactions of fumarate with freely available cysteines, cysteine residues of small molecules[58,59] and cysteine residues of proteins[55] (**Figure 5**). Together, these two events have been shown to induce a plethora of oncogenic signalling cascades that are thought to support cellular growth and transformation (**Figure 3**), which I discuss in detail below.

The first type of oncogenic signalling triggered by fumarate is related to its structural similarity to succinate and aKG, both involved in mediating reactions catalysed by the superfamily of αKG-dependent-dioxygenases (aKGDDs)[57,60], enzymes involved in a variety of cellular processes including protein hydroxylation, DNA and histone demethylation, and RNA modifications[57,60]. Fumarate can act as a competitive inhibitor of aKGDDs, with important biological implications. For instance, the inhibition of aKGDD prolyl-hydroxylases (PHDs) stabilises Hypoxia Inducible Factor (HIF)[61], leading to the transcription of genes involved in the hypoxic response, even at normal oxygen conditions, a phenomenon known as pseudohypoxia[46]. Amongst the HIF target genes are several metabolic enzymes including the glucose transporter 1 (GLUT1)[61], pyruvate dehydrogenase kinases (PDKs)[62,63] and lactate dehydrogenase A (LDH-A)[64]. Together, these enzymes could support the glycolytic shift observed in FH-deficient cells, and block the entry of glucose into the mitochondria[56]. Whilst this hypoxic response is known to regulate tumour proliferation, metabolism and

angiogenesis[65], the role of HIFs in FH deficiency is still debated. Indeed, the genetic deletions of both *Hif1*and *Hif2* in Fh1-deficient mice did not prevent carcinogenesis, which suggests that HIF proteins are dispensable for tumorigenesis[66]. Another class of aKGDDs are DNA and histone demethylases that fine-tune chemical changes of DNA and histones within the

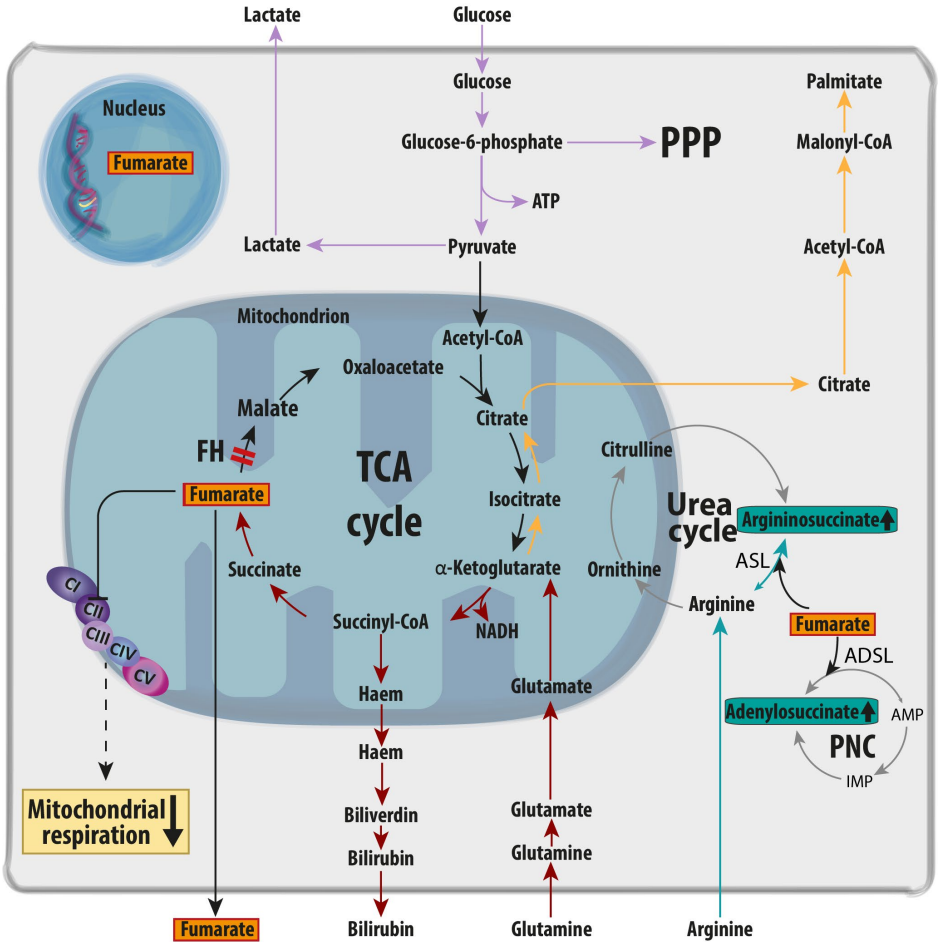


Figure 4: Metabolic adaptation in FH loss.

FH loss leads to TCA cycle truncation and fumarate accumulation (highlighted in orange). The disruption of the TCA cycle and the inhibition of Succinate Dehydrogenase (SDH, also known as Complex II of the respiratory chain) by fumarate reduces mitochondrial respiration. FH-deficient cells shift towards aerobic glycolysis, reducing the oxidation of glucose in the mitochondria (lilac arrows). Moreover, glucose is shunted into the pentose phosphate pathway (PPP) to maintain redox homeostasis (lilac arrows). To regenerate NADH in the mitochondria, FH-deficient cells increase glutamine oxidation and the glutamine-derived carbons are further metabolized to fumarate and, through the haem pathway, to biliverdin and bilirubin, which is secreted into the medium (red arrows). Additionally, glutamine-derived carbons are used to generate lipogenic acetyl-CoA *via* reductive carboxylation (yellow arrows). Fumarate permeates the various intracellular compartments and is released into the extracellular milieu. Fumarate accumulation also leads to the aberrant production of argininosuccinate *via* the reversal of the urea cycle enzyme argininosuccinate lyase (ASL) (turquoise arrows). Moreover, fumarate leads to the accumulation of adenylosuccinate, likely via the reversal of adenylosuccinate lyase (ADSL) within the purine nucleotide cycle (PNC).

Figure was adapted with permission from Schmidt et. al.[47]*. Copyright 2019 Elsevier.*

nucleus[67,68]. The demethylation of cytosine residues on DNA is catalysed by Ten-Eleven Translocation (TETs) proteins, which are inhibited upon FH loss[69]. The inhibition of TETs induces a distinct hypermethylation profile in FH-deficient cells and tumours, which triggers the suppression of cyclin dependent kinase inhibitor 2A (*CDKN2A*)[70,71] and epithelial to mesenchymal transition (EMT)[69]. The latter is a signature associated with poor clinical outcome in cancer patients[72], and thought to be triggered *via* the epigenetic suppression of antimetastatic miRNAs, *MIR200*, in FH-deficient cells[69]. *CDKN2A* encodes for p16, which is a known inducer of senescence (a permanent growth arrest)[73] and it has been discussed that the suppression of p16 inhibition might be a way for the cells to overcome fumarate-induced senscence[70,71,73]. In this context, it is important to note that the depletion of the antioxidant glutathione (GSH), a small molecule including a cysteine residue that reacts with fumarate, together with the profound oxidative stress caused by the disruption of mitochondrial function, could promote senescence in FH-deficient cells[58]. However, given that FH-deficient cells can form tumours, the cells must have mechanisms to circumvent senescence. It has been previously discussed that fine-tuning of the antioxidant response, which is orchestrated by different oncogenic signalling cascades in FH-deficiency (discussed in detail below), could be crucial to overcome senescence[47]. Together with the antioxidant response, the epigenetic silencing of *CDKN2A* and activation of EMT can be beneficial to evade senescence and at the same time promote cell migration and invasion.

The second type of oncogenic signalling is based on the reaction of fumarate with freely available cysteines, cysteine residues of small molecules such as GSH[58,59] and cysteine residues of proteins[55] in a process called succination leading to S-(2-succino) cysteine (2SC)[74] (**Figure 5**). The latter leads to an irreversible, post translation modification (PTM)[74,75] which can be detected by anti 2SC antibodies and is an established marker of FH loss in cancer patients[76]. Thus far, various proteins have been identified as targets of succination in FH-deficient cells[55], one of which is the Kelch-like ECH-associated protein1 (KEAP1), the negative regulator of the antioxidant gene Nuclear Factor Erythroid 2 Like 2 (NRF2). Succination inactivates KEAP1, preventing the degradation of NRF2, which in turn mounts a powerful antioxidant response[77,78] mediated by genes such as haem oxygenase 1 (HMOX), NAD(P)H dehydrogenase quinone 1 (NQO1), and glutamate-cysteine ligase catalytic subunit (GCLC)[77,78], and could be part of the fine-tuned antioxidant response mentioned above.

Besides KEAP1[66,78], a multitude of succinated targets have been identified, including iron regulatory protein 2 (IRP2)[79], the iron-sulfur-cluster (Fe-S cluster) biogenesis family of proteins[55], and mitochondrial aconitase (ACO2)[53]. IRP2 succination inactivates the protein and

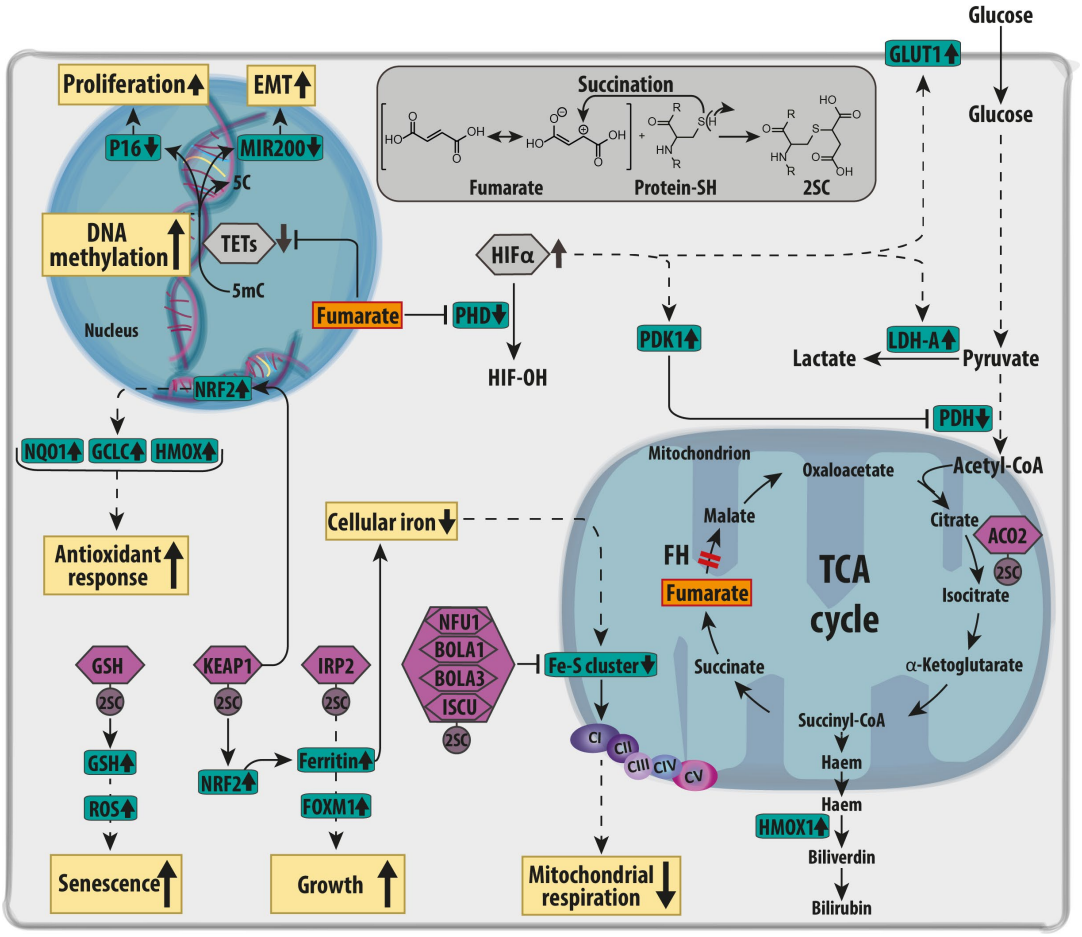


Figure 5: Oncogenic signalling via epigenetic, transcriptional and PTM reprogramming.

Fumarate accumulation leads to a PTM that affects cysteine residues of small molecules/proteins (violet hexagons) called succination. The chemical reaction between fumarate and reactive thiol residues of proteins is depicted in the grey insert. Succination of KEAP1 induces the NRF2-mediated antioxidant response. One of the targets of NRF2 is Haem Oxygenase 1 (HMOX1), which is required for the haem biosynthesis and degradation, an essential pathway for the survival of FH-deficient cells. Succination of Iron Responsive Element Binding Protein 2 (IRP2) inhibits the translation of ferritin. Subsequent ferritin increase causes a drop in free cellular iron. In parallel, ferritin promotes the expression of Forkhead box protein M1 (FOXM1), a pro-mitotic protein that supports cell growth. Succination of the Fe-S cluster proteins Nfu1, Bola and Iscu impairs the Fe-S clusters assembly required by the electron transport chain complex I, contributing to defects in mitochondrial respiration. The reduction of iron and the succination of key cysteine residues in its catalytic core also inactivates the TCA cycle enzyme Aconitase 2 (ACO2). GSH succination causes the depletion of glutathione (GSH) stores, increasing oxidative stress and triggering senescence.

Fumarate mediates oncogenic signalling *via* aKGDDs inhibition (grey hexagons). Inhibition of prolyl hydroxylases (PHDs) stabilises of the alpha subunit of a family of hypoxia-inducible factors (HIFs). The transcriptional response elicited by HIFs promotes aerobic glycolysis via increased expression of the glucose transporter GLUT1, and lactate dehydrogenase (LDH-A). HIF triggers the expression of pyruvate dehydrogenase kinase 1 (PDK1), which phosphorylates and inhibits pyruvate dehydrogenase complex (PDH), a gatekeeper of glucose-derived pyruvate in the mitochondria. In the nucleus, fumarate accumulation induces a profound epigenetic reprogramming due to the inhibition of DNA demethylases (TETs). Hypermethylation

of miR200 was shown to trigger an epithelial-to-mesenchymal transition (EMT) in FH-deficient cells, and hypermethylation of CDKN2A (p16) increases proliferation and helps to overcome senescence.

Figure was adapted with permission from Schmidt et. al.[47]. *Copyright 2019 Elsevier.*

leads to the translation of ferritin causing depletion of freely available iron[79]. Additionally, succination of multiple family members of the Fe-S cluster assembly proteins reduce the activity of Fe-S clusters-dependent enzymes such as the respiratory chain complex I (CI)[55]. Aco2 succination and subsequent enzyme inactivation has also been observed in *Fh1*$^{-/-}$ mouse embryonic fibroblasts, potentially preventing the use of glutamine for citrate formation through reductive carboxylation[53].

In summary, FH loss results in compensatory metabolic adaptations and fumarate accumulation, which elicit a plethora of pro-oncogenic signals that could contribute to transformation (**Figure 3**). Yet, the contribution of metabolic rewiring in the process of cellular transformation is still unclear. Moreover, it is not fully understood which of these metabolic changes occur at the early phases of transformation or at the later stages of tumour progression. One problem to adequately address this question is the definition of tumorigenesis. Understanding at which point a cell becomes cancerous is crucial to study the impact of metabolic rewiring on tumorigenesis. A way to overcome this issue is to investigate the presence of the "Hallmarks of Cancer" by pathway analysis. However, this approach is mostly based on transcriptomics data, which are known to correlate poorly with protein abundance, which in turn defines the cellular phenotype. Finally, it is emerging that the metabolic phenotype of a cell strictly depends on extracellular nutrient cues. Therefore, it is essential to assess the impact of the tissue environment on the metabolic rewiring caused by FH loss to fully understand the connection between dysregulated metabolism and oncogenesis *in vivo*.

1.3. Cancer phenotype and environmental cues

1.3.1. Pathway analysis to define the cancer phenotypes

Technological advances like genome-wide sequencing technologies have led to large collections of datasets from tumour specimens[80]. One of the largest endeavours to systematically collect and analyse these data is The Cancer Genome Atlas (TCGA)[80]. The TCGA provides publicly available large-scale genome sequencing data, including DNA sequencing (DNAseq), DNA methylation-seq, RNAseq, microRNAseq, Reverse-phase protein array (antibody based functional and quantitative proteomics method) and many more[80]. To

find dysregulated groups of genes, bioinformatic tools have been developed to perform differential expression analysis between e.g. tumour and healthy tissue and subsequent gene set enrichment (GSE) methods[81,82]. Gene set collections are collections of biological pathways or molecular network information about a biological system, such as the Molecular signatures database (MsigDB)[83]. This gene set collection extracted signatures from original research publications and imports entire collections such as Gene ontology[84] (GO), a hierarchy of controlled terms to describe individual gene products, or the Kyoto encyclopedia of genes and genomes[85] (KEGG)[83]. Performing GSE methods using these gene sets reduces the data to smaller, more interpretable sets of altered signalling pathways or processes[81].

The different genomic assays capture a narrow view of the complex biological system and omits interactions and events that are regulated at multiple levels (e.g. post-transcriptional, or post-translational)[86]. Consequently, several efforts to combine different omics data to unravel coherent biological signatures and to make predictions about the phenotype were made[86]. In this context, data scientists have been capitalizing on machine learning approaches to analyse cancer data, which has improved the accuracy of survival analysis and diagnosis[87]. One of these approaches are Variational Autoencoders (VAEs), an unsupervised method that provides a latent representation of integrated data[88]. Similar to principal component analysis (PCA), VAEs perform dimensionality reduction, but they are not restricted to linear transformation and can also model complex non-linear functions and correlate the different input features[88,89]. For cancer data, VAEs have been used for data integration to combine the data types at different scales to identify cancer traits, but also to learn compressed representations of the input data by learning the underlying distribution of the input data[87]. For instance, applied to TCGA pan-cancer RNA-seq data, VAE was able to separate the data according to the underlying cancer type[90].

Given there is no standard method for integrating multi-omics data, and the challenges that integration presents, the transcriptome (RNAseq data) is commonly used to perform GSE analysis (GSEA). However, since gene expression does not always correlate with protein abundance and the cellular phenotype, inferring pathway activities based exclusively on transcriptomics data may not be sufficient to predict the phenotype accurately. Indeed, translational control of gene expression plays a crucial role during tumorigenesis to support the transformed phenotype and ensure the cancer cell function[91]. Despite observing a general increase in global translation rates, which is important for cancer cells to sustain their increased proliferation, oncogenic signalling has been observed to induce transcript-specific changes in translation[91]. This oncogenic regulation of translation drives features of cancer, including

altered metabolism *via* the translational control[92], coupling nucleotide metabolism with protein synthesis[92]. Another example is the resistance to oxidative stress *via* eIF4E (eukaryotic initiation factor 4E, initiation factor for mRNA translation)-driven oncogenic translation of ROS scavengers that fuel tumorigenesis[93]. Given that translation control is the most energetically expensive molecular process in the cell, it is not surprising that it plays a key role in cell growth and metastasis in response to environmental changes like nutrient stress[94,95]. Hence, translational regulation allows cancer cells to adapt to environmental stressors such as nutrient and oxygen deprivation[94,95].

1.3.2. The impact of the environment on the metabolic phenotype of cancer

Cancer cells need to survive and adapt to the harsh environmental conditions, including low oxygen levels and constrained nutrient availability[23], which can be secondary to reduced tumour vascularization[96]. Low oxygen levels have been shown to trigger multiple adaptational processes important in cancer. For instance, hypoxia triggers vascularization, which in comparison to healthy tissue, has a chaotic, altered structure and function[96]. Moreover, hypoxia leads to increased cell mobility and metastasis, and alters cancer cell metabolism[62,97]. Changes in intracellular metabolism affect also the extracellular milieu through the excretion of metabolites such as lactate, which, in turn, can alter tumour microenvironment and cell-cell interactions[98]. This metabolic rewiring is also influenced by the tissue origin, the location, and the size of the tumour[28,45,99,100], and consequently, it has been argued that some features of cancer metabolism observed in 2D cell culture do not fully recapitulate the metabolic changes *in vivo*, and might instead be affected by culture methods and culture media[27,101]. For instance, recent experiments showed that replacing the traditional cell culture media with a media that mimics the metabolic content of the tumour microenvironment can significantly affect the overall metabolic landscape of cancer cells, and, as a consequence, the sensitivity to anticancer drugs that target metabolic enzymes[102]. These results led to the development of "physiological media", including Gibco Human Plasma-like Media (HPLM) and Plasmax, which recapitulate nutrient and metabolite concentrations found in the blood of healthy individuals[103]. Plasmax has been shown to increase colony formation capacity in breast cancer and to differentially regulate gene expression[103]. In line with the fact that environmental cues affect the cancer cell's phenotype, it was shown that the inhibition of certain metabolic routes can suppress oncogenic pathways and reverse the tumour phenotype in 3D, but not in 2D cultures[104]. Moreover, it has been shown that within solid tumours, metabolic cooperation arises: For instance, lactate secreted by glycolytic cells can be taken up and used by oxidative cancer cells[105,106].

Consequently, growing cells in 3D culture can better reflect the metabolic features of cancers and enable us to understand metabolic cooperation[27].

To shed light on the role of the microenvironment in cancer cell behaviour, a multitude of different 3D systems have been generated. In 3D systems, cells are usually cultured in aggregates, grown on various scaffold materials or embedded in gels[107]. Cancer cells cultured in aggregates are called tumour spheroids[107]. They mimic the phenotype of the human tissues including cell polarity, apical brush border and receptor-mediated transport[107,108]. Moreover, they recapitulate the gradient of nutrients and oxygen as observed in patient tumours[107,108]. This aspect is pivotal to study cellular metabolism and the metabolic plasticity observed in cancer[27]. A more complex structure compared to spheroids are organoids, which derive from stem cells or human tissue grown in an organised network mimicking a specific organ[108]. The complexity of the organoid depends on the developmental potential of the starting stem cells[109,110]. For instance, kidney organoids have been shown to include individual nephron segments into distal and proximal tubules, early loops of Henle and glomeruli[108,111–113]. Importantly, they have been already used to mimic renal diseases and study both tumour metabolism and epigenetic reprogramming such as histone-modifications[108,111–113]. Recent evidence showed that renal carcinoma organoids could be used as a novel tumorigenesis gene discovery model[114]. Yet, in both systems, spheroids and organoids, the collection of cells from defined areas within the tumour model is not possible[115]. To enable the rapid collection of cells from different areas of a 3D tumour model that mimics the cell-to-cell contact and gradients of oxygen and nutrients, a Tissue Roll for Analysis of Cellular Environment and Response (also known as TRACER) was developed[116]. This model consists of a permeable biocomposite scaffold strip that is populated with cells embedded in collagen[116]. This strip is wrapped around a mandrel and allows the generation of oxygen and nutrient gradients that mimic those observed in a solid tumour[116]. Rapid disassembly allows the analysis of cells grown under different environmental conditions whilst preserving the oxygen and nutrient gradients for downstream analyses in a layer specific manner[116]. Using this system, it was shown that cells exhibit spatially-defined metabolic signatures, in an oxygen-dependent fashion[115]. Hence TRACER can spatially correlate the different tumour environments to real-time snapshot metabolic signatures.

To sum up, studying cancer metabolism using standard 2D culture systems and commercial media cannot mimic selective environmental pressures such as nutrient and oxygen gradients as present in a tumour. Consequently, to understand the link between dysregulated mitochondrial function and tumour formation and progression in a physiological context, novel 3D culture systems need to be used. Given that the metabolic state of a cells is significantly

altered by nutritional and environmental cues, it is crucial to understand to what extent the metabolic signature of FH loss is affected by these cues. Addressing this question is essential to better understand the consequences of FH loss *in vivo* and could reveal new cascades which are not present in 2D cultures.

Aims of the study

THE reprogramming of mitochondrial metabolism has been associated with poor clinical outcome and mutations in some mitochondrial genes lead to hereditary and sporadic forms of cancer. Understanding the mechanisms responsible for cellular transformation and subsequent tumour formation in these hereditary, metabolically-impaired tumours could reveal novel molecular links between dysregulated mitochondrial function and tumour formation.
Here, I used the loss of FH as a genetic model to investigate this connection. Although fumarate accumulation in FH-deficient cells is thought to elicit a multi-layer cellular reprogramming that contributes to tumorigenesis, it is still unclear how these signals are integrated and coordinate transformation. Recently, it has become clear that the tumour microenvironment affects the phenotype of cancer cells and that culturing cells in standard culture media under 2D condition does not mimic environmental pressures and can alter the metabolic phenotype. How the signals elicited by FH loss are shaped by environmental cues such as oxygen and nutrient availability is currently unknown.

My study is aimed at:

1) Establish and characterize a non-transformed human model of FH loss.
2) Perform a computational analysis of the consequences of FH loss by integrating multiple omics approaches and investigating the regulatory levels (methylation, transcription, translation) that induce the FH phenotype.
3) Understand the effects of environmental cues on the hallmarks of FH loss using a 3D model.

Methods

2.1. Experimental model and cell culture

HK2 cells, an immortalised epithelial cell line deriving from the proximal tubule of normal human kidney were purchased from ATCC (cat. No. CRL-219). $FH^{+/+}$ and $FH^{-/-}$ were generated using CRISPR/Cas9-based genome editing (detailed in 2.2. Genome editing). Cell culture was performed in a sterile environment with sterilised materials. Cells were cultured in an incubator at 37°C, 5% CO_2 and 100% air moisture with Dulbecco's modified Eagle's medium (DMEM, Life Technology cat. no. 41966) supplemented with 10% v/v fetal bovine serum (FBS, Gibco cat. no. 10270-106). Before sub-passaging, cells were allowed to grow for two to three days until 90% confluent. Cell lines were authenticated using STR profiling and routinely tested for mycoplasma contamination.

2.2. Genome editing: FH-deficient cells

The single guide RNA targeting the *FH* gene (exon 4: CCAGTCTGCCATACCACGAG and exon 2: GCGCCATAATACTTATCATT) introduced into the lentiCRISPR V2 vector with puromycin resistance cassette (Zhang lab, plasmid #52961) was a generous gift of Dr. James Nathan laboratory (Cambridge Institute for Medicine (CIMR), Cambridge, UK). For virus production, 4 µg CRISPR-vector, 24 µl Lipofectamine 2000 (Life Technology) and 296 µl Optimem (Life Technology) were added to the cell culture media of the packaging cells HEK293T following the manufactures instructions. After 24 h, media was changed for 24 h before the viral supernatant for cell transduction was obtained from the filtered growth media. HK2 cells seeded into 6 cm dishes were infected at 60% confluency with 3 ml of neat viral supernatant in the presence of 8 mg/ml polybrene for 48 h. Afterwards, the media was replaced with selection medium containing 1 µg/ml puromycin for 3 days. Antibiotic-selected cells were picked by diluting single-cell clones into a 96-well plate through serial dilution. Then, cells were counted and diluted to 1 cell/100 µl, followed by pipetting of 100 µl cell suspension per well of a 96-well plate. Clones were expanded for three weeks. In total 22 $FH^{+/+}$ clones were picked and 25 $FH^{-/-}$ clones. The FH expression of all clones was estimated by Western blotting

(WB, method details in 2.4.) and epithelial marker expression (method details in 2.5), oxygen consumption rate (OCR, method details in 2.6.) as well as fumarate level (method details in 2.8.) were analysed for 5 $FH^{-/-}$ and 5 $FH^{+/+}$ clones. To test off target effects a common method is the reconstitution of the target gene in the knockout clones, yet often this does not reverse all changes since changes induced by e.g. aging cannot be reversed. Hence it is difficult to use this principle to evaluate off target effects in this system. To account for changes that could result from off target effects, two independent CRISPR clones were used for all downstream experiments. In detail, two $FH^{+/+}$ clones with a different degree of epitheliality and two $FH^{-/-}$ clones that had a complete loss of FH protein and high fumarate accumulation, whilst retaining decent proliferative capabilities, were selected.

2.3. Sanger Sequencing to confirm FH knockout

Primers where designed using Primer3 and Blast. Genomic DNA was extracted using DNeasy Blood and Tissue kit (Qiagen cat no. 69504) and the area around exon 2/exon 4 was amplified by genomic PCR using Phusion High-Fidelity DNA Polymerase kit (New England Biolabs cat no. M0530L) running the following protocol:

Reaction	Amplification		Primer pairs
Phusion buffer 5x	98°C - 30s	1x	**Exon 2** (66.1°C annealing)
10mM dNTPs	98°C – 10s		fw-ATTCCTCGAACTCCCTGCTC
10µM fw primer	XX°C – 30s	34x	rv-GAACCTCTTATTACTCACGAAGC
10µM rv primer	72°C – 2 min		**Exon 4** (70°C annealing)
0.2µl Phusion DNA	72°C – 5 min	1x	fw-AATAGGGGCAAATCTGGGCA
100ng gDNA target	4°C - ∞		rv-TGAGGACAGAAAAGATGGC
ad. 20µl ddH_2O			

Next, the PCR product was purified using the Qiaquick PCR purification Kit (Qiagen, cat no. 28106) according to the manufacturer's instructions and quality checked by running samples mixed with 3 µl 6x loading dye on 1% agarose gel (1g agarose, 100ml TAE buffer, 10 µl Sybr safe dye (Invitrogen, cat no. PINS33102)). Samples were sent with the respective primers for Sanger sequencing (Eurofins genomics).

2.4. Western blotting (WB)

Proteins were extracted using radioimmune precipitation assay (RIPA) lysis buffer (150 mM

NaCl, 50 mM Tris, 1 mM EGTA, 1 mM EDTA, 1% (v/v), Triton X-100, 0.5% (w/v) sodium deoxycholate, 0.1% (v/v) SDS, pH 7.4) supplemented with 1 µl benzonase *per* sample. Protein concentration was measured by bicinchoninic acid (BCA) assay (Thermo Fisher Scientific), according to the manufacturer's instructions. 25 µg of protein extracts were mixed with Bolt loading buffer (Life Technologies cat. no. B0007) with 4% β-mercaptoethanol (Sigma-Aldrich cat. no. M6250)) and incubated for 10 min at 70°C. Samples were loaded and run at 200 V for 30-40 min using Bolt gel 4%–12% Bis-Tris (Invitrogen cat. no. NW04122BOX) and 1x MES running buffer (Life Technologies cat. no. B0002). Proteins were transferred onto a nitrocellulose membrane using a dry transfer system IBLOT2 (Life Technology), which was run for 12 min at 20 V. The membrane was stained for total proteins using the Revert 700 total protein kit (Licor, cat.no. 926-11010) according to the manufacturer's instructions. The image was acquired in the 700-channel using Odyssey CLx (Licor). Afterwards, the membranes were blocked 30 min at room temperature using blocking buffer (5% milk in 1xTBS). Antibodies were diluted in blocking buffer supplemented with 0.01% Tween 20. The membrane was incubated in the primary antibodies on a shaking platform at 4°C typically for 16 h or 2 h at room temperature. Secondary antibodies were diluted 1:2000 and incubated on a rocking platform 1 h at room temperature. Images were acquired and analysed using Odyssey CLx (Licor).

2.5. RNA extraction and quantitative PCR

The RNA was extracted using RNeasy Kit (Qiagen) and the miRNA using the miRCURY RNA Isolation Kit (Exiqon), following manufacturer's instructions. The quantification of the RNA was performed using Nanodrop (Thermo Fisher Scientific) following the manufacturer's instructions. Next cDNA was produced *via* reverse transcription of RNA using Quantitect Reverse transcription kit (Qiagen) or miScript PCR kit (Qiagen) using 300 ng total RNA. qPCR was performed using Quantitect Syber Green master mix (Qiagen) on a Step One Plus real-time PCR system (Life Technology) and experiments were analysed using the StepOne software (v2.3).

Housekeeping genes used for internal normalisation:

TBP	mRNA	fw: GAACATCATGGATCAGAACAACA rv: ATAGGGATTCCGGGAGTCAT
SNORD95	miRNA	Qiagen cat no. MS00033726
SNORD61	miRNA	Qiagen cat no. MS00033705

Target genes:

E-CADHERIN	mRNA	fw: TGGAGGAATTCTTGCTTTGC rv: CGCTCTCCTCCGAAGAAAC
EPCAM	mRNA	Qiagen cat no. QT00000371,
NQO1	mRNA	Qiagen cat no. QT00050281
MIR200A	miRNA	Qiagen cat no. MS0003738
MIR200B	miRNA	Qiagen cat no. MS00009016
MIR200C	miRNA	Qiagen cat no. MS00003752

2.6. Oxygen consumption rate (OCR) and extracellular acidification rate (ECAR)

$2*10^3/2.4*10^3$ cells ($FH^{+/+}/FH^{-/-}$) were seeded into XFe96 cell culture microplate in 100 µL standard culture media and incubated overnight. Cells were then washed with 1x PBS and 675 mL of bicarbonate-free DMEM (Sigma-Aldrich, D5030) supplemented with 25 mM glucose, 1 mM pyruvate, 4 mM glutamine and 1% v/v FBS was added. To eliminate residues of carbonic acid from the medium, cells were incubated for 30 min at 37°C without CO_2 in the BioTek Cytation 1/5 (BioTek) and brightfield images where taken during this incubation. OCR and ECAR were assayed in a Seahorse XFe96 Analyzer (Seahorse) by a measurement cycle of 2 min mix, 2 min wait, and 4 min measure at basal condition. The last injection included 100µM Hoechst (Thermo, cat.no 33342) in the injection port. After 30 min of incubation fluorescence images were taken using the BioTek Cytation 1/5 and images integrated into Wave software (Agilent) for OCR and ECAR value normalisation by cell count.

2.7. Respiratory chain complex measurements

$10*10^6$ cells were harvested by trypsinisation and washed with 1xPBS (4°C). Crude mitochondria were extracted on ice using 1.5 mL/$6*10^6$ cells of isotonic buffer (containing 0.25 M sucrose, 10 mM Tris–Cl, pH 7.5, and 0.1 mM phenylmethylsulfonyl fluoride) and homogenised using a glass teflon homogeniser. Unbroken cells and nuclei were removed by centrifugation at 3000g for 10 min at 4°C. The supernatant containing the mitochondria was centrifuged at 13000g for 20 min at 4°C and the pellet was suspended in isotonic buffer and aliquots stored at −80°C. Protein quantification was performed as described in 2.4. “Western Blotting”.

To measure the native respiratory complex activity crude mitochondria dissolved in a 1 ml complex measurement solution (for details see table below) were measured at 37°C by using a

dual-wavelength spectrophotometer (V550 Jasco Europe, Italy) using the materials and protocol described in Ghelli et al[117].

Complex measurement solution:

Buffer	Components	Concentration	pH
A	KH_2PO_4	50 mM	7.6
	EDTA	1 mM	
	BSA	3.5 mg/ml	
B	KH_2PO_4	50 mM	7.6
	EDTA	1 mM	
	BSA	2.5 mg/ml	
C	Tris	100 mM	8.1
	Triton X-100	0.1%	

Complex	Components	Concentration	Details	Measurement settings
CI	DCIP	60 µM	Mastermix	λ: 600 nm εDCIP: 19.1/mm/cm Temp.: 37°C F: 52.3
	Antimycin A	1 µM		
	KCN	0.3 mM		
	NADH	200 µM		
	Sample	10 µM		
	Buffer A	Ad. 1 ml		
	Rotenone	1 µM	Repeat +/- rotenone	
	DB	70 µM	Start reaction	
CII	DCIP	80 µM	Mastermix, incubate 1 min at 37°C	λ: 600 nm εDCIP: 19.1/mm/cm Temp.: 37°C F: 52.3
	Antimycin A	1 µM		
	KCN	0.3 mM		
	Rotenone	1 µM		
	ATP	200 µM		
	Na-succinate	10 mM		
	Sample	20 µM		
	Buffer A	Ad. 1 ml		
	DB	50 µM	Start reaction	
	K-malonate	5 mM	Stop reaction	
CIII	KCN	0.3 mM	Mastermix	λ: 550 nm εDCIP: 19.1/mm/cm Temp.: 37°C F: 52.6
	Cytochrome C oxidised	20 µM		
	Sample	10 µM		
	Buffer B	Ad. 1 ml		
	DBH_2	50 µM	Start reaction	
	Antimycin A	1 µM	Stop reaction	

CIV	Cytochrome C reduced	20 µM	Mastermix	λ: 550 nm εDCIP: 19.1/mm/cm Temp.: 37°C F: 52.6
	Buffer B	Ad. 1 ml		
	Sample	10 µM	Start reaction	
	KCN	0.3 mM	Stop reaction	
CS	DTNB	100 µM		λ: 360 nm εDCIP: 13.6/mm/cm Temp.: 37°C F: 73.5
	Acetyl-CoA	300 µM		
	Sample	20 µM		
	Buffer C	Ad. 1 ml		
	Oxaloacetate	50 µM	Start reaction	

2.8. Metabolomics: Liquid chromatography–mass spectrometry (LC-MS)

2.8.1. LC-MS sample preparation

Steady-state metabolomics and carbon tracing experiments were performed using liquid chromatography–mass spectrometry (LC-MS) analysis. $5*10^4$ cells ($FH^{+/+}$) and $6*10^4$ cells ($FH^{-/-}$) were seeded in 6-well plates and incubated overnight for 60% confluency. Afterwards, the media of choice was supplemented with stable isotope labelled substrates and added for 24 h. The cells were placed at ambient oxygen tension (21%) in a standard cell culture incubator or at 1% oxygen in a hypoxia chamber (H35 Hypoxystation, Whitley). For carbon tracing experiments in the standard culture media either 580 mg/ml U-^{13}C-glutamine (Cambridge Isotope Laboratories, cat. no. CLM-1822-SP-PK) was added to glutamine free DMEM (Life Technology cat. no. 21969-0.35) supplemented with 10% v/v FBS, or 4500 mg/ml U-^{13}C-glucose (Cambridge Isotope Laboratories, cat. no. CLM-1396-MPT-PK) and 110 mg/l sodium pyruvate were added to glucose and pyruvate free DMEM (Life Technology cat. no. 11966-025) with 10% v/v FBS. At the experimental endpoint the cells had 80% confluency and one well from each condition was used to estimate cell number and cell volume (CASY cell counter, Innovatis). For consumption-release experiments the growth factor was estimated by additionally counting one well from each condition at the start of the experiment.

2.8.2. LC-MS sample extraction

Intracellular metabolites were extracted by washing the wells 2x with 1xPBS (room temperature), placing the plate into a cooling bath (dry ice and methanol) and adding 1 ml of metabolite extraction solution (50% v/v methanol, 30% v/v acetonitrile, 20% v/v ddH2O, with 5 µM d8-valine as the internal standard) per $1*10^6$ cells followed by a 15 min incubation.

Afterwards, cells were scraped and transferred into Samples were incubated and mixed in an Eppendorf Thermomixer at maximum speed for 15 min at 4°C. Eppendorf tubes and incubated at -20°C to further precipitate proteins. Subsequently after, samples were centrifuged at 16000g for 10 min at 4°C and the supernatant transferred into autosampler vials and stored at -80°C until further analysis. Cell culture medium (extracellular) extracts were prepared by adding 750 µL of extraction solution to 50 µL of centrifuged cell culture medium, and further processed as described for the intracellular extracts. For the derivatisation of D-2HG and L-2HG, the protocol of Quing-Yun Cheng et all was followed[118]. In brief, after following the standard extraction protocol for intracellular metabolites as described above, 200 µl of each sample was dried using the speedvac (Savant SPD121P with the Universal Vacuum System UVS400A). Afterwards, the pellet was resuspended in 160 µl derivatisation buffer (1.25 mmol/L N-(p-Toluenesulfonyl)-L-phenylalanyl chloride (TSPC) in ACN plus 2µl Pyridine), mixed in an Eppendorf Thermomixer at maximum speed for 30 min at RT, and dried using the speedvac. Lastly, the pellet was resuspended in 100 µl metabolite extraction solution by sonication, centrifuged at maximum speed for 15 min at 4°C, and supernatant transferred into autosampler vials for analysis.

2.8.3. LC-MS sample measurement

Sample measurement and peak integration was performed by the LC-MS facility of the Frezza Laboratory (Ana Sofia Henriques da Costa, Laura Tronci, Efterpi Nikitopoulou and Ming Yang). In detail, samples were randomised to avoid bias due to machine drift and processed blindly with an injection volume of 5 µl. Pooled samples were generated from an equal mixture of all individual samples and analysed interspersed at regular intervals within sample sequence as a quality control.

HILIC chromatographic separation of metabolites was achieved using a Millipore Sequant ZIC-pHILIC analytical column (5 µm, 2.1 × 150 mm) equipped with a 2.1 × 20 mm guard column (both 5 mm particle size) with a binary solvent system. Solvent A was 20 mM ammonium carbonate, 0.05% ammonium hydroxide; Solvent B was acetonitrile. The column oven and autosampler tray were held at 40 °C and 4 °C, respectively. The chromatographic gradient was run at a flow rate of 0.200 mL/min as follows: 0–2 min: 80% B; 2-17 min: linear gradient from 80% B to 20% B; 17-17.1 min: linear gradient from 20% B to 80% B; 17.1-22.5 min: hold at 80% B. Metabolites were measured with a Thermo Scientific Q Exactive Hybrid Quadrupole-Orbitrap Mass spectrometer (HRMS) coupled to a Dionex Ultimate 3000 UHPLC. The mass spectrometer was operated in full-scan, polarity-switching mode, with the spray voltage set to +4.5 kV/-3.5 kV, the heated capillary held at 320 °C, and the auxiliary gas heater held at 280 °C.

The sheath gas flow was set to 25 units, the auxiliary gas flow was set to 15 units, and the sweep gas flow was set to 0 unit. HRMS data acquisition was performed in a range of m/z = 70–900, with the resolution set at 70,000, the automatic gain control (AGC) target at 1 × 106, and the maximum injection time (Max IT) at 120 ms. Metabolite identities were confirmed using two parameters: (1) precursor ion m/z was matched within 5 ppm of theoretical mass predicted by the chemical formula; (2) the retention time of metabolites was within 5% of the retention time of a purified standard run with the same chromatographic method.

The HPLC separation for TSPC derivatised D-2HG and L-2HG was performed on an Inertsil ODS-3 column (250 mm × 2.0 mm i.d., 5 µm, VWR) with a binary solvent system. Solvent A was 0.1% formic acid in water (v/v) and solvent B was 50:50 (v/v) ACN and MeOH. The chromatographic gradient was run at a flow rate of 0.200 mL/min as follows: 0–3 min: 30% B; 3-10 min: linear gradient from 30% to 70% B; 10-25 min 70% B; 25-26 min linear gradient from 70 to 30% B; 26-40 min: held at 30% B. Samples were randomised and analysed with an injection volume of 3 µl. The mass spectrometry detection was performed on a Thermo Scientific Q Exactive Hybrid Quadrupole-Orbitrap Mass spectrometer using a full-scan method in the negative mode with the spray voltage set to -3.5 kV. The heated capillary was held at 200 °C, and the auxiliary gas heater held at 250 °C. The sheath gas flow was set to 45 units, the auxiliary gas flow was set to 10 units, the sweep gas flow was set to 0 unit and S-lens RF level was set to 55. Data acquisition was performed in a range of m/z =135-500, with the resolution set at 17,500, the AGC target at 1 × 106, and the Max IT at 50ms.

Chromatogram review and peak area integration were performed using the software Tracefinder 5.0 (Thermo Fisher). Absolute quantification of selected intracellular metabolites was calculated by interpolation of the corresponding standard curve obtained from serial dilutions of commercially available standards running with the same batch of samples and by dividing it by the measured cell volume. For ^{13}C-tracing analysis, the theoretical masses of ^{13}C isotopes were calculated and added to a library of predicted isotopes. These masses were then searched with a 5 ppm tolerance and integrated only if the peak apex showed less than 1% difference in retention time from the [U-^{12}C] monoisotopic mass in the same chromatogram.

2.8.4. LC-MS data analysis

Metabolomics analysis was performed according to the workflow presented in **Figure 6**. In brief, mass isotopologue distribution of metabolites was determined by integration of the corresponding peaks and correction for natural abundance was performed using the Accucor Package[119] (v.0.2.3) and the fractional enrichment was visualised using stacked bar graphs. The total pool of metabolites was obtained by applying the “Filtering 80% Rule”, half minimum

missing value imputation, and normalising the samples using total ion count normalisation. Sample were excluded after performing principal component analysis (PCA) and testing for

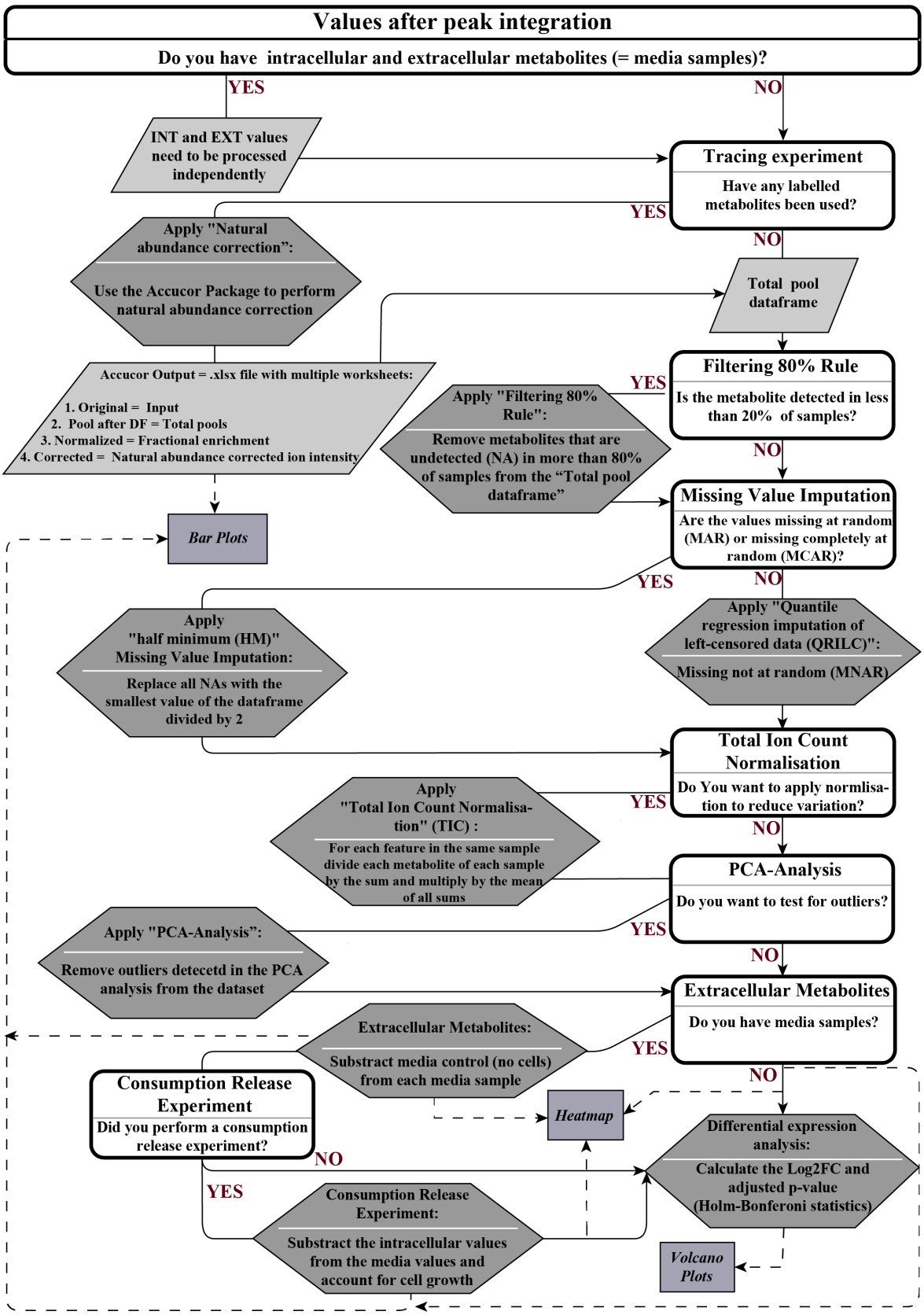


Figure 6: Flowchart depicing the processing of metabolomics data.

outliers based on geometric distances of each point in the PCA score plot as part of the muma package[120] (v1.4). Afterwards, differential expression analysis was performed. The results were visualised as bar graphs, heatmap, or volcano plot. In case of consumption-release experiments the pre-processed total pool or the fractional enrichment of each metabolite detected in the fresh culture medium (incubated in the absence of cells) were subtracted from the metabolites detected in the media samples. This will result in the extracellular metabolite levels that were visualised using bar graphs (total pool) or stacked bar graphs (fractional enrichment). In order to understand the metabolic flux, the cell volume measured at the timepoint of harvest (CASY cell counter, Innovatis) and the cellular growth rate or at least the cellular growth factor has to be taken into account.

2.9. Proteomics

$1*10^6$ cells for FH-proficient and $1.2*10^6$ cells for FH-deficient clones were seeded on a 10 cm culture dishes and incubated for 24 h until ~80% confluency was reached. For the harvest, cells were washed twice with 1xPBS and scraped in 500 µl, 4° C 1x PBS supplemented with protease- and phosphatase inhibitors. Pellet was obtained by 4 min centrifugation at 3600 g at 4° C, snap frozen, and stored at -80° C until shipment.

2.9.1. Label-free proteomics

Label-free proteomics measurement was performed by Dr. Alexander von Kriegsheim (University of Edinburgh, HTPU and Mass Spectrometry Facility Manager). First, the cell pellets were lysed in 50 µl buffer comprised of 200 mM Tris pH 8.5, 10 mM Tris (2-carboxyethyl) phosphine (TCEP), 20 mM Chloroacetamide and 6 M Guanidine-HCl. Each sample was sonicated for 10 s and then incubated at 95°C for 5 min. Protein concentration was determined using a modified Bradford assay (Pierce). Proteins were digested for 4 h at 1/200 substrate/enzyme with LysC (Wako) then diluted 1/10 with water and digested overnight at 37° C with MS grade trypsin (90057, Thermo Scientific) at protease/protein ratio of 1/100. Samples were acidified to 1% TFA final volume and clarified by spinning on a benchtop centrifuge (15k g, 5 min). Sample clean up to removes salts was performed using C18 stage-tips[121]. Samples were eluted in 25 µl of 80% Acetonitrile containing 0.1% TFA and dried using a SpeedVac system at 30°C. Samples were resuspended in 0.1% (v/v) TFA such that each sample contained 0.2 µg/ml. All samples were run on an Orbitrap Fusion™ Lumos™ Tribrid™ or QExactive plus™ mass spectrometer coupled to a uHPLC (Ultimate 3000, RSL-Nano). 5 ul of the samples was injected onto an emitter packed with C18 material (35 cm, 75um ID 360 OD packed with UChorm 1.8um) and heated to 50C. Peptide were separated by a 150 min

gradient from 5-40% Acetonitrile in 0.5% acetic acid. Data was acquired as data dependent acquisition with following settings. MS resolution 240 k, cycle time 1 s, MS/MS HCD ion-trap rapid acquisition, injection time 28 ms (for Lumos) or MS resolution 70k, MS/MS 15k, top 24, injection time 45ms (for QExactive plus).

The data was searched using the MaxQuant 1.6 software suite (https://www.maxquant.org/) by searching against the human Uniprot database with the standard settings enabling LFQ determination and matching. The data was analysed using the Limma-pipeline[122].

Using this data, I have performed gene set enrichment analysis (GSEA) by recapitulating on the gene sets published on the MsigDB[82] using the packages fgsea[123] (v1.8.0) and GSEABase[124] (v1.44.0). The EMT genset was generated by manually curating the gene list published by Taube et al[125]. Plots were generated using the EnhancedVolcano package[126] (v.1.0.1).

2.9.2. Succination TMT proteomics

Succinated protein residues were measured using TMT proteomics by Eva Papachristou and Clive D'Santos (CRUK Proteomics Core Facility, Cambridge, UK) according to their standard protocol published in Papachristou et. al[127]. The database search was performed as previously described in Petros et. al[55].

The statistical analysis was performed by Chandra Sekhar Reddy Chilamakuri (CRUK Bioinformatics Core facility, Cambridge, UK). In detail, peptide intensities were normalised using median scaling and protein level quantification was obtained by the summation of the normalised peptide intensities. A statistical analysis of differentially-regulated proteins was carried out using the qPLEXanalyzer, R-bioconductor package[127]. Multiple testing correction of p-values was applied using the Benjamini-Hochberg method[128] to control the false discovery rate (FDR).

Using this data, I have pulled down the information about the position of the modified cysteine residue within the protein using uniprot peptidesearch (https://www.uniprot.org/peptidesearch/). Next, the differential-succinated proteins (Log2FC ($FH^{-/-}$ versus $FH^{+/+}$)) were normalised for the differential total protein expression by substracting the Log2FC of the total proteome from the Log2FC of the corresponding succinated protein residue. To check the change induced by this normalisation step I performed a correlation analysis ($R^2 = 0.97$). Lastly, these succinated cysteine residues were compared to previously reported cysteine residues in human and cysteine residues in mouse that are conserved in human[129].

2.10. RNA sequencing

The mRNA was extracted using RNeasy Kit (Qiagen) following the manufacturer's instructions. Quantification of the RNA was done using the Qubit RNA broad range assay kit (Invitrogen, cat.no. Q10210) and the Qubit Fluorometer 3.0 (Invitrogen). The samples were prepared in Qubit assay tubes (Invitrogen cat.no. Q32856) according to the manufacturer's instructions. The RNA integrity number (RIN) was determined using the D2200 ScreenTape system (Agilent) running the manufacturers RNA protocol using the RNA ScreenTape (Agilent, cat.no. 5067-5576). Samples for the "RNA ScreenTape" assay were prepared according to the manufacturer's instructions. Only with samples that had RIN values above 8.5 was proceeded to library preparation. For RNAseq polyA capture library preparation 550 ng total RNA (10 ng/µl in ddH_2O) was submitted in a randomised plate layout to the CRUK Genomics Core facility (Cambridge, UK) for library preparation and sequencing. Library preparation was done using the Trueseq stranded mRNA kit (Illumina) and single-read sequencing was preformed using the HiSeq 4000 (Illumina). The first sequencing run from CRUK Genomics Core facility produced low read coverage, as such they performed a second run which resulted in a read depth of approximately 5 million.

The downstream sequencing alignment, QC and conversion to count data was done by Ariane Mora (visiting PhD-student in the Frezza Lab). Quality control (QC) was performed using FastQC (v0.11.9, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and MultiQC[130] (v1.9). Since overrepresented sequences (UD indexes of the Illumina-TruSeq kit) were found, trimming was performed using Cutadapt[131] (v2.10). Adapters were removed from the reverse strand with minimal error tolerance and removing poor quality bases below pfred score of 20 (-g AGATCGGAAGAGCACACGTCTGAACTCCAGTCA -e 0.05 -q 20,20). Afterwards, FastQC and multiQC was re-run, which showed lower over-representation with a shorter sequence affecting 0.1% of sequences which was considered sufficient to pass QC. The reads were mapped to the human Hg38.18 genome using Hisat2[132] (v2.2.1) using at most 5 distinct reads (k parameter), allowing a maximum restart of 5 times (--max-seeds and -k 5 -p 10 -q --max-seeds 5 --rna-strandness "R"). As noted above, there were two runs from the sequencer, these bam files were merged and sorted using "merge" and "sort" from samtools[133] (v1.10-88). To summarise the counts to annotated genes, FeatureCounts[134] (v2.0.1) from subread was run also using the annotation from GRCh38.100 human genome. In detail, primary reads are counted to exons (-F GTF -t exon -T 10 -s 2 -g gene_id –primary).

Next, I used the Deseq2 (v1.22.2) package[135] to compute differentially expressed genes between the two conditions ($FH^{-/-}$ versus $FH^{+/+}$) applying Walds statistics and setting the significance level "alpha" for independent filtering to 0.05. Prior to running DEseq2, genes with fewer than five reads were removed. Gene IDs from Ensembl were mapped to gene names using scibiomart[136] (v.1.0.2), a wrapper around the API from BioMart. In case multiple ids were mapped to a gene, the gene with the highest log2 fold change was retained. To determine the functional enrichment of the change in mRNA expression, GSEA was performed on the gene sets published on the MsigDB[82] using the packages fgsea[123] (v1.8.0) and GSEABase[124] (v1.44.0). The EMT geneset was generated by manually curating the gene list published by Taube et al[125]. Plots were generated using the EnhancedVolcano package[126] (v.1.0.1).

2.10.1. Alternative isoform regulation

While featureCounts provides details of changes in expression that occur at the gene level, to perform isoform analysis, we used StringTie[137–139] to count the reads mapping to different transcripts (https://ccb.jhu.edu/software/stringtie/). The merged and sorted Bam files (details in 2.10 RNA sequencing) were mapped to the GRCh38.100 human genome using StringTie (v2.1.4) with parameter -e for downstream use in IsoformSwitchAnalyzeR. Using the -e parameter, the primary output of StringTie are tab files which contain details of the assembled isoforms from the RNA-Seq data. IsoformSwitchAnalyzeR uses these outputs from StringTie to compute the differential usage of isoforms[140–144]. We used the function isoformSwitchTestDEXSeq with alpha 0.05 and dIFcutoff of 0.1 to quantify the mean change and significance between $FH^{-/-}$ and $FH^{+/+}$ at the transcript level.

2.11. Bisulfite sequencing

This experiment was done in collaboration with the Massie Lab (Oncology Department at the University of Cambridge) and performed by Sabrina Rossi (PhD-student) and Sara Pita (Research assistant). DNA samples (10 ng/µl, 500 ng total) were sheared by S220 Focused-ultrasonicator (Covaris). The mean dsDNA fragment size of 180-200 bp was checked using the Agilent D1000 ScreenTape System (>60% of DNA fragments were between 100 and 300 bp). Tissue methylation analysis of a capture-based method targeting 3 million CpGs was performed using the TruSeq Methyl Capture EPIC Library Preparation Kit (Illumina) according to the manufacturer's instructions. Sequencing was performed using the HiSeq4000 Illumina Sequencing platform (single end 150bp read) using two lanes per library pool. Technical replicates were performed for cell line data to assess assay reproducibility (R2 = 0.97). Quality

control (QC) was performed using FastQC and MultiQC. The reads were trimmed using TrimGalore (v0.4.4) and Cutadapt (v1.8.1) using standard parameters. To extract DNA methylation, Bismark (v0.22.1) was used with the bowtie aligner (v2.3.4.1) and the parameters: --ignore_3prime 1 –ignore.

The downstream methylation analysis was performed in collaboration with Ariane Mora (visiting PhD-student in the Frezza Lab) (**Figure 7**). First, MethylSig[145–148] (v1.0.0) was used to identify differentially methylated regions (DMRs) in the promoter region of known genes from the TxDb.Hsapiens.UCSC.hg38 database. Loci with less than 5 counts in either $FH^{+/+}$ or $FH^{-/-}$ were removed prior to running the DMR analysis. The function "diff_methylsig" was called on the filtered data, and an approximate squared t was used for the likelihood ratio statistic. Next, differentially methylated cytosine (DMC) analysis on the resolution of CpGs was performed using the MethylKit[149] (v1.14.2). Coverage files were read in using the GRCh38 genome and the bismark coverage pipeline at the base-pair resolution, using CpG methylation context with a minimum coverage set at 10. To improve the test statistic, CpGs were filtered and removed when coverage was considered "very high" (top 0.1% of CpGs) or "very low" (counts < 10). The function "calculateDiffMeth" (default parameters) was used to calculate

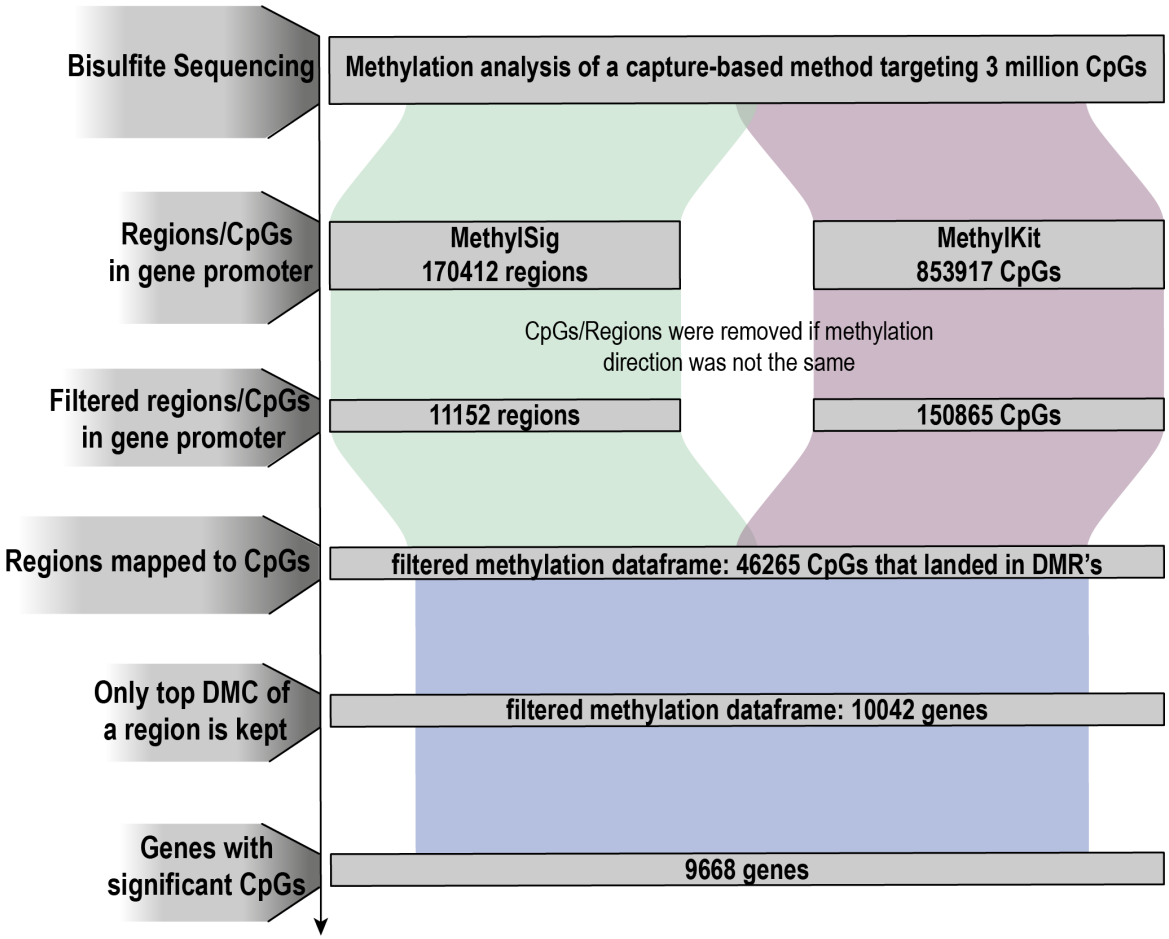


Figure 7: Alluvian plot depicting the methylation analysis workflow.

differential methylation at each CpG. Coverage, number of C's, number of T's, and percentage (number of C's/ coverage) were recorded for downstream analysis. The DMRs from MethylSig and differential CpGs from MethylKit were annotated to genes if they fell within 1000 base pairs of the Transcriptional start side (TSS) or overlapped the gene body by 500 base pairs using sciepi2gene[150] (v1.0.0). The results from MethylSig and MethylKit were annotated to genes and merged using scidmg[151] (v1.0.0) with the default parameters. In detail, only significant (q<=0.1) DMRs with strong concordance to the DMCs (q < 0.1) in terms of direction of methylation change (> 60%) were retained. If multiple DMR's mapped to a single gene yet were not concordant in terms of the directionality of change then these were removed. For the remaining DMRs the CpG with the highest methylation difference in the direction of change was assigned as the methylation value (change and padj) for that gene (i.e. as the driver CpG behind the gene's change in DNA methylation).

2.12. SiRCle (Signature Regulatory Clustering) model

This tool was developed in collaboration with Ariane Mora (visiting PhD student).

Depending on which level the gene expression is influenced, genes were clustered into regulatory groups. Regulatory clusters were created using scircm v1.0.0 (unpublished) using the following parameters: rna_padj_cutoff=0.05, prot_padj_cutoff=0.05, meth_padj_cutoff=0.05, rna_logfc_cutoff=0.5, prot_logfc_cutoff=0.3, meth_diff_cutoff=10, bg_type='P|(M&R)'. In detail, genes were grouped by less than, or greater than the amount with one level of redundancy (i.e. only had significant changes in two of the three groupings). The only exemption from this rule are genes encoding for non-coding RNA and genes measured at the protein level. In these cases, significant changes in one of the three groups were sufficient to be assigned to a cluster. This resulted in 24 different flows that were driven by 10 different regulatory drivers (e.g. mRNA increase) resulting in 10 individual regulatory clusters. Non-coding gene annotations were downloaded from GO (http://geneontology.org/) by searching for "Non-coding" in the genes and products section, filtering organism for Homo-sapiens and choosing "Gene/product (bioentity)", "Source" and "Type" for customised download.

2.12.1. Over representation analysis of the SiRCle clusters

Over representation analysis (ORA) was performed on each SiRCle clusters using the "enrich_GO" function (parameters: keyType = "ENTREZID", OrgDb = org.Hs.eg.db, ont = "ALL", pAdjustMethod = "BH", qvalueCutoff = 0.1) of the clusterProfiler package[152] (v3.10.1).

The background data as defined in the scircm package “bg_type” were used as the “universe” in the “enrich_GO” function, these were all genes that either have a significant change in protein, or a significant change in RNA or DNA methylation (bg_type='P|(M&R)'). The output data are plotted using the “emapplot” function or the “heatplot” function of the clusterProfiler package.

2.12.2. Overlaying other omics data onto the SiRCle clusters

To overlay other omics data, the Fisher’s exact test was used to quantify the likelihood of the overlaid data falling onto specific regulatory clusters. This analysis's input data are the results from scircm and the filtered (e.g. for adjusted p-value, Log2FC) omics data that are overlaid. Here we overlay genes with significant changes in succination (calculated using succination proteomics where succination may occur at any site, p <= 0.05). Similarly, genes which were found to be alternatively spliced from the isoform analysis (2.10.1. Alternative isoform regulation) were also overlaid.

2.12.3. Transcription factor (TF) analysis based on the SiRCle clusters

To identify potential binding sites of motifs, we downloaded the upstream flanking sequences from Ensembl (https://www.ensembl.org/biomart/martview/, human genes GRCh38.p13) for the genes in the background dataset of our SiRCle model (+100bp, unique results only). Motif position weight matrices (PWM‘s) were downloaded in meme format from the HOCOMOCO database (https://hocomoco11.autosome.ru/), using the core dataset from v11. We used fimo from the online version of meme suite (v1.3.0, http://meme-suite.org/tools/fimo) to find transcription factor (TF) binding sites enriched in our background dataset (Ensembl sequences and HOCOMOCO PWMs as input). The output tsv file was used as the input to scimotf[153] (v1.0.0), which in turn identifies common TFs in each SiRCle cluster. We used the following parameters for scimotf: RNASeq padj, RNAseq Log2FC, tf_in_dataset=TRUE, fimo_pcol="p-value", cluster_pcutoff=0.05). The results were filtered to only contain TFs that were uniquely assigned to one SiRCle cluster.

2.12.4. Integration of VAE into SiRCle

To assign ranks to genes in SiRCle regulatory groups, we use a Variational Autoencoder (VAE) to learn an integrated rank for each gene. There were 2805 genes assigned to SiRCle clusters and these were used as the training set of genes. The input values for each gene was the methylation difference, Log2FC for RNAseq and proteomics data, and also three replicates of

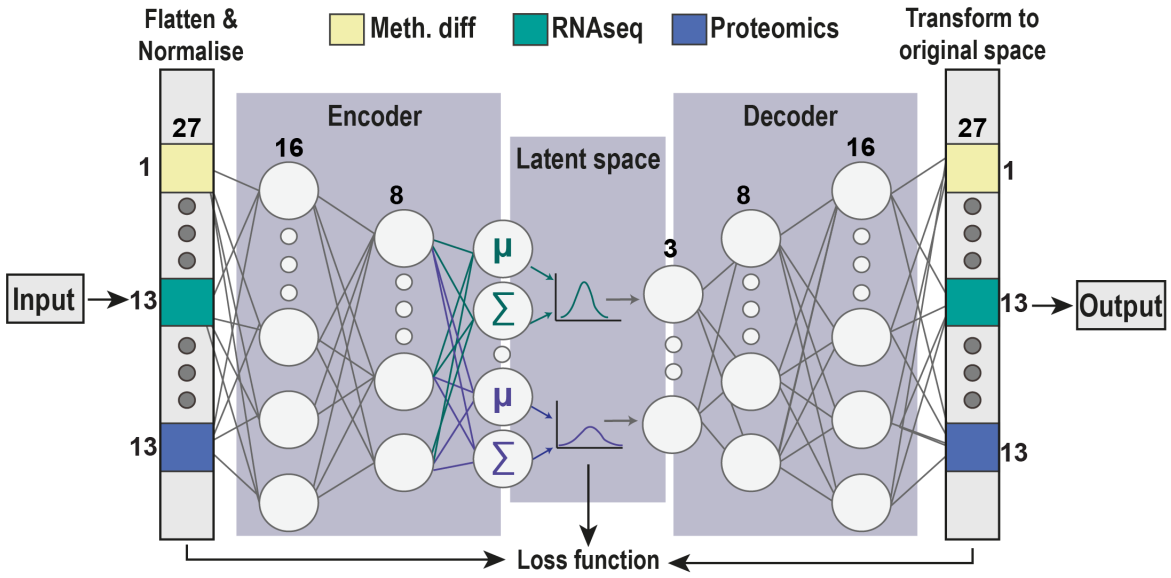


Figure 8: The VAE model.

The VAE is a machine learning approach that focuses on combining different data types before applying a learning algorithm to get a common latent representation of the data. The data input are the methylation difference (yellow), RNAseq data (green) including Log2FC and three replicates of the normalised values of the single clones, and proteomics data (blue) including Log2FC and three replicates of the normalised values of the single clones, which results in a total of 27 different input features. In the encoder the input features are reduced into a smaller number of nodes to get a compressed representation into the latent space. Within the latent space the VAE tries to learn the distribution of our data (latent variable σ and ∑). The decoder is used to decode the latent variables and approximates the observed variable given to our latent space. After the decoding the loss will be calculated and in a feedback loop propagated back to ensure as less loss of information as possible (loss function). This regularisation applied to the latent space is maximum mean discrepancy (MMD), which takes into account the clonal differences and ensures the best fit of the data without overfitting. In the end each gene will go through this projection method and the VAE will project the data into a unique position in the latent space on three nodes corresponding to x, y, z planes taking into account all our input features.

the normalised values of the single clones for RNAseq (normalised using EdgeR's (v3.30.3, https://rdrr.io/bioc/edgeR/man/calcNormFactors.html, TMM method) and proteomics (in total 27 features, **Figure 8**). We then used the rcm package (v1.0.0, unpublished) to compute the VAE (scivae v1.0.0, unpublished) using three nodes in the latent space, with selu activation functions and MMD loss metric. The rcm package assigned three ranks to each gene, one for each node in the latent space. These ranks were used to run GSEA using fgsea[123] (v 1.14.0, 10000 permutations, „simpleGSEA" function) and GSEABase[124] with the publicly available gene sets from the Molecular Signatures Database (MsigDB)[82], namely KEGG, Biocarta, Rectome, hallmarks, GO and NEF2L2, as well as the manually curated EMT signature based on the gene list published by Taube et al[125].

2.13. TRACER

2.13.1. Fabrication

The 3D cell culture system TRACER requires different components to enable assembly of the model[116]: the biocomposite seeding mould, the cell spreader for monoculture, the mandrel, the mandrel handle, the tissue securing clip, the 6-well inserts and the 6-layer scaffold strips, which all were manufactured together with the Laboratory of Alison McGuigan (University of Toronto, Canada) using the material and procedures described by Rodenhizer et al[116]. Before usage, the biocomposite seeding mould, the cell spreader for monoculture, the mandrel handle, the tissue securing clip and the 6-well inserts were sterilised by wet autoclaving. The mandrel, the 6-layer scaffold strips and standard paper tissue were sterilised by backing around 170°C overnight.

2.13.2. Seeding, assembly and disassembly

The TRACER seeding was adapted and optimised based on Rodenhizer et al.[116]. During this process, the best cell number, reproducibility of cell density, and homogeneous distribution were assessed using the SRB staining protocol described below (2.13.3. SRB staining). Before harvesting the cells, the biocomposite seeding chamber was assembled by placing autoclaved paper tissue into a 500 cm^2 (Nunc, cat.no.166508) culture dish, soaking them with 1x PBS and humidifying it by incubating at 37 °C in the cell culture incubator (**Figure 9**, Step 1). At the same time, the heating pad and 6-well plates filled with 10 ml culture media are placed in the incubator and 1x PBS is heated to 37°C. The $FH^{+/+}$ and $FH^{-/-}$ clones were harvested by trypsinisation, filtered through a cell strainer (45 µm, Corning), counted (CASY cell counter, Innovatis) and centrifuged for 4 min at 3600 g at room temperature (**Figure 9**, Step 2). In the meanwhile, the collagen hydrogel was prepared by mixing 1 ml PurCol (Sigma, cat.no C7657) with 125 µl 10x MEM (1 pot from Gibco, cat. no 194388, supplemented with 22 g $NaHCO_3$, ad. 1 l ddH_2O) in a 2 ml Eppendorf tube until the solution turns from pink to yellow. Next, 60 µl of 0.1 N NaOH are added to neutralise the pH, which turns the solution pink again. The collagen hydrogel was stored on ice. Depending on the number of strips needed for the experiment, all clones were processed together or if more than a total of 6 strips where required, the step 3-7 (**Figure 9**) were repeated for each clone and the cell suspensions were stored at 37°C until usage. The cell pellet obtained after centrifugation was resuspended in the collagen solution ($50*10^6$ cells/ml $FH^{+/+}$ or $60*10^6$ cells/ml $FH^{-/-}$, avoid air bubbles), transferred to a 2 ml Eppendorf tube and placed on ice (**Figure 9**, step3). The biocomposite seeding mould was placed onto ice and 90 µl of the collagen-cell-suspension was distributed along one lane

covering the total length of a strip (**Figure 9**, step 3). After all lanes are filled, the cell suspension is spread evenly within the lanes using a cell spreader (**Figure 9**, step 4). It is important to use a clean cell spreader for each cell type that is seeded. The nitrocellulose scaffold strip is placed onto the lane using tweezers, and the cell suspension will soak in (**Figure 9**, step 5). After all strips are placed on the lanes, they are pressed down using the cell spreader to ensure an equal distribution of the cell suspension across the strip (**Figure 9**, step 6). For the collagen to solidify without drying out, the strip is placed in the warm, humidified

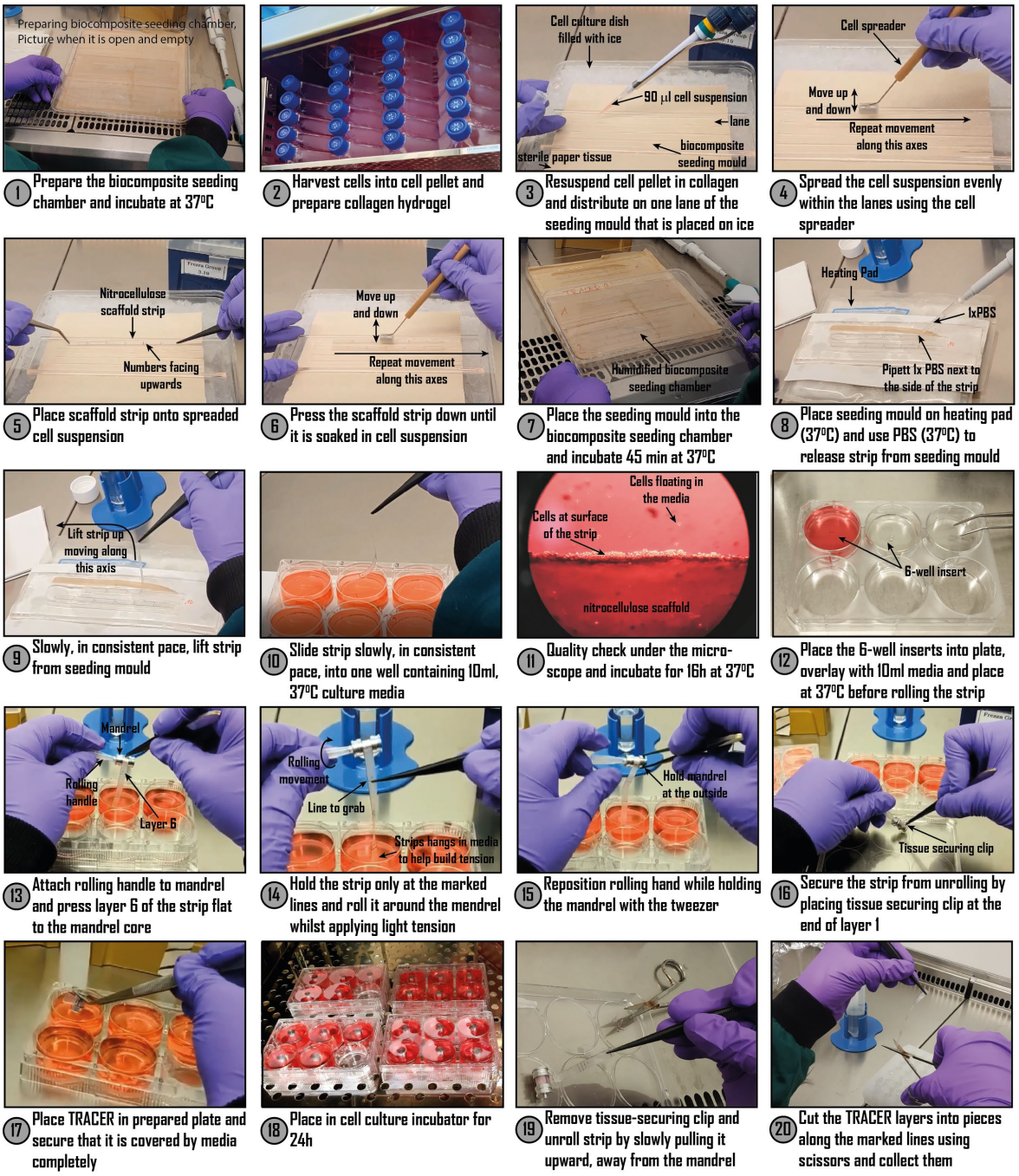


Figure 9: Key TRACER seeding, assembly and disassembly steps. The detailed description for the steps 1-20 can be found in the text.

seeding chamber and incubated for 45 min at 37°C in the cell culture incubator (**Figure 9**, step 7). Here it is crucial to keep on time to ensure reproducibility between the batches. Next, the seeding mould was placed on a heating pad (37°C) and 1x PBS (37°C) is carefully pipetted along the edges of the strip until it is completely soaked and swims off the seeding mould (**Figure 9**, step 8). The strip was picked up using tweezers in a slow, consistent pace to minimise the number of cells that remain on the seeding mould (**Figure 9**, step 9). The slide was immediately slided into the prewarmed culture media and similar to the previous step it is important to keep a slow consistent pace to reduce the loss of cells (**Figure 9**, step 10). This can be checked under the microscope by looking at the standing strip. At the side that was faced down onto the seeding mould the cells can be observed and should build an even layer (**Figure 9**, step11). Moreover, floating cells that have been lost when sliding the strip into the culture media are visible (this should be a small number as indicated in the picture; otherwise the strip was excluded from downstream analysis). For the cells to attach and establish cell-to-cell contacts, the strip was incubated overnight, precisely for 16 h. After this incubation, the strips can be used for experiments that do not require the gradient of oxygen and nutrients as for instance, the growth curve as described in 2.13.3. "SRB staining". Otherwise, the 6-well inserts are placed in a 6-well plate, overlaid with 10 ml cell culture media and incubated at 37°C to warm up (**Figure 9**, step 12). Next, the rolling handle is attached to the mandrel and layer 6 facing upwards is placed at the carve of the mandrel (**Figure 9**, step 13). The strip is rolled around the mandrel by applying settled tension using the tweezers that grab the strip at the lines (only at the lines and nowhere else) and turning the mandrel using the rolling handle (**Figure 9**, step 14). An important quality control for the right tension and rolling is that after each full turn of the mandrel, the line on the strip should overlay exactly onto the previous line. During this process, the rolling hand need to reposition, which is done by holding the mandrel with the tweezer (**Figure 9**, step 15). Once the strip is completely rolled around the mandrel, the end is secured from unrolling by placing the tissue securing clip at the end of layer 1, where all lines come together (**Figure 9**, step 16). Next, the mandrel is held at the top, the rolling handle is removed, and the mandrel is stuck into the 6-well insert (**Figure 9**, step 17). The TRACERs are stored in the cell culture incubator for 24 h prior to harvest (**Figure 9**, step 18). For the downstream experiments, the TRACER is unrolled by removing the tissue securing clip and pulling on the end of layer 1 (**Figure 9**, step 19). Next the TRACER layers are cut into pieces using scissors and placed into 2ml Eppendorf tubes (**Figure 9**, step 20).

2.13.3. SRB staining

To assess the cell seeding uniformity within the strip (inter-batch variability), the reproducibility between strips (batch-to-batch variability), and the cell growth (growth curves), Sulphonamide B (SRB) staining was performed. In comparison to the MTT-assay (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazoliumbromide) described in Rodenhizer et al.[116], where the dye stains the cells based on the reduction by mitochondrial SDH and NADH/NADPH, the SRB stain can also be used for cells that harbour mitochondrial dysfunction (like FH-deficient cells) since it binds to the protein constituents of the cells in a stoichiometric manner.

To evaluate inter-batch and batch-to-batch variability the entire strip was placed in a 6-well plate containing 4 ml 4°C cold 0.1% trichloroacetic acid (TCA), and incubated for at least 1 h at 4°C. For the growth curve, the strips were cut into pieces after finishing step 11 (**Figure 9**), placed into a 24-well plate containing 2 ml cell culture media per well and stored in the incubator until harvest. The media was replenished every 24 h. At the harvest, the piece of

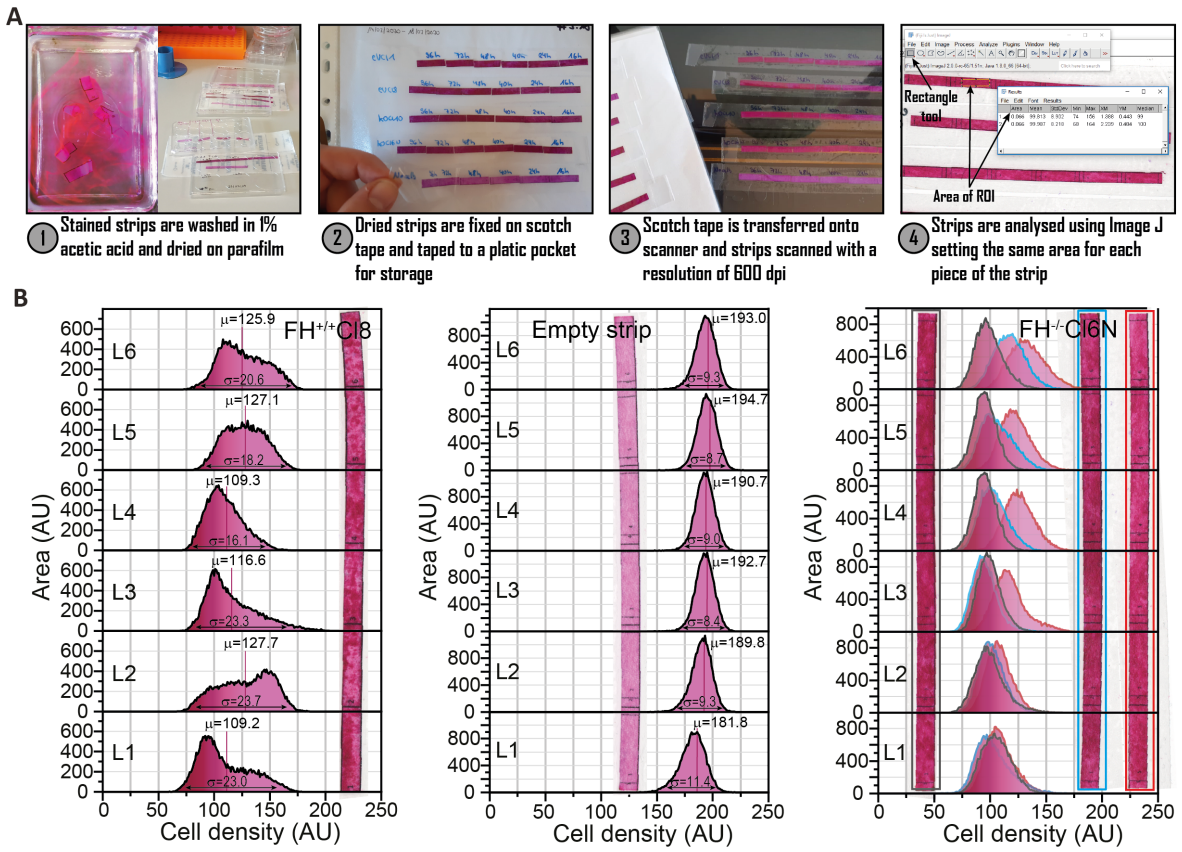


Figure 10: Key TRACER steps after SRB staining.

(**A**) The detailed description for the steps 1-4 can be found in the text. (**B**) Example of the quantification showing a badly seeded strip (left), an empty strip with perfect distribution of the stain (middle) and the comparability of the batches seeded on different days (right).

interest was placed in a 24-well plate containing 1 ml 4°C cold 0.1% TCA, and incubated for at least 1 h at 4°C. Once the TCA was removed the plate was stored at 4°C until staining. Noteworthy, for the growth curve, all pieces were harvested and stained together. For the staining, SRB-solution (0.057% SRB in 1% acetic acid) was added to the strip/pieces and incubated at RT for 1 h on a shaking platform. Afterwards, the strips/pieces were washed in 1% acetic acid and put to dry on parafilm (**Figure 10A**, step 1). The dried strips/pieces where stored with scotch tape fixed on a plastic pocket (**Figure 10A**, step 2). To scan the pieces, the scotch tape was fixed to the scanner and images were saved as TIF with 600 dpi resolution (**Figure 10A**, step 3). Image analysis was performed using Image J following the published protocol[116]. To ensure that the measurements are comparable between scans, the region of interest (ROI) was always set to an area of 0.66 (**Figure 10A**, step 4). The downstream analysis was performed as described previously[116] and showed that our seeding is uniform and can be used for cell quantification (**Figure 10B**).

2.13.4. Isolation of live cells from the TRACER strip

To perform downstream experiments such as cell counting, fluorescence-activated cell sorting (FACS) or protein extraction, live cells have to be extracted from the strip. The previously described assay takes a long time and is based on digesting the collagenase the cells are embedded in[116]. To reduce this time, a new protocol was established that enables the extraction of live cells in 25 min when handling 4 strips (24 pieces) at a time. In detail, 800 µl TrypLE Express 1x (Gibco, cat.no 12605-010) are pipetted into 2 ml Eppendorf tubes, the TRACER roll is disassembled and cut into pieces that are added into the solution. The pieces are incubated for 15 min at 37 °C on a thermomixer shaking at 600 rpm. During this period the tubes are vortexed 4 times for 3 s. Afterwards, 200 µl FBS (37°C) is added to stop the digest, the strip is removed, and the cell mixture is centrifuged for 4 min at 4400 g at RT. If the cells are harvested for counting the centrifugation step is skipped and the cells are counted using the CASY cell counter (Innovatis). It is important to note that this protocol has been optimised for a cell density reached after 24 h of TRACER culture (which equals total amount of 40 h during which cells were embedded in the strip) and needs to be optimised for each time point and cell number used. The complete extraction of the cells from the strip is controlled by SRB staining and by comparing the pieces to an empty strip.

2.13.5. RNA extraction and qPCR

The pieces were harvested into 2 ml Eppendorf tubes containing 350 µl RLT lysis buffer (RNeasy Kit, Qiagen) according to **Figure 9**, step 20, and frozen at -80 °C for at least 1h. Next,

the TRACER pieces were removed, and the mRNA was extracted according to the description above (2.5 RNA extraction and quantitative PCR). Following the protocol of 2.5. "RNA extraction and quantitative PCR", cDNA was produced using 500 ng total RNA. qPCR was performed using TaqMan Fast advanced MasterMix (Thermo fisher, cat.no. 4444557) and duplexing Fam and Vic on a Step One Plus real-time PCR system (Life Technology). Experiments were analysed using the StepOne software 2.3 and the standard deviation was calculated using excel.

Housekeeping genes used for internal normalisation:

TBP_Vic	Thermo Fisher, cat.no. HS00427620_m1

Target genes:

EPCAM_Fam	Thermo Fisher, cat.no. HS00901885_m1
NQO1_Fam	Thermo Fisher, cat.no. HS01045993_g1
BNIP3_Fam	Thermo Fisher, cat.no. HS00969291_m1

2.13.6. LC-MS

Steady-state metabolomics and carbon tracing experiments were performed using LC-MS analysis as described above (2.8. Metabolomics: Liquid chromatography–mass spectrometry (LC-MS)). TRACER strips were prepared as described in Figure 9 and the strips were placed in DMEM or DMEM supplemented with stable isotope labelled substrates for 24 h (**Figure 9**, step 17). Afterwards, a TRACER from each condition was used to estimate cell number and cell volume as described above and 1 ml/25*10^4 cells metabolite extraction buffer was aliquoted for each sample. Media samples were taken and frozen at -80°C before intracellular metabolites were collected. To extract intracellular metabolites the strip was unrolled (**Figure 9**, step 19), carefully washed with 1x PBS by dipping the strip into a petri dish and placed onto a metal plate on dry ice. The strip was cut into pieces, which were transferred into the corresponding Eppendorf tube containing the metabolite extraction buffer and placed into the cooling bath for 15 min. The strip is removed before the final centrifugation step of the protocol in "2.8.2. LC-MS sample extraction". After centrifugation, the supernatant was transferred into a fresh Eppendorf tube and the pellet was dried using a speedvac. Afterwards, the pellet was resuspended in 1 ml/1*10^6 cells metabolite extraction buffer, sonicated until completely resolved, transferred into autosampler vials, and stored at -80°C until further analysis.

2.13.7. Metabolite distribution: Linear modelling

Metabolomics analysis was performed according to the workflow presented in 2.8.4. "LC-MS data analysis". For the linear modelling, the R base package stats (v3.5.1) (https://www.r-

project.org/) was used. In detail, the normalised total pool metabolomics results were used and for each biological replicate the pearson R^2 with corresponding p-value was calculated using the function "cor.test" (method = "pearson", conf.level = 0.95) . The pearson linear regression results of the biological replicates of clones with same FH status ($FH^{+/+}$ or $FH^{-/-}$) were pooled by calculating the median of the pearson R^2 using the function "median" and the adjusted p-value of the individual p-values by using the "sumlog" function, which aplies the Fisher's method (sum of logs method) as part of the metap package[154] (v1.4) . Metabolites that had a median pearson $R^2 > 0.7$ were categorised "UP", whilst metabolites with a median R2 < -0.7 were categorised "DOWN", otherwise they fell into the category "No Linear Change" (**Figure 11**). Additionally, metabolites were excluded from the category "UP" or "DOWN" with an adjusted p-value > 0.05 or with a R^2 variance > 0.2 (**Figure 11**). The variance was calculated using the R base function "var".

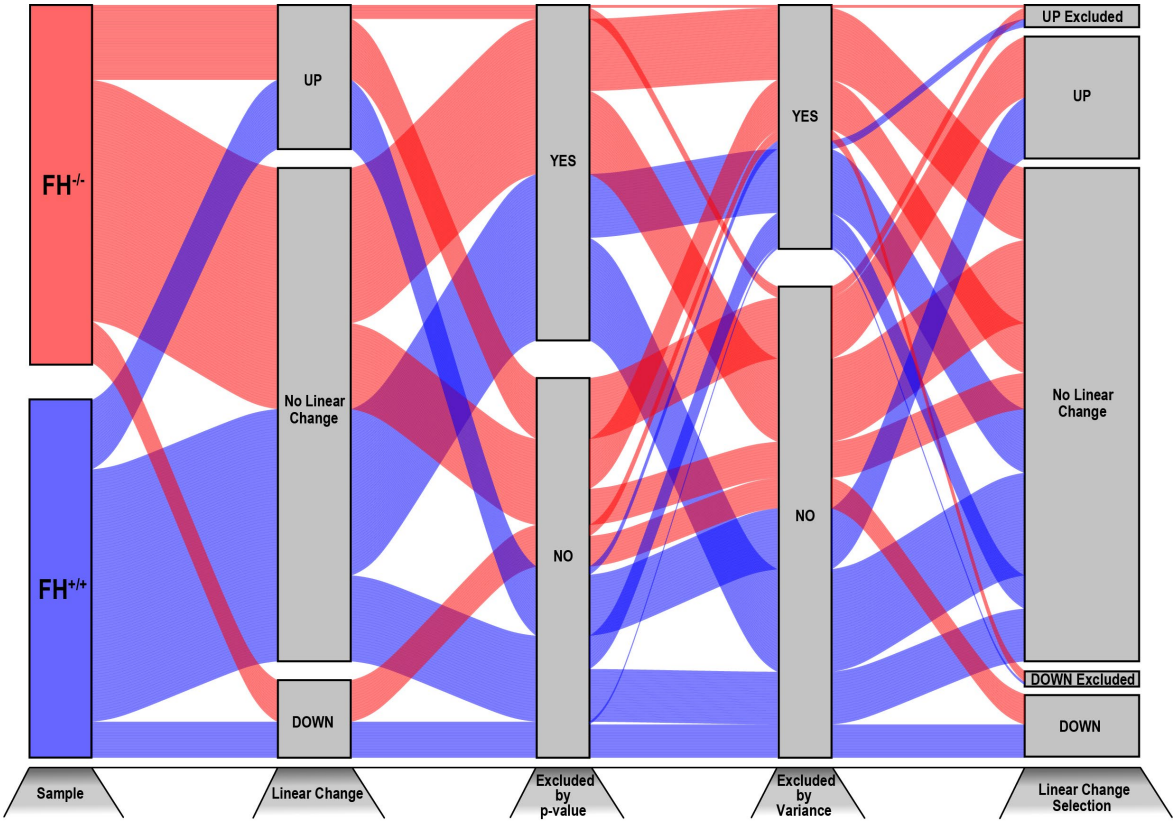


Figure 11: Linear modelling workflow using the LC-MS data.

Results Chapter 1

3.1. HK2 cells as a model of FH-deficiency

The current human models of FH loss, as the HLRCC-derived UOK cells[155,156], originate from primary or metastatic tumours, which represent an advanced phase of tumour formation. To better understand the early steps of transformation, I generated a human model of FH loss in non-transformed renal epithelial cells, HK2 cells, an immortalized proximal tubule cell line obtained from healthy adult human kidney[157]. To this aim, I performed CRISPR/Cas9-based genome editing of FH and characterized the single-cell clones. In detail, I generated a control cell line using an empty vector that expresses SpCas9, hereafter indicated as $FH^{+/+}$, and an FH-deficient cell line, hereafter indicated as $FH^{-/-}$, where the FH protein was depleted by targeting *FH* exon 2 and 4 with two independent sgRNAs (**Figure 12A-B**, **Supplementary Figure 1A-B**).

3.1.1. Metabolic rewiring in FH-deficient cells

To systematically elucidate the metabolic changes upon FH loss, I performed LC-MS-based metabolomics of intracellular and extracellular metabolites (**Figure 12C-I**, **Supplementary Figure 1C-D**). As observed in Fh1-deficient mouse cells[69], fumarate reaches millimolar levels in the $FH^{-/-}$ clones and it is secreted in the extracellular milieu (**Figure 12C-D**). This response is accompanied by a change of TCA cycle metabolites including the decrease of malate, the product of the FH catalysed reaction, and α-ketoglutarate (aKG), the increase of intra- and extracellular succinate, and increase of argininosuccinate (**Figure 12E**, **Supplementary Figure 1C**), as previously described in other models of FH deficiency[49,51]. We and others have shown that FH-deficient cells increase their glycolytic rate, and instead of fully oxidising glucose in the mitochondria, they shunt it into lactate production[49]. Consequently, glucose is replaced by glutamine as the main source of carbons for the TCA cycle[49]. Confirming these observations, I detected a minor increase in lactate release in the FH-deficient clones compared to the FH-proficient counterpart and a change in the TCA cycle metabolites (**Figure 12E-F**). To examine the metabolic rewiring in more detail, I capitalised on ^{13}Carbon (^{13}C) tracing

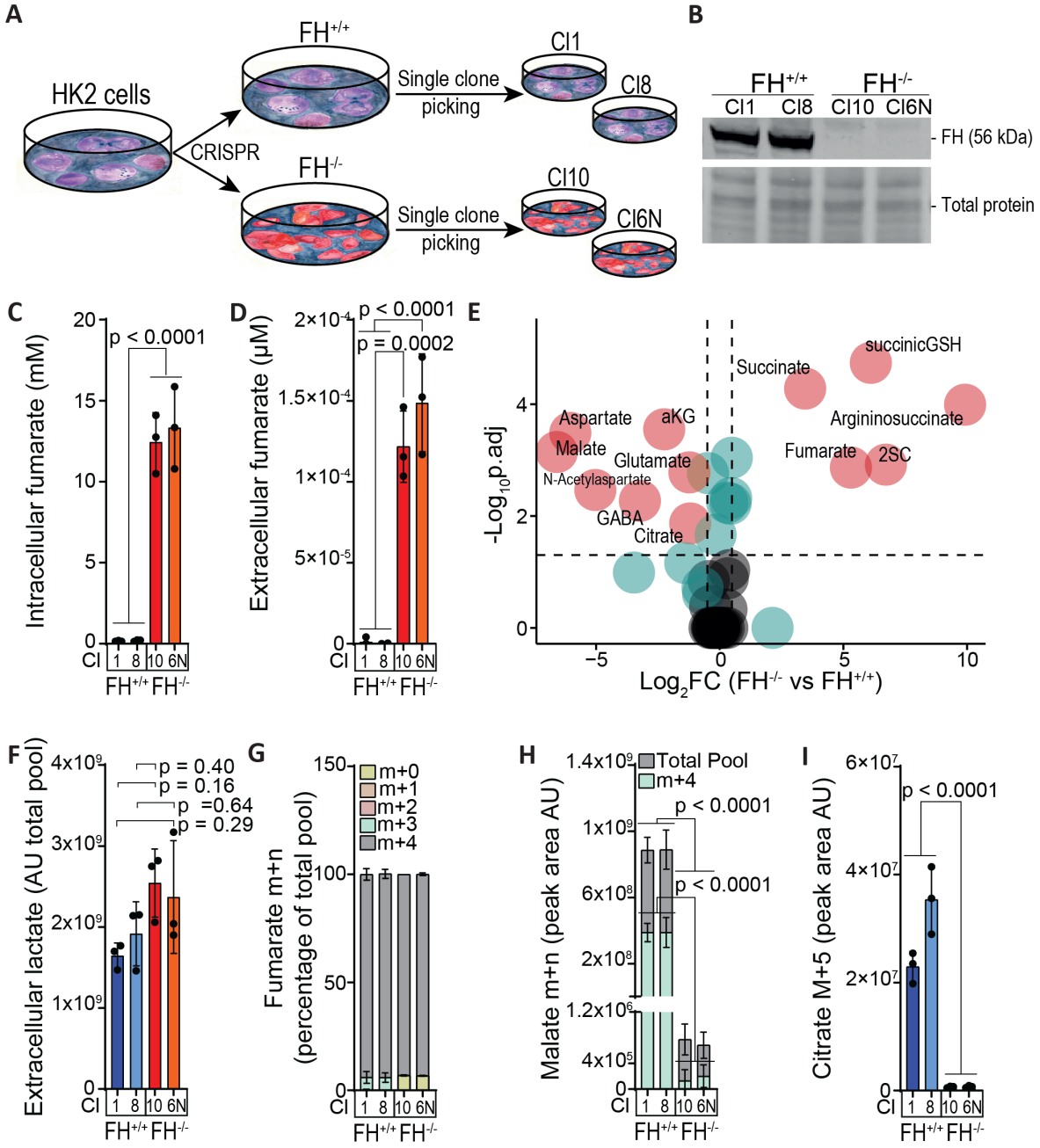


Figure 12: Metabolic adaptation phase after FH loss.

(A) Schematics of the generation of Fumarate Hydratase (FH) knockout cells using CRISPR/Cas9-based genome editing of a human non-transformed epithelial cell line, HK2. **(B)** FH protein expression measured using western blotting (Representative image of n=3, details see **Supplementary Figure 1A-B**). **(C+D)** Intracellular and extracellular fumarate levels, **(E)** differential intracellular metabolite levels and **(F)** extracellular lactate measured by LC-MS. **(G)** Proportion of total pool of m+n fumarate, **(H)** malate and **(I)** citrate m+5 were measured by LC-MS using U-^{13}C-glutamine in the culture media. Results were obtained from three biological replicates and presented as mean ± s.d. after performing total ion count normalisation and missing value imputation (for details see 2.8.4 LC-MS data analysis). P-value (p) is calculated using one-way ANOVA and the Tukey's multiple comparison test.

experiments[158], using uniformly ^{13}C-labelled glutamine (U-^{13}C-glutamine) (**Supplementary Figure 1D**). This metabolite enters the TCA cycle *via* glutamate, generating aKG enriched of 5 ^{13}C, shifting its molecular weight of 5 Daltons (the aKG thus generated is indicated as mass+5 dalton (m+5) isotopologue). aKG can then be oxidized within the TCA cycle generating m+4 succinate, fumarate, and malate, or can undergo reductive carboxylation, generating m+5 citrate. First, I observed that most of the fumarate is fully labelled (m+4) under these conditions, indicating that it derives from glutamine **(Figure 12G-H**). Patient-derived FH-deficient cells, UOK262 cells, use glutamine to generate lipogenic acetyl-CoA *via* the reversal of isocitrate dehydrogenase (IDH) and ACO2 in a process called reductive carboxylation [52]. In contrast, this reaction was not observed in mouse Fh1-deficient epithelial cells[49] or fibroblasts[53]. Similar to the observations made in mouse Fh1-deficient cells[49], I did not detect reductive carboxylation in our FH-deficient clones (**Figure 12I**). It has previously been proposed that the lack of reductive carboxylation could be due to the inactivation of ACO2 via succination of ACO2 itself or due to the lack of Fe-S clusters that are integrated into ACO2[47,53]. The higher levels of fumarate in our FH$^{-/-}$ cells compared to the UOK262 could lead to the inhibition of ACO2 and might explain the suppression of reductive carboxylation.

To sum up, these results show that FH loss in HK2 cells leads to profound metabolic changes and to fumarate accumulation, similar to the observations made in mouse and other human FH-deficient cells.

3.1.2. Biochemical changes and fumarate buffering

The aberrant accumulation of fumarate has been implicated in the dysregulation of several cellular processes. We recently demonstrated that fumarate affects mitochondrial respiration in a multi-pronged fashion[55]. For instance, fumarate blocks the activity of complex II *via* product inhibition, causing an overall suppression of oxidative phosphorylation (OXPHOS) in FH-deficient cells[55]. In accordance with these previous findings, the FH-deficient clones exhibit a lower oxygen consumption rate (OCR) compared to the FH$^{+/+}$ clones (**Figure 13A**). Moreover, I detected a decrease in the activity of respiratory chain complexes performing spectrophotometry (**Figure 13B**) and downregulation of a variety of OXPHOS signatures performing GSEA using RNA sequencing data comparing FH$^{-/-}$ *versus* FH$^{+/+}$ (**Supplementary Figure 2**).

Another mechanism through which fumarate can affect mitochondrial bioenergetics is succination[159]. This reaction affects proteins involved in the synthesis of Fe-S clusters, proteins required for the proper activity of many enzymes, including complex I and II of the respiratory

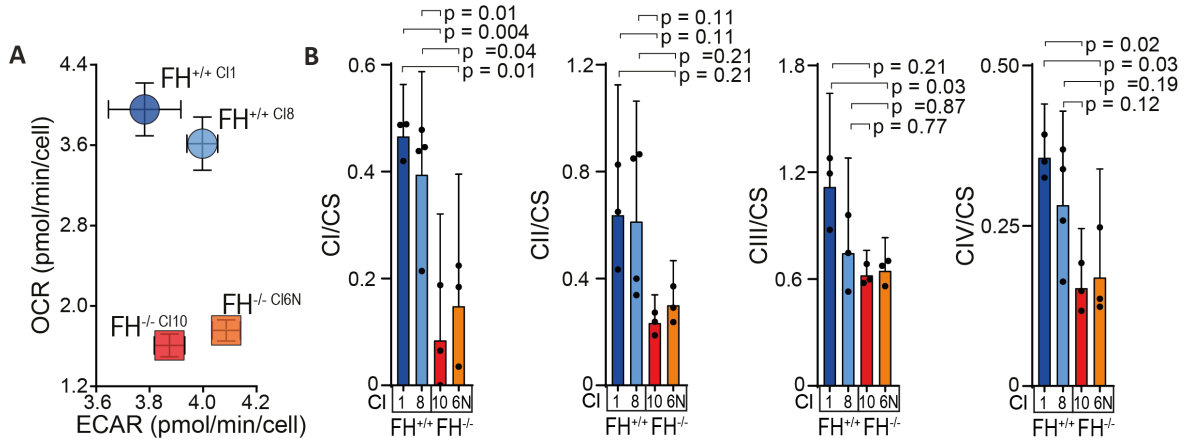


Figure 13: Impairment of OXPHOS in FH-deficient clones.

(A) Basal extracellular acidification rate (ECAR) and oxygen consumption rate (OCR) of three biological replicates measured using Seahorse. **(B)** Mitochondrial respiratory chain activity of isolated mitochondria normalised to protein amount and citrate synthase activity (CS) using spectrophotometry. Results were obtained from three to four biological replicates and presented as mean ± s.d., adjusted p-value (p) is calculated using one-way ANOVA and the Tukey's multiple comparison test.

chain[55]. In our FH deficient model, I detected an increase in 2SC and in the succination of protein residues determined by western blotting (**Figure 14A-B**, **Supplementary Figure 3A**). Performing succinated proteomics further confirmed the increase in succinated protein residues (**Figure 14C**, **Supplementary Figure 3B**), yet I did not detect the succination of Fe-S cluster assembly family of proteins. Nevertheless, these results confirmed the features of mitochondrial dysfunction and increased succination previously described in FH-deficient cells.

3.1.3. FH loss promotes oncogenic signalling cascades

Besides the above-described metabolic reprogramming and fumarate-induced mitochondrial dysfunction, FH loss and fumarate accumulation activate a plethora of pro-oncogenic signalling cascades, as described in the introduction. To investigate the downstream effects of protein succination in our clones, I focused on the well-studied NRF2 stabilisation and subsequent activation of an antioxidant response signature[66,78]. To this end, I applied an enrichment analysis on the proteomics and RNA landscape using the NFE2L2.V2 signature[160]. This analysis showed a significant upregulation of the NRF2 pathway in the $FH^{-/-}$ clones on RNA and protein expression level (**Figure 14D-E**). Consistent with this finding, we confirmed the upregulation of one of the NRF2 target genes, NQO1, in the $FH^{-/-}$ using qPCR (**Figure 14F**).

Fumarate accumulation is also known to inhibit the family of aKGDDs, enzymes involved in various cellular processes, including protein hydroxylation, RNA modifications, and DNA and histone demethylation[47]. In agreement with these findings, I detected overall

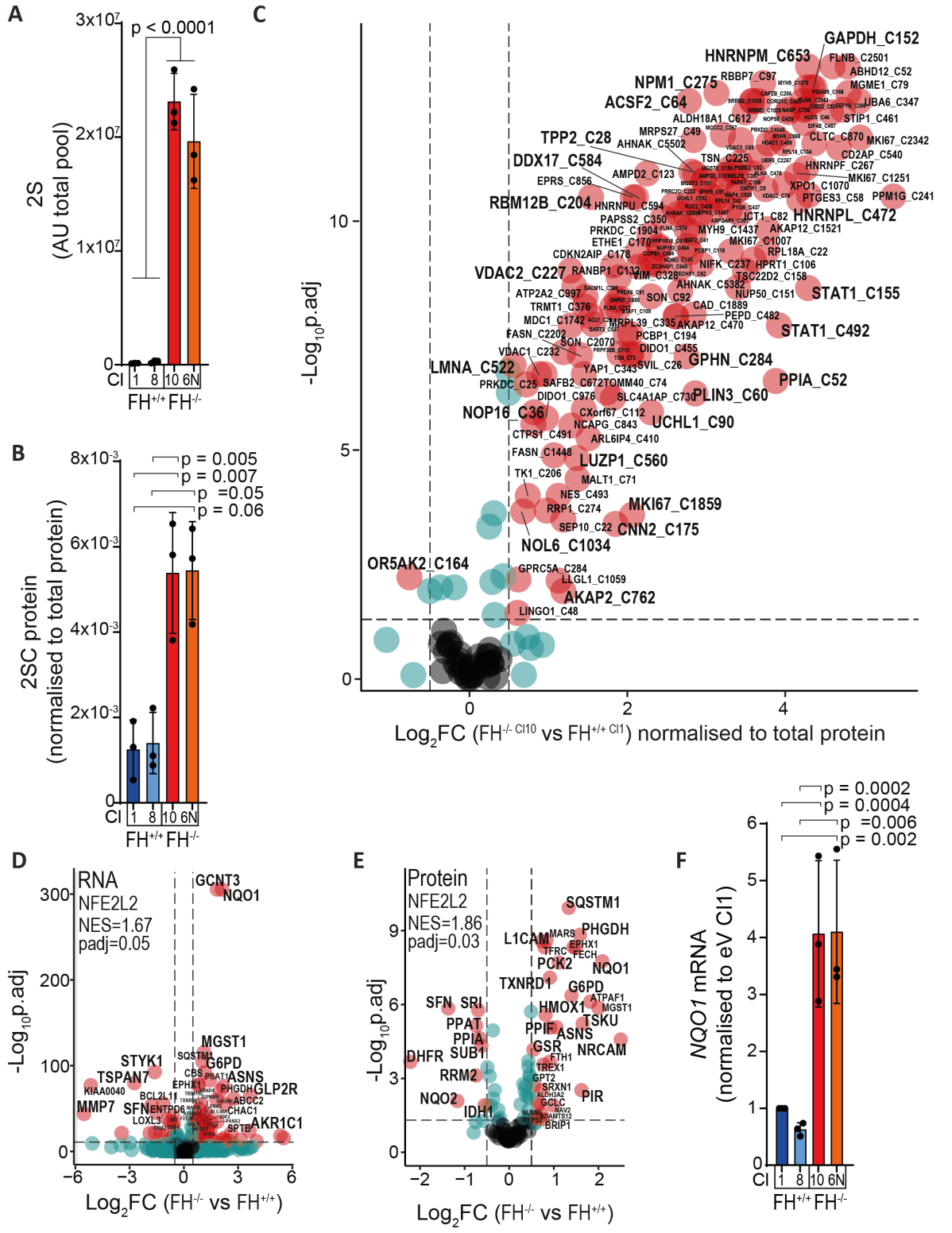


Figure 14: Fumarate accumulation results in succination and promotes antioxidant response.

(A) Intracellular 2SC concentration measured using LC-MS. **(B)** Succinated protein residues measured by western blotting (for image see **Supplementary Figure 3**). **(C)** Volcano plot of the detected succinated protein residues comparing $FH^{-/-}$ to $FH^{+/+}$. **(D-E)** Volcano plot illustrating the expression profile of the antioxidant response (GSEA including NFE2L2.V2 signature) on (E) RNA and (E) protein level detected in $FH^{-/-}$ compared to $FH^{+/+}$. **(F)** NQO1 mRNA levels measured by qPCR. TBP was used as a calibrator for the qPCR analysis. Results for A, B, F were obtained from three biological replicates and presented as mean ± s.d. Adjusted p-

value (p) is calculated using one-way ANOVA and the Tukey's multiple comparison test. Results for C, D and E were obtained from five biological replicates. Adjusted p-value (p.adj) is calculated using the DESeq2 package (RNAseq data) or the Limma-pipeline (proteomics data).

DNA methylation changes with 63% differentially hypermethylated and 24% differentially hypomethylated regions (**Figure 15A**). Hypermethylation suppresses the family of antimetastatic *MIRNA200*, leading to EMT[69]. Consistently, these miRNAs were suppressed in the FH-deficient cells (**Figure 15B**). This downregulation is in accordance with the enrichment analysis of the proteomics and RNAseq data using the manually curated "EMT DOWN" and "EMT UP" signatures[125], which showed a significant downregulation of a plethora of epithelial markers both at the protein and RNA level, when comparing $FH^{-/-}$ *versus* $FH^{+/+}$ (**Figure 15C**, **Supplementary Figure 2**). The decrease of epithelial markers upon FH loss was confirmed by qPCR, which showed a consistent decrease of mRNA levels of epithelial markers such as *E-CADHERIN* and *EPCAM* (**Figure 15D**).

Overall, these results show that our novel model of FH loss faithfully recapitulates the molecular features previously observed in mouse and human FH-deficient cells and tumours.

3.1.4. Discussion

In this Chapter, I have described the generation and characterisation of a novel non-transformed human model of FH loss. This model fully recapitulates the biochemical and phenotypic features of the previously described FH-deficient human and mouse models. Indeed, I detected the accumulation of fumarate and the metabolic hallmarks typical for FH loss. Moreover, I detected protein succination, NRF2 activation, DNA hypermethylation, and activation of EMT. However, there are also some notable exceptions.

For instance, I did not detect the succination of *KEAP1*, *ACO2,* or the Fe-S cluster assembly family members, whilst still observing the downstream activation of the antioxidant pathway (ascribed to KEAP1 succination[66,78]), the lack of reductive carboxylation (previously explained by ACO2 succination[53]), and the decrease in OXPHOS (previously linked to Fe-S cluster assembly family succination[55]). This observation can have multiple explanations: First, for the detection of succinated residues, there is no enrichment method, whereas KEAP1 succination was only detected after enrichment for this protein *via* pulldown[66,78]. Therefore, other succinated residues could have been missed due to their low abundance. Second, not all protein residues in mouse cells are conserved in human cells, yet excluding the cysteine residues that are not conserved in human only showed a minor overlap of the succinated proteins[129] (**Supplementary Figure 3C**). This further supports the problem of detection of succination: Given that FH-deficient cells accumulate high levels of fumarate driving succination,

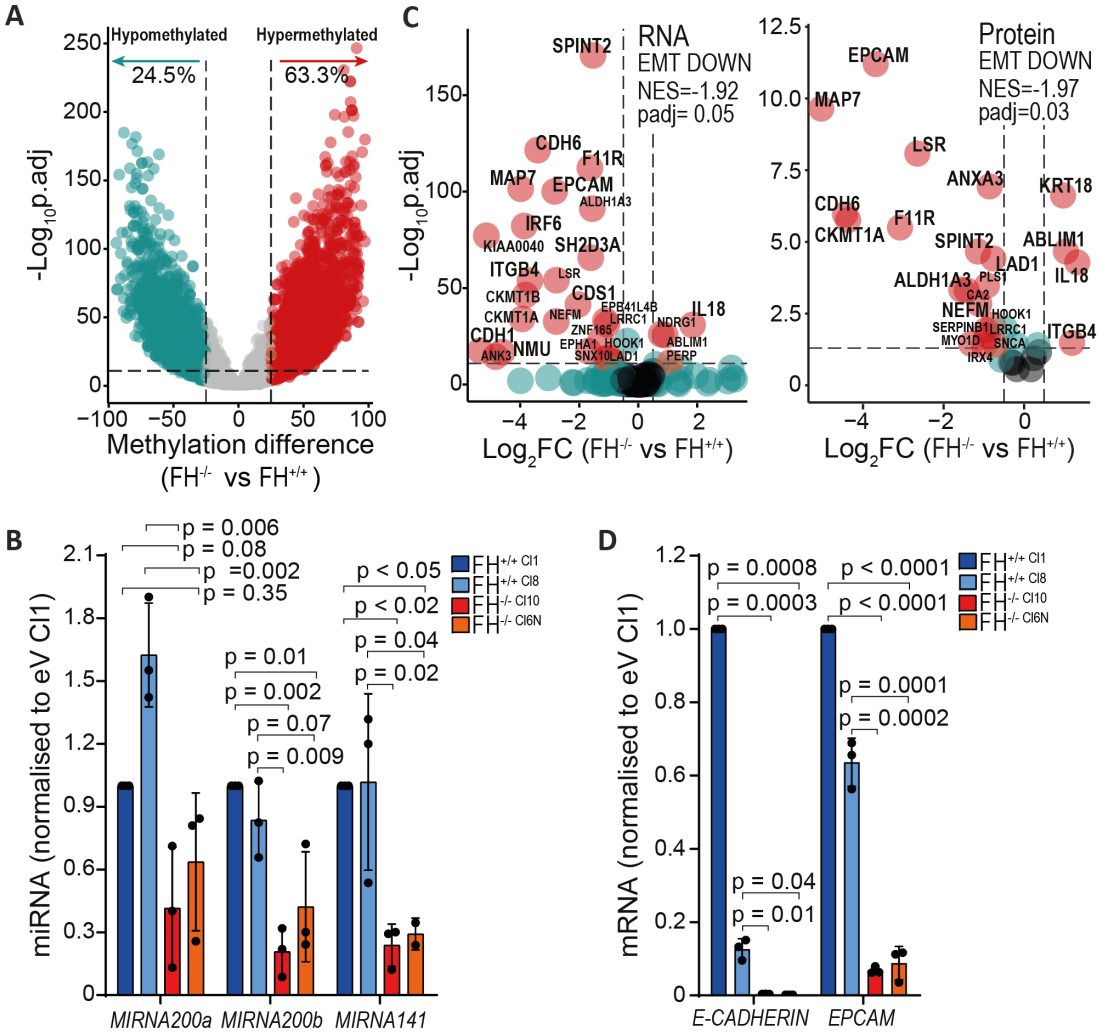


Figure 15: FH loss induces EMT through inhibition of MIR200 family members.

(A) Volcano plot of the DNA methylation difference between $FH^{+/+}$ and $FH^{-/-}$. **(B)** miRNA expression measured by qPCR normalised to $FH^{+/+}$ Cl1. SNORD 59 and SNORD 61 were used as calibrators for the qPCR analysis. **(C)** Volcano plots illustrating the expression profile of EMT DOWN (GSEA including EMT DOWN signature) on RNA and protein level detected in $FH^{-/-}$ compared to $FH^{+/+}$. **(D)** mRNA expression measured by qPCR normalised to FH+/+ Cl1. TBP was used as a housekeeping control. Results for A, B and D were obtained from 3 biological replicates. Results for B and D are presented as mean ± s.d. Adjusted p-value (p) is calculated using one-way ANOVA and the Tukey's multiple comparison test. Results for C were obtained from five biological replicates. Adjusted p-value (p.adj) is calculated using the DESeq2 package (RNAseq data) or the Limma-pipeline (proteomics data).

succination would be expected to occur on most amenable cysteine residues and one would expect a substantial overlap of detected succinated sites. Third, given that the cellular pH affects the succination reaction[129], the succination landscape can also be influenced by the cell-density, media acidification and general environmental cues.

Another difference between the mouse and human model is the lack of hypermethylation of *MIR200* family members [69] and *CDKN2A* hypermethylation[70,71]. Interestingly, whilst in the FH-

deficient cells, this hypermethylation was not detected (data not shown), the downstream change in epithelial marker expression and EMT, both ascribed to MIR200 hypermethylation, were a key feature of these cells.

The observation that despite these differences, FH-deficient cells maintain a similar phenotype is worth noting. Some of the signalling cascades discussed here are often sustained by converging mechanisms, and alternative routes may be explored in the different models to elicit those cascades. For instance, the decrease in OXPHOS in FH-deficient cells is not only caused by the suppressed synthesis of Fe-S clusters but also *via* the direct inhibition of SDH-dependent mitochondrial respiration by fumarate[55]. Similarly, the EMT is achieved by multiple transcriptional mechanisms, and the hypermethylation of *MIR200* family members could be bypassed by redundant pathways.

I also noticed important differences when comparing the two FH-proficient clones. In comparison to $FH^{+/+}$ clone 1, $FH^{+/+}$ clone 8 has decreased epithelial marker expression, which underlines an intrinsic heterogeneity of the wild-type population and the possibility that the impact of FH loss and downstream transformation may vary depending on the phenotype, rather than genotype, of the cell of origin.

Although these cell lines are a useful tool to investigate the consequences of FH loss, they do not appear to be fully transformed. Indeed, the FH-deficient cells grow nearly two-fold slower than their FH-proficient counterpart. This paradoxical finding can be explained by the fact that these cells may need additional genetic and/or tissue-specific environmental cues to become fully transformed. Therefore, studying FH-deficient cells in a more physiological context by using a 3D model or performing orthotopic kidney injection could help to understand if they can form tumours. Finally, these observations highlight the complexity of biological systems and the difficulty to study them using *in vitro* methods and 2D cultures. The cellular origin, age and state (cell-cycle phase and epigenetic organisation), and the environmental cues will all play a role in how a single cell will cope with a stressor such as the loss of FH. These different factors might also dictate which signalling cascades have the greatest contribution to the emerging phenotype. To understand the contribution of the various molecular events (as described above) to the FH signature, I developed a computational approach, which will be discussed in the next chapter.

Results Chapter 2

3.2. SiRCle (Signature Regulatory Clustering) model

Given the complexity of gene expression and the plethora of regulators that can impact this process, it is not surprising that the novel model of FH deficiency does not preserve all the previously observed cascades, despite presenting a similar phenotype. For instance, despite observing a change in EMT makers, I could not attribute this expression change to the hypermethylation of *MIR200* family members as previously described[69]. Consequently, the change in expression of EMT markers could be regulated at a different level in the HK2 model. To dissect at which level (methylation, transcription, translation) the gene expression is regulated, genes were clustered into regulatory groups based on the level that dictates the cellular phenotype (**Figure 16**). For example, genes corresponding to the dark red flow are hypermethylated (methylation status), have downregulated mRNA expression and downregulated protein expression, and hence we can assume that increased DNA methylation drives (regulation driver) the downstream phenotype. In this case, genes that follow this flow are assigned to the cluster "Methylation-driven suppression", short "MDS" (**Figure 16**).

3.2.1. Regulatory clusters that drive specific cellular phenotype

By sorting the genes into regulatory clusters based on their expression change at the level of DNA methylation, mRNA, ncRNA and protein comparing $FH^{-/-}$ versus $FH^{+/+}$, I observed that only some flows are direct (e.g. hypermethylation, lower mRNA expression, lower protein expression), whilst there are also flows that are regulated at multiple levels (e.g. hypermethylation, high mRNA expression, low protein expression) (**Figure 16**). Summarising the different flows into groups resulted in 12 regulatory clusters, where the methylation-driven clusters included the lowest number of genes. To assess if these clusters of genes can be associated with the phenotypic changes observed in the FH-deficient cells, I performed over representation analysis (ORA) on each cluster. Importantly, 10 out of the 12 clusters were

associated with distinct biological signatures (**Figure 17**). The biological signatures associated to each of the 10 clusters are tightly interconnected and have at least one gene in common (depicted by the connecting lines), which further shows that clustering the genes based on their regulatory level leads to clusters that could drive a distinct part of the cellular phenotype. For instance, the "Transcription and Processing Driven Suppression" (TPDS) cluster, which is regulated at the transcriptional level (decreased mRNA expression), includes genes that play a key role for cytoskeleton organisation and their altered expression profile after FH loss is

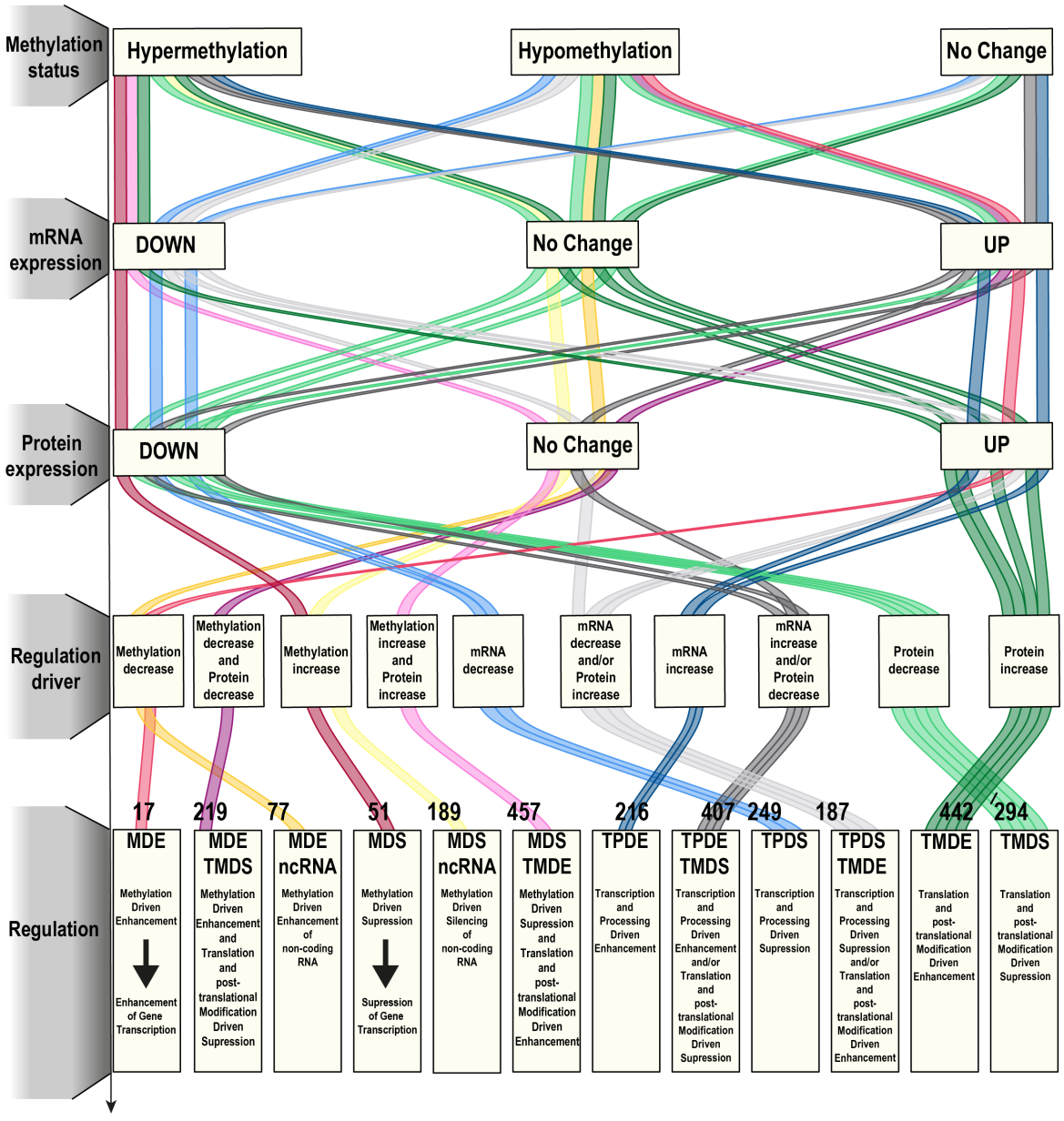


Figure 16: Alluvian plot of the SiRCle model.

Genes are clustered into the regulatory clusters („Regulation") following the flows (lines) through the alluvian plot, which results in clusters dependent on which level (methylation status, mRNA landscape, protein landscape) gene expression is influenced and hence the regulation is driven („Regulation driver"). The total number of genes placed in each cluster is labelled on the flows that result in the cluster.

predicted to be caused by suppressing their transcription (**Figure 17**). This cluster includes genes such as EPCAM, which has previously been associated with the EMT signature

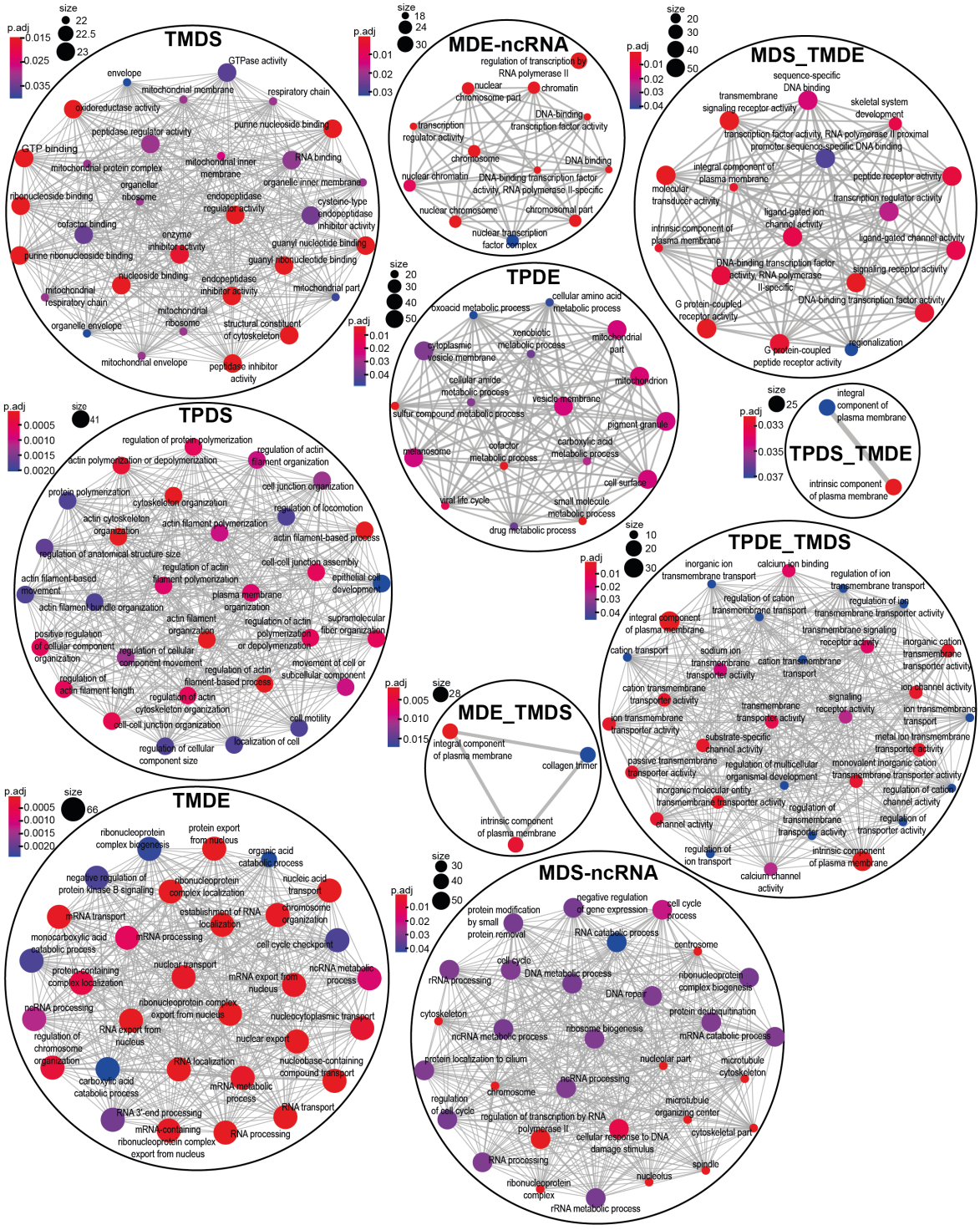


Figure 17: Emapplots of the ORA results shows phenotypic changes of FH loss.

ORA performed on each regulatory cluster resulted in biological pathways that are altered comparing $FH^{-/-}$ *versus* $FH^{+/+}$. The dot size corresponds to the number of genes found in the cluster that are part of the biological pathway. The colour of the dot shows the p-adjusted value (p.adj) of the ORA. The connecting lines (grey) show that the biological pathways have genes in common.

(**Figure 15**, **Supplementary Figure 4A**). Another interesting example is the opposingly regulated cluster "Transcription and Processing Driven Enhancement" (TPDE), which includes genes associated with mitochondria and metabolism that are predicted to be altered in FH loss by enhancing their transcription (**Figure 17**). Reassuringly, this includes genes such as HMOX1 an enzyme of the Haem biosynthesis and degradation pathway and NRF2 target gene, which had been previously shown to be essential for survival in FH-deficient cells[49] (**Supplementary Figure 4B**).

Next, I wanted to elucidate which genes *per* cluster are most differentially regulated upon FH loss. To do so, a VAE was used to compress the data into three nodes to learn an integrated rank for each gene within a cluster using all genes that were assigned to a cluster as the training data. Additionally, the VAE enabled the different features such as the differential expression in the different detection methods (Bisulphite sequencing, RNAseq, Proteomics) and the gene expression differences between the individual clones to be taken into account. This analysis resulted in three ranks for each gene corresponding to the three nodes in the latent space of the VAE. In comparison to the feature correlation of the input data selection, which shows that the different data types cluster best with each other even for the different clones than showing good inter-data correlation (**Supplementary Figure 5A**), the latent space of the VAE captured the data separation based on the underlying omics datasets (**Supplementary Figure 5B-C**). Reassuringly, this approach captured and separated the different clusters resulting from our SiRCle method (**Figure 18A**). Using the VAE gene ranking, the key genes of a cluster can be identified. For instance, looking at the top ten and bottom ten genes of each node in the TPDE cluster shows a direct correlation between mRNA expression and downstream protein expression, whilst increased hypermethylation does not affect the increased mRNA expression (**Figure 18B**). Next, I used the ranked list of genes to perform GSEA for the TPDE cluster, which includes genes predicted to be expressed due to enhanced transcription and showed that these genes play a role in the "Hallmarks_mTORC1_ Signalling" pathway when ranking the genes of the TPDE cluster *per* node 0 (**Figure 18C**). The mammalian/mechanistic target of rapamycin complex 1 (mTORC1) and the associated signalling pathways regulate numerous cellular processes implicated in proliferation, mitochondrial metabolism, and cell growth[161]. The mTORC1 associated pathways are in line with the observation that genes in the TPDE cluster affect mitochondria morphology and metabolism. Moreover, amongst the targets of the mTORC1 pathway detected in the TPDE cluster are metabolic enzymes such as PHGDH, ASNS (Glutamine-Dependent Asparagine Synthetase) and SHMT2 (mitochondrial Serine Hydroxymethyltransferase 2) (**Figure 18C**),

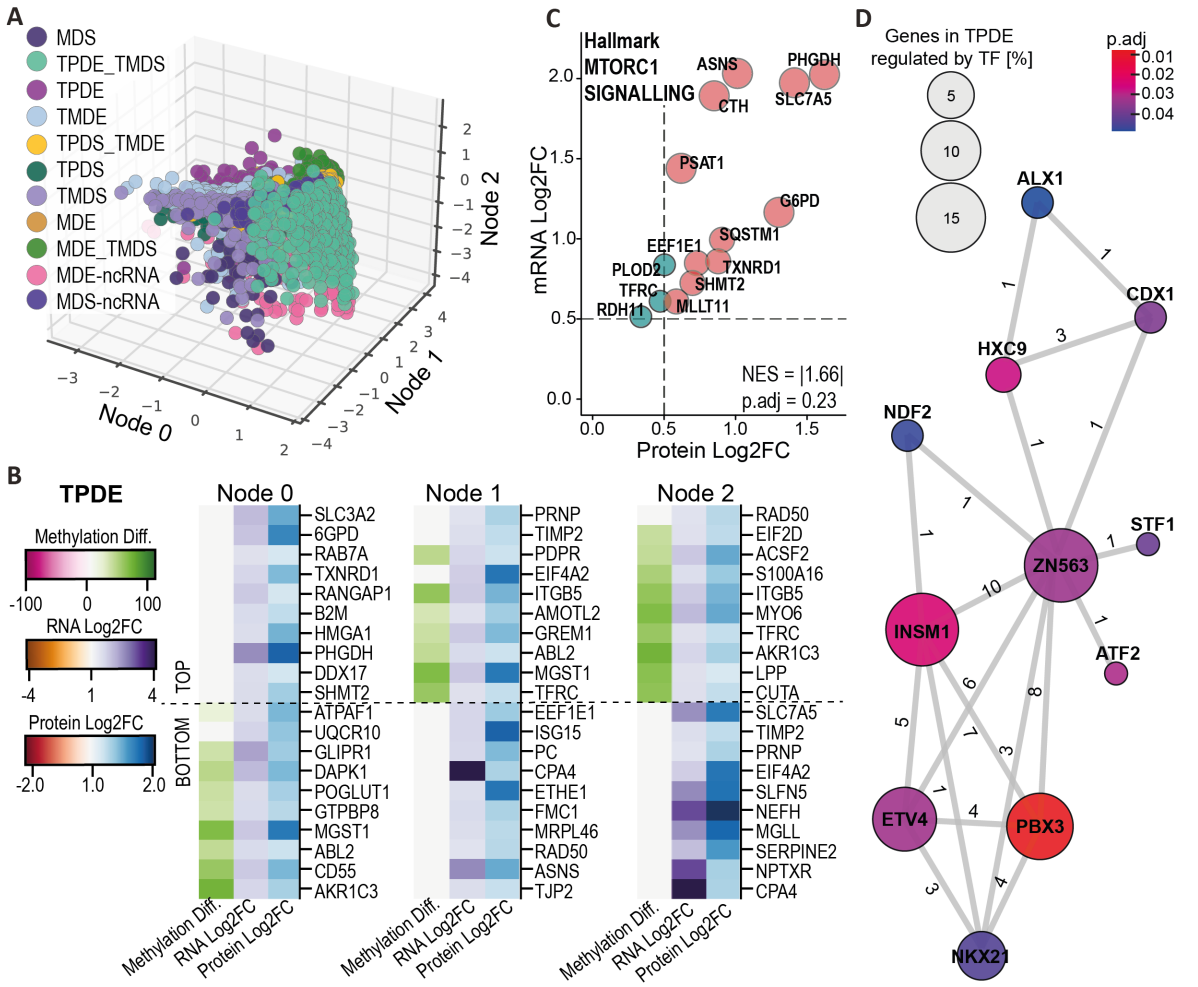


Figure 18: Gene ranking using the VAE model and TF analysis.

(A) Dotplot of the genes in the latent space of the VAE colour coded for the regulatory clusters. **(B)** Heatmap of nodes' top/bottom ten genes of the TPDE cluster. **(C)** Correlation plot comparing mRNA expression versus protein expression of the Log2FC comparing $FH^{-/-}$ *versus* $FH^{+/+}$. Genes included are part of the „Hallmark MTORC1 Signalling" pathway, which was enriched in GSEA ranking for node 0 of TPDE. **(D)** TF plot of unique TFs that are predicted by motif analysis to drive genes in the cluster TPDE. The dot size corresponds to the percentage of genes of the cluster that are targeted by the TF. The colour of the dot shows the p-adjusted value (p.adj). The connecting lines (grey) shows the number of common genes the connected TFs regulate.

which is consistent with the ORA results that assigned multiple metabolic pathways to be altered in the TPDE cluster.

Conventional methods that perform ORA or GSEA based on transcriptomics data only, fail to capture on which level the expression change is taking place and if this change is translatable into the phenotype present on the protein landscape. The SiRCle method overcomes this by using all available information to generate a model of the data, which can be applied to a multitude of biological datasets. Together, these results show that the concept of "Signature Regulatory Clustering" (SiRCle) can be used to disentangle at which level a cellular phenotype

is altered and at which level a phenotype must be targeted to block/revert a specific phenotypical trait.

3.2.2. Transcription factors that could drive the regulatory clusters

To investigate which transcription factors (TFs) drive the SiRCle clusters, I used motif analysis of the gene flank (100 bp upstream of TSS) and identified groups of enriched motifs within the clusters based on the Fisher's exact test. I found several TFs for each cluster that are uniquely driving a group of genes within the associated cluster. For clusters that are regulated at the level of transcription, TF analysis can predict which factors drive the expression change on the mRNA landscape. Moreover, methylation changes on the DNA can alter the binding ability of TFs and hence influence the expression landscape. We identified a set of TFs in the TPDE cluster that could drive the changes in gene expression involved in mitochondria morphology and metabolism observed in FH-deficiency (**Figure 18D**). Some genes are regulated by multiple TFs as depicted by the connecting lines (grey lines **Figure 18D**). Importantly, several TFs that drive the genes in the TPDE cluster, such as PBX3, ATF2 and ETV4 are detected at the RNA level in the samples (data not shown).

Given that TF activation is not just a matter of TF expression, but TFs are often regulated *via* phosphorylation or other mechanisms, it is important to validate this finding with wet-lab studies such as Chip-Seq. The advantage of performing the regulatory clustering prior to the TF analysis is that we can exclude TFs that are predicted to regulate genes in clusters regulated at the level of translation (e.g. TMDE, TMDS), since TFs regulate targets at the level of transcription. TFs are known to drive a plethora of genes and the regulatory clustering can dissect which of these TF target genes are actually regulated by the TF and hence which phenotypical change is driven by the TF. In turn, it also allows the user to understand which TF target genes are altered at other levels of regulation (e.g. translational regulation) independent of the TF itself.

3.2.3. Multi-omics data integration

Given that a protein's function can be influenced by alternative splicing or PTMs such as phosphorylation or succination, I wanted to understand whether genes' activity in a regulatory cluster are altered *via* either of those mechanisms. To quantify the likelihood of the overlaid data falling onto a specific regulatory cluster, I used the Fisher's exact test. Succination plays a crucial role in FH deficiency and has been shown to affect many proteins in the $FH^{-/-}$ clones (**Figure 14**). Interestingly, I found that proteins belonging to the TMDE cluster are significantly enriched to contain succinated proteins (**Figure 19A**). TMDE is associated with genes that play

a crucial role in RNA biology processes, including nuclear export/import processes and mRNA processing (**Figure 17**). Amongst the associated succinated proteins are proteins involved in nuclear export/import processes such as NPM1 (Nucleophosmin 1), NUP153 (Nucleoporin 153) and NUP50 (Nucleoporin 50). This result could hint at a feedback loop that alters the protein function, and hence affects the pathways associated with TMDE. This means, despite the upregulation of the proteins in a certain pathway predicted for the TMDE cluster using ORA or GSEA, the processes the pathway is usually driving might not be altered due to succination.

RNA processing pathways are differentially regulated comparing $FH^{-/-}$ *versus* $FH^{+/+}$ and the expression of genes involved in these pathways are regulated at the translational (TMDE) and methylation (MDS-ncRNA) level (**Figure 17**). To understand if the alternative spliced isoform expression is altered, differential transcript analysis using StringTie and IsoformSwitchAnalyzeR was performed. A plethora of genes were assigned to isoform switching in FH-deficient cells, indicating altered functionality (**Figure 19B**). Overlaying these genes onto the SiRCle clusters, showed that alternative splicing affects genes of the TPDE cluster involved in mitochondria and metabolism (**Figure 19A**). Amongst those alternative isoforms, I found genes known to be crucial in FH-deficient cells such as *NQ01*, but also genes involved in mitochondria morphology and metabolism, namely *CHCHD6* (Coiled-Coil-Helix-Coiled-Coil-Helix Domain Containing 6, Mitochondria), *ASNS* and *PHGDH*. The latter two are part of the MTORC1 signalling, where we observed enhanced gene expression for several members of the pathway (**Figure 18C**). I also found that upon FH loss, alternative isoforms of *EEF1E1* and *TXNRD2*, both members of the MTORC1 signalling pathway, are expressed. Intriguingly, MTOR is also alternatively spliced in FH-deficient cells in two protein-coding isoforms (decreased usage of ENST00000376838 with -17% and increased usage of ENST00000361445 with +15%), which further shows that this key bioenergetics pathway is affected *via* alternative splicing occurring at the regulatory level of transcription (TPDE).

Observing the landscape of genes affected by succination showed that the majority of succinated proteins are associated with transcription, translation, cytoskeleton and metabolism (**Figure 19C**). Interestingly, succination is detected on a plethora of Heterogeneous nuclear ribonucleoproteins (HNRNP) a family of RNA-binding proteins (RBPs) that are crucial for alternative splicing, mRNA stabilization, and transcriptional and translational regulation[162]. This result could explain, on the one hand, the increase in different isoform usage upon FH loss and, on the other hand, the change in RNA biology regulated at the translational level (TMDE).

To sum up, overlaying post-translational modification data, isoform expression data or other biologically relevant data over the SiRCle model can further explain how the genes of the

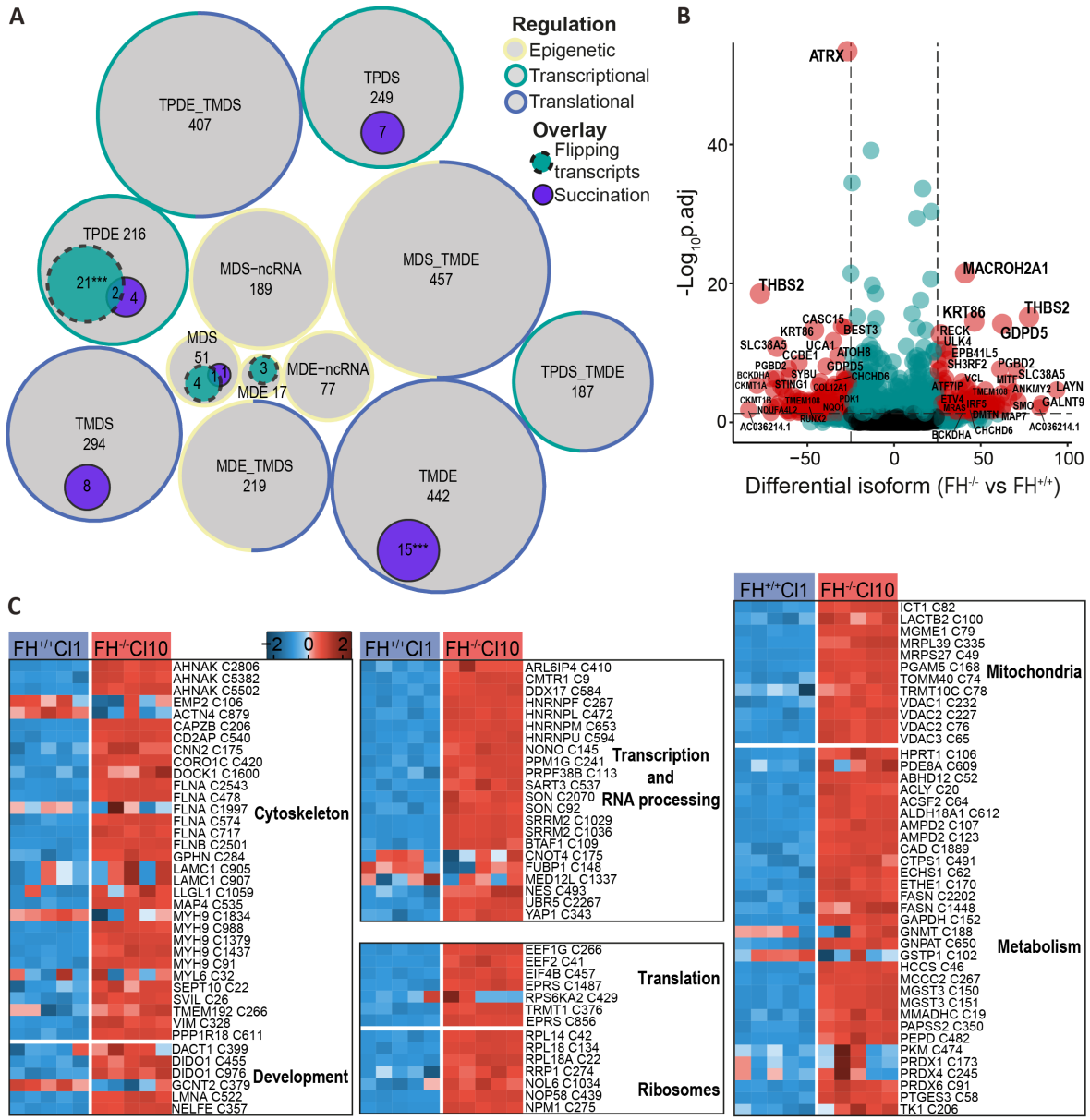


Figure 19: Overlay of succination and alternative isoform information.

(A) SiRCle plot showing the different clusters and the overlay. **(B)** Volcano plot of the differential isoform expression comparing FH$^{-/-}$ *versus* FH$^{+/+}$. **(C)** Heatmap of protein succination divided into pathways.

individual clusters are regulated.

3.2.4. Discussion

Inferring signalling pathway activity from gene expression data is a common approach to study cancer biology. Often, the most differentially activated pathways are followed up by studying the possible underpinning molecular mechanisms. Yet, these findings are often not confirmed in patient samples. This is also what I experienced in the novel model of FH-deficiency: For instance, I observed the loss of epithelial markers as a consequence of FH loss without being able to detect the upstream mechanistic alteration that had been ascribed to drive these changes, such as the epigenetic suppression of MIR200. Given the complexity through which a signalling

pathway can be regulated, it is not surprising that each patient, cellular model and even the severity of e.g. FH loss can alter *via* which mechanisms a signalling pathway is regulated. In turn this also implies that there is an interplay of multiple mechanistic cascades that result in the differential activation of a pathway and potential downstream phenotypical hallmarks of a disease.

Another problem researchers in the field recently started to discuss is the (mis)use of GSE methods, such as GSEA, to study cancer and infer the differentially regulated pathways only basing the analysis on transcriptomics data[163]. Ultimately, the protein expression profile determines the phenotype of a cell and hence inferring pathway activity solely based on transcriptomics data might be misleading. Indeed, it was hypothesised that pathway based methods may be effective because pathway gene sets overlap with genes regulated by TFs and not because the transcriptomics landscape is translated into the protein landscape[163]. In turn, this shows that GSE methods do not predict the pathway activity, but reflect the changes in TF activity that regulated the genes of a pathway[163]. The SiRCle method overcomes this limitation by using an integrated rank from methylation, transcriptome and proteome analyses for GSEA as opposed to using the mRNA response alone.

In this chapter I have introduced SiRCle, a novel approach to overcome these challenges, and I applied this model to dissect if the gene expression after FH loss is influenced at the level of methylation, transcription or translation and to understand if this is translated into phenotypical changes in FH-deficient cells. Importantly, I could show that SiRCle helps to disentangle at which level the gene expression is influenced, and more importantly, if it translates into the phenotype present at the proteome level. For instance, I found that in this FH-deficient model, the cytoskeletal markers, such as EPCAM, are predicted to be regulated *via* transcriptional suppression independent of hypermethylation. Moreover, combining SiRCle with TF analysis, PTM data or alternative splicing data gives the ability to further dissect the intricate network of gene regulation. Although overlaying the succination data onto the SiRCle clusters revealed several clusters affected by succination, this could not capture the overall cellular pathways succinated proteins are involved in, since succination proteomics is a small data set. The overlay of a small dataset onto the SiRCle clusters can help to understand if a cluster is significantly altered, yet one should also consider checking the general biological impact of e.g. this PTM since this gives an additional type of information. In other words, with SiRCle one can understand at which level the genes are regulated, but for small datasets and PTMs one needs to also consider that genes while regulated at different levels, thus falling into several SiRCle clusters, may share the same PTM such as succination. Succination has been

found to affect proteins of the TMDE cluster that are involved in nuclear export/import processes and some correspond to the nuclear pore complex. The nuclear pore complex consists of nucleoporins (NUPs), which can form disulphide bonds in oxidative environments[164] that eventually control the molecular crowding barrier of the nuclear pore in a redox-sensitive manner[165]. Mutation analysis of cysteine residues of NUPs revealed that mutations lead to mislocalisation for several NUPs, whilst mutations of NUP153 in the 21 cysteines of the zinc finger region (C585-C874) leads to the loss of adaptive crowding control[165]. In the $FH^{-/-}$ clones we observed succination of NUP153 (C404) and NUP50 (C151), which have not been tested by mutation studies to alter crowding or to mislocalise. It is tempting to speculate that succination could alter the molecular transport of RNA species through the nuclear pore complex.

Investigating the succination landscape has revealed that succination affects a plethora of proteins associated with transcriptional regulation and alternative splicing after FH loss. Indeed, I detected a multitude of alternatively spliced isoforms that are found in FH-deficient cells, which is in line with the recent finding of an alternative splicing signature in papillary renal cell carcinoma[166]. Moreover, the SiRCle model detected multiple RNA biology pathways to be altered in the clusters TMDE and MDS-ncRNA. In particular, the succination of HNRNP family members, key regulators of splicing[162] that have previously been implicated in different types of cancers[167], could drive the observed changes after FH loss. Previous analysis of the TCGA data showed that HNRNPs are differentially expressed in KIRP (kidney renal papillary carcinoma) in comparison to normal tissue, and that protein expression was increased for most of the HNRNPs in RACA (renal cancer) using "The Human Protein Atlas" database[167]. Moreover, HNRNPs were associated with cancer-related pathways, which showed that HNRNPF, HNRNPH2, HNRNPU and HNRNPUL1 are more likely to be implicated in those cancer related processes[167]. Interestingly, we detected succination of HNRNPF, HNRNPL, HNRNPM and HNRNPU, two of which have been tightly related to oncogenic changes in tumorigenesis. Finally, amongst these regulated cancer-related pathways are "MTORC1_Signalling", "HEME-Metabolism" and other metabolic processes[167]. This could explain the increase in alternative splicing of MTOR targets and potentially of MTOR itself. Noteworthy, mTOR signalling pathways have a close interplay with alternative splicing in cancer[168] and the isoform mTORβ (ENST00000376838.5/MTOR-202) that is an activated form of mTOR has been shown to promote cell proliferation and tumorigenesis[169]. In our FH-deficient cells, I detected a decrease of 17% in the mTORβ variant, whilst the full length mTOR

transcript (ENST00000361445.8/MTOR-201) was 15% increased. It would be interesting to investigate the role of this transcript switch of mTOR in our cellular model.

Together, these results show that SiRCle is a novel tool to dissect the interconnected landscape of gene expression regulation and combining it with TF analysis, PTM data and isoform expression can reveal novel regulatory cascades and connections. Additionally, this approach can disentangle at which level a cellular phenotype is altered and hence modifiable *via* genetic or pharmacological approaches, which can be applied to a multitude of biological datasets. Although the SiRCle analysis revealed an enormous complexity and interconnection between the levels of regulation, this does not take into account how external environmental cues can impact on this regulation. This question can be addressed by multi-omics analyses of cells exposed to varying levels of nutrient and oxygen using the 3D model that I will describe in the next chapter.

Results Chapter 3

3.3. Towards the generation of a 3D model of FH loss (TRACER)

The loss of FH causes the accumulation of fumarate, which elicits a variety of signalling cascades triggered at the epigenetic, transcriptional, and translational level. Thus far, most of these findings were made by culturing cells in commercial media under atmospheric oxygen conditions. Given that the metabolic rewiring and fumarate accumulation has been described to directly impact the downstream signalling cascades, I wanted to understand if environmental cues such as oxygen and nutrient concentrations influence the hallmarks of FH loss. To shed light on the role of the microenvironment in FH-deficiency, I recently capitalised on a novel 3D scaffold, a Tumour Roll for Analysis of Cellular Environment and Response (also known as TRACER), which consists of a permeable strip as a scaffold populated with cells and then wrapped around a mandrel (**Figure 20A**)[170]. This device allows the generation of oxygen and nutrient gradients that mimic those observed in a solid tumour and can be rapidly disassembled allowing the analysis of cells grown under different environmental conditions. In the next chapters I will describe how I took advantage of TRACER as a 3D model of FH loss.

3.3.1. Hallmarks of FH loss are robust to environmental changes

Using TRACER, I first assessed cell proliferation. Reassuringly, both $FH^{+/+}$ and $FH^{-/-}$ clones proliferate within the strip (**Figure 20B-C**, **Supplementary Figure 6A**). I then assessed the expected activation of the hypoxic response in the innermost layers of TRACER. Culturing the clones for 24 h leads to the expression of the HIF target gene BNIP3 and metabolite markers of hypoxia, such as 2HG and lactate, in a layer-dependent manner (**Figure 20D-F**). Together, this data confirms the validity of TRACER to mimic the effects of hypoxia.

Next, I assessed the behaviour of the metabolic hallmarks of FH loss, such as the accumulation of fumarate and the previously reported metabolic rewiring (**Figure 20G**, **Supplementary Figure 6B**). As in 2D cultures, I observed the accumulation of fumarate, succinate , and 2SC, whilst malate levels decreased. These changes were consistent across the

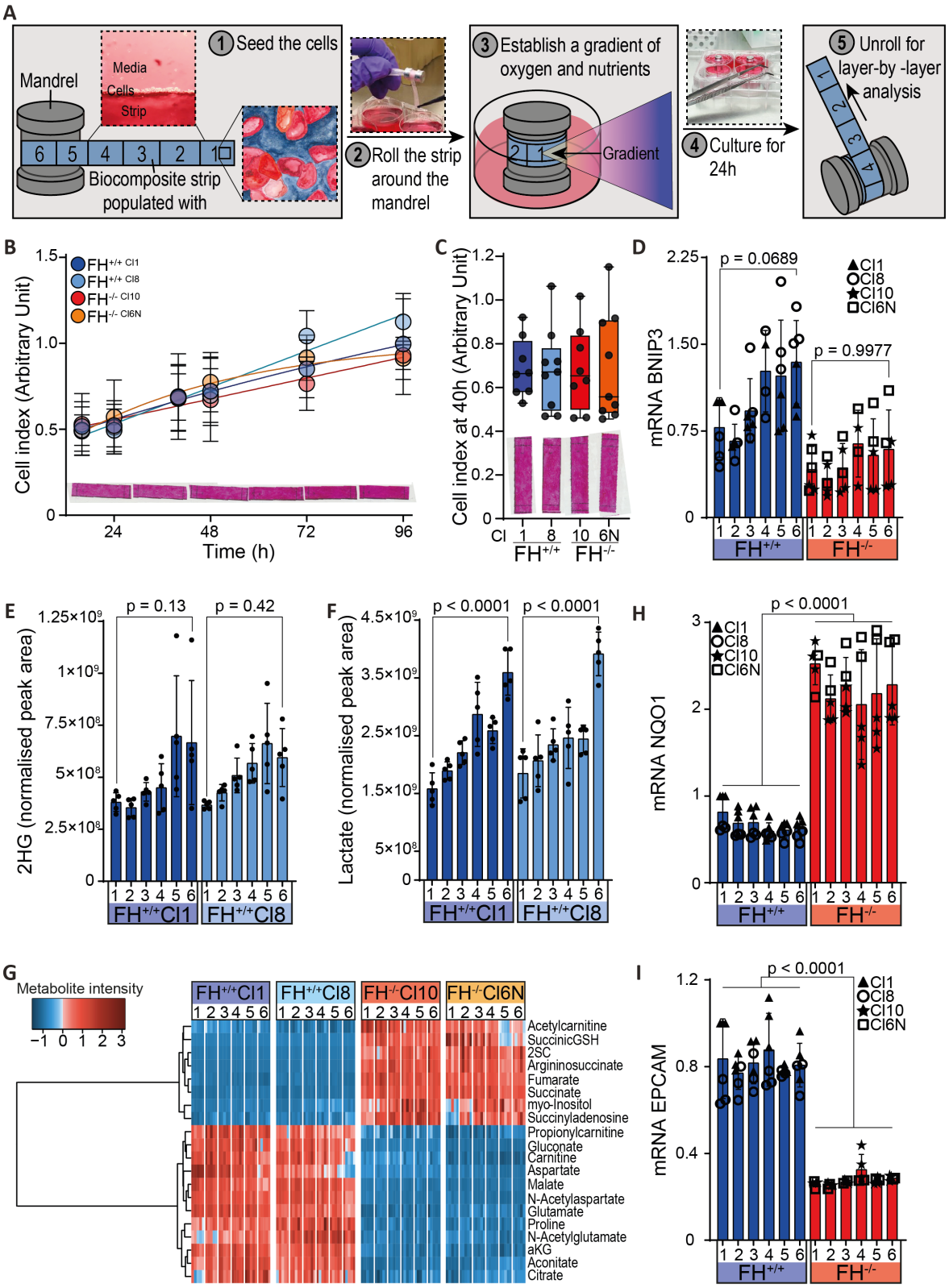


Figure 20: Introduction to the 3D Model TRACER and quality controls.

(**A**) Schematics of the oxygen-impermeable mandrel, the basis of the structure around which the biocomposite strip (violet) populated with cells is rolled. This strip is permeable to oxygen and nutrients, and allows cell migration through the different layers. After wrapping the strip around the mandrel, a gradient of nutrients and oxygen establishes. For the layer-by-layer analysis the strip can be quickly disassembled. (**B**) Growth curve of the clones cultured on the biocomposite strip and (**C**) quantification of cell amount at the time point of harvest

obtained *via* SRB staining. (**D**) BNIP3 mRNA expression measured by qPCR. (**E-F**) Intracellular metabolites (**E**) 2HG and (**F**) lactate and (**G**) the metabolic FH-deficient signature measured using LC-MS. (**H**) NQO1 and (**I**) EPCAM mRNA expression measured by qPCR. **B** and **C** are obtained from 8-10 biological replicates. **D**, **H** and **I** were obtained from three biological replicates. TBP was used as a housekeeping control. **E**, **F** and **G** are obtained from 5 biological replicates. All replicates are presented as mean ± s.d. P-value (p) is calculated using one-way ANOVA and the Tukey's multiple comparison test.

TRACER layers and independent of the gradient of nutrient and oxygen, showing that the metabolic defects caused by FH loss are not affected by oxygen or nutrient levels. To understand if this oxygen- and nutrient-independent fumarate accumulation also triggers downstream oncogenic signalling cascades, such as EMT or antioxidant response, I performed qPCR of the main targets as identified in our previous RNAseq and proteomics results (**Figure 20H-I**). In line with the hallmark FH-deficient metabolic signature over the TRACER layers, the expression of the antioxidant response target NQ01 and the epithelial marker EPCAM are not altered by oxygen or nutrient levels, which confirms that the molecular signature of FH loss is robust to environmental changes.

3.3.2. Linear model of the TRACER metabolite distribution reveals metabolic defects

Besides assessing the specific metabolic signature of FH loss, I wanted to determine whether oxygen and nutrient availability could affect other properties of FH-deficient cells. To this end, I performed semi-quantitative LC-MS analysis of 129 metabolites of five independent TRACER roles for each of the clones (**Supplementary Figure 6C**). Reassuringly, I detected a correlation between the two $FH^{+/+}$ and the two $FH^{-/-}$ clones, but at the same time I observed reduced correlation when comparing the different TRACER layers of the same clone (**Figure 21A**, **Supplementary Figure 6C**). This result indicates that the gradient of nutrient and oxygen influences the metabolic landscape of cells in a layer-specific manner, whilst not altering the hallmarks of FH loss.

To elucidate the behaviour of the metabolic signature in response to the TRACER gradients, I applied linear modelling to the metabolite distribution over the layers (**Supplementary Figure 6D**). A metabolite that accumulates in a linear manner (median $R^2 > 0.7$) is labelled "UP", whilst the one that decreases linearly across the layers (median $R^2 > -0.7$) is labelled "DOWN". If there is no linear metabolite distribution, the metabolite is assigned to "No Linear Change". Moreover, metabolites that follow a linear distribution ("UP" or "DOWN") but have a sample variance > 0.2 and/or a median R^2 adjusted p- value > 0.05 are excluded. Following this approach, I sorted the 129 metabolites into

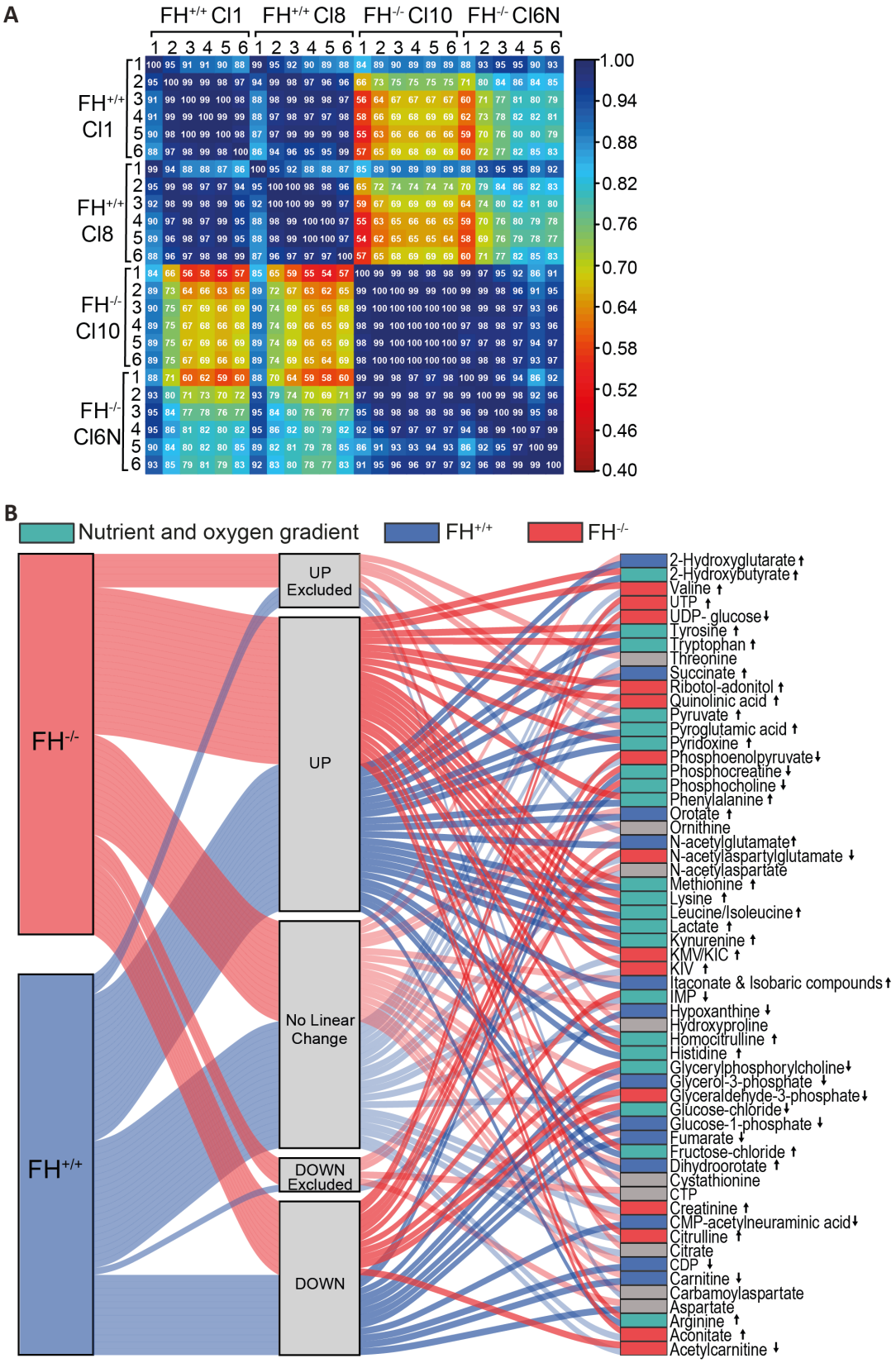


Figure 21: Layer-dependent metabolic signature in the TRACER.

(**A**) Correlation matrix comparing the patterns of 129 metabolites. (**B**) Alluvian plot of the metabolites that follow a linear pattern in $FH^{+/+}$ and/or $FH^{-/-}$ cells highlighting metabolites distribution driven by oxygen and nutrients gradients (green), by FH-proficiency (blue) and by FH loss (red).

behavioural patterns to understand which metabolites accumulate/decrease in a linear manner with the gradient of oxygen and nutrients, and selected metabolites that change linearly in at least one of the conditions ($FH^{+/+}$ or $FH^{-/-}$) (**Figure 21B**). The metabolites that follow the same linear distribution in $FH^{+/+}$ and $FH^{-/-}$ represent metabolites affected by the nutrient and oxygen gradients independent of the FH status (green, **Figure 21B**). As expected, I observed a linear increase of pyruvate and lactate (**Supplementary Figure 6E**) across the layers, which is likely due to increased aerobic glycolysis under reduced oxygen availability. Additionally, this results further strengthen the validity of the TRACER to mimic physiological conditions given that the lactate accumulation is detected intracellularly, whilst in 2D cultures increased lactate production under hypoxia can mostly be detected as lactate released into the media.

Together, these results show that the TRACER, combined with linear modelling, is a powerful tool to investigate how the extracellular microenvironment affects the metabolic landscape of cells.

3.3.3. D-2HG production is limited under physiological conditions in FH-deficient cells

Using this combined approach of TRACER and linear modelling, I then focused on metabolic features that are differentially altered in $FH^{+/+}$ (blue) and $FH^{-/-}$ (red) cells (**Figure 21B**). I observed that 2HG accumulates linearly in $FH^{+/+}$ cells, consistent with its putative role as a hypoxia marker, but not in $FH^{-/-}$ cells. To investigate whether this unexpected behaviour is caused by oxygen levels independently of the nutrient availability, I performed 2D metabolomics experiments comparing hypoxia (1% oxygen) *versus* normoxia (**Figure 22A-B**). Similar to the observations made in the TRACER, I detect an FH-deficient signature that is independent of hypoxia (**Figure 22A**). Yet, some metabolites are sensitive to hypoxia (**Figure 22B**). Noteworthy, 2HG and aKG are increased under hypoxia in both $FH^{-/-}$ and $FH^{+/+}$ clones (**Figure 22B-D**), whilst in the TRACER we only detected a linear metabolic accumulation of 2HG in $FH^{+/+}$ clones (**Figure 21B**, **Figure 22E**). This result is in line with the aKG pattern in $FH^{-/-}$ over the TRACER layers (**Figure 22F**), which is not due to nutrient deprivation since the glutamine level across the TRACER layers is consistent between $FH^{+/+}$ and $FH^{-/-}$ and does not change with the gradient of nutrients and oxygen (**Figure 22G**). As discussed above, 2HG is a metabolic marker of hypoxia and has been previously shown to accumulate under hypoxic conditions to facilitate the physiological adaptation to hypoxic stress[171]. Its production can be carried out *via* enzymes that convert aKG into the enantiomers L-2HG or D-2HG, including LDH and MDH, and PHGDH[171], respectively (**Figure 22H**). This conversion is based on enzyme promiscuity to catalyse this side reaction in addition to their

primary reaction and is favoured under low oxygen levels[171]. Moreover, acidification (lower pH), as often observed in hypoxic cells, has been connected to stimulate 2HG production[171]. I

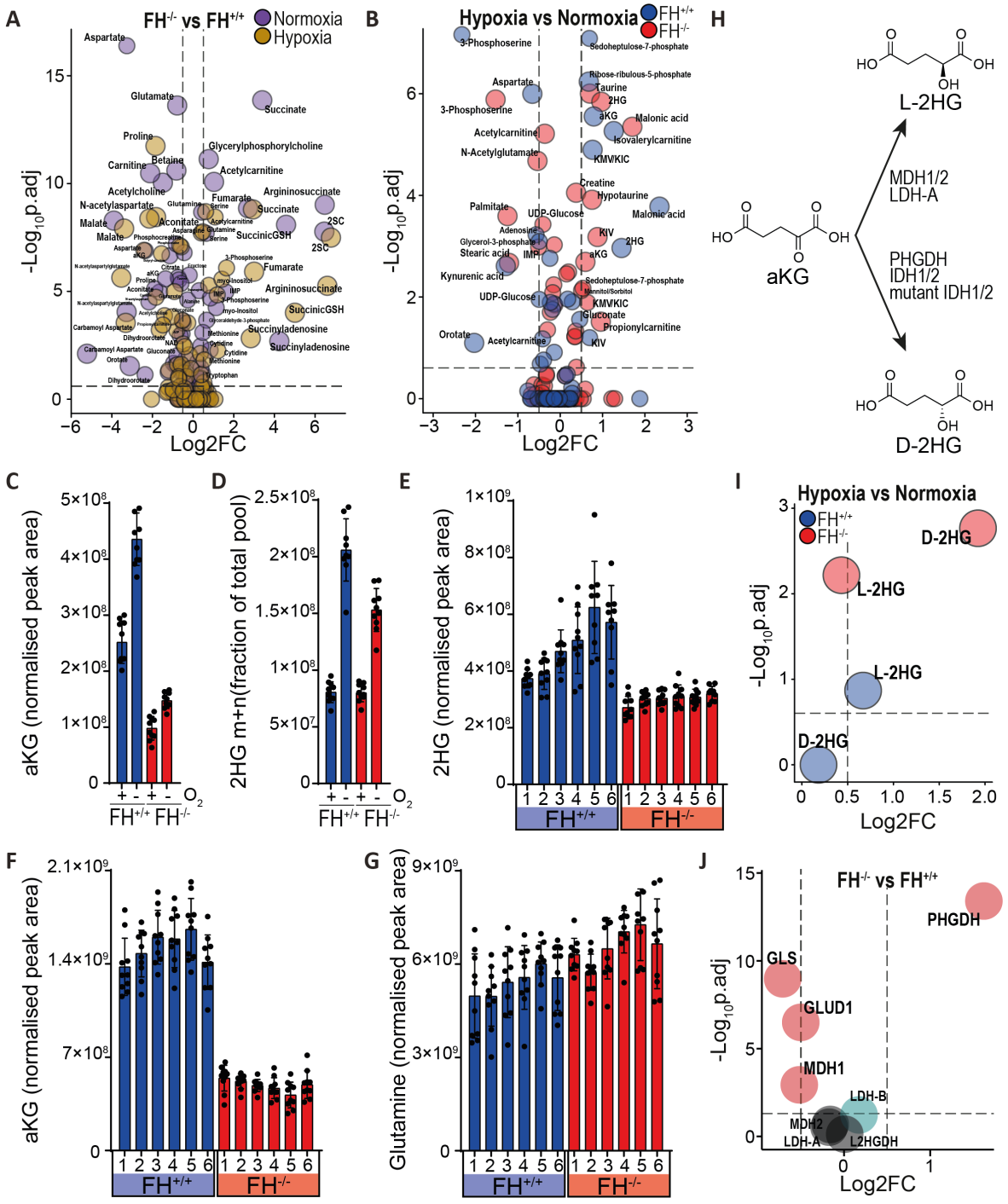


Figure 22: D-2HG is produced in FH-deficient cells and limited under physiological conditions.

(**A-B**) Volcano plots of 2D differential metabolomics results comparing $FH^{-/-}$ *versus* $FH^{+/+}$ (**A**) and hypoxia *versus* normoxia (**B**). (**C-D**) Intracellular metabolite levels of cells cultures in 2D under normoxia and hypoxia measured using LC-MS. (**E-G**) Intracellular metabolite levels of the TRACER measured using LC-MS. (**H**) Schematics of D-2HG and L-2HG production. (**I**) Volcano plot of differential metabolomics analysis of derivatised 2HG comparing hypoxia *versus* normoxia in 2D cultures measured using LC-MS. (**J**) Protein expression measured *via* TMT proteomics comparing $FH^{-/-}$ *versus* $FH^{+/+}$.

first performed 2HG derivatisation followed by mass spectrometry to understand which isomer is produced under these conditions. This approach revealed that under hypoxia D-2HG accumulates in FH-deficient cells, whilst in $FH^{+/+}$ clones it is L-2HG (**Figure 22I**). This result is in accordance with an increase in PHGDH protein expression in $FH^{-/-}$ clones (**Figure 22J**). PHGDH is an enzyme of the serine biosynthesis pathway and has previously been shown to produce D-2HG[171].

In summary, here I showed that in 2D cultures under hypoxia 2HG accumulates in $FH^{+/+}$ and $FH^{-/-}$ cells. However, this change is underpinned by different metabolic routing, with MDH/LDH likely giving rise to L2HG in $FH^{+/+}$ cells and PHGDH to D-2HG in $FH^{-/-}$ cells. Furthermore, the TRACER revealed that under more physiological 3D culture conditions, 2HG does not increase drastically in $FH^{-/-}$. Together, this revealed a potential role of PHGDH in FH-deficiency and how the function is altered under physiological environment compared to standard cell culture.

3.3.4. Discussion

During the process of tumour initiation and progression, cancer cells are exposed to harsh environmental conditions such as hypoxia and nutrient depletion, which are known to affect the metabolic landscape of a cell. Most of the studies carried out in FH-deficient cells were performed in 2D cultures exposed to atmospheric oxygen and non-limiting nutrients, potentially limiting their relevance to the human disease. To understand if environmental cues such as oxygen and nutrient concentrations influence the hallmarks of FH loss, I applied a novel analysis pipeline to the 3D model TRACER.

I observed that changes in oxygen or nutrient levels do not affect fumarate accumulation or the more general metabolic signature of FH loss. This result shows that FH loss is robust to environmental changes and the results obtained in 2D culture can predict cellular changes due to FH loss in patients. Indeed, most of the molecular signatures that we had previously identified in mouse and human cells recapitulate those obtained from human tissues. Additionally, I observed that, independently of the FH status, lactate accumulates across the TRACER layers in a linear manner, which is likely due to increased aerobic glycolysis under reduced oxygen availability. Interestingly, I also observed a linear accumulation of several amino acids, such as methionine, lysine, leucine/isoleucine, arginine, kynurenine and tryptophan, across the TRACER layers. This change was independent to the FH status and hence likely due to the nutrient and oxygen gradients. It has been previously proposed that under hypoxia the catabolic process of autophagy enables the breakdown of protein and lipid cells to generate nutrients that compensate for the energetic defect caused by impaired mitochondrial function[172]. It is tempting

to speculate that the $FH^{+/+}$ cells might undergo mitophagy, a specific form of autophagy of the mitochondria that has been linked to BNIP3 expression[173]. Yet, given that BNIP3 does not increase in the $FH^{-/-}$ cells across the TRACER, it is possible that the amino acids accumulation in $FH^{-/-}$ is not due to autophagy/mitophagy.

Investigating the general metabolic response to the combined gradient of oxygen and nutrients using linear modelling revealed a layer-dependent metabolic reprogramming that would have been difficult to detect using traditional 2D cultures or other 3D systems such as spheroids. For instance, D-2HG increases in 2D cultures under hypoxia in FH-deficient cells, whilst in the TRACER system this increase is negligible. Since reductive carboxylation is impaired in the HK2 $FH^{-/-}$ cells, it is likely that D-2HG is produced by PHGDH[171], which showed increased expression in these cells. Noteworthy, PHGDH is alternatively spliced in the FH-deficient cells and part of the mTOR response we detected using the SiRCle model. There is not much known about the different functions of PHGDH isoforms, but it has been shown that different levels of PHGDH protein expression potentiate cancer cell dissemination and metastasis in breast cancer[174]. The increase in PHGDH protein expression could be important to regenerate GSH *via* serine/glycine synthesis[175], which is crucial since GSH biosynthesis has been observed to be increased in FH-deficiency due to increased oxidative stress caused by GSH succination[58]. Moreover, PHGDH is a target of NRF2 and mTORC1, both pathways that I have discussed in this thesis to be dysregulated in the $FH^{-/-}$ cells. Given the many regulators of mTORC1, such as growth factors and nutrient sensing (e.g. amino acids)[161], and the many pathways it is known to target, it is hard to draw any conclusions here. Yet, it is important to investigate mTOR regulation based on the alternative splicing signatures I observed (Results Chapter 2) and understand its connection to the PHGDH expression in the future. It was recently shown that PHGDH protein expression plays a role in cancer cell dissemination and metastasis[174]. Moreover, PHGDH coordinated serine synthesis has been connected to the one-carbon metabolism and to reduce the serine-derived incorporation of nucleotides of one-carbon units[176]. Given that the mitochondria play a key role in one-carbon metabolism and that the nucleotide biosynthesis is altered in FH-deficient cells (**Figure 21B**), it would be important to further understand the role of serine biosynthesis in this system. Finally, why 2HG production is impaired in the $FH^{-/-}$ cells in the TRACER remains unclear. It is possible that when nutrient and oxygen are scarce, the diversion of glucose towards serine catabolism *via* PHGDH is limited allowing a higher glycolytic flux. It will be interesting to investigate what are the biochemical determinants of this possible metabolic diversions

Perspectives

THIS study presents a novel human FH-deficient cell model based on non-transformed human epithelial kidney cells. The extensive cellular characterisation and comparison to other models highlight the hallmarks of FH loss. My findings expand the current understanding of the gene expression landscape underlying the epigenetic, transcriptional and translational rewiring after FH loss, as well as the metabolic rewiring in physiological environment.

These results raise critical points regarding the current analysis methods used to elucidate the signalling pathways responsible to activate oncogenic cascades during the process of tumorigenesis. Indeed, I could show that upon FH loss the cellular rewiring occurs at different regulatory levels and the phenotype-defining protein expression does not always correlate with the transcriptional or methylation landscape. These results enable a deeper understanding of the regulatory origin *via* which a signalling pathway is induced. It is important to mention that my study is not devoid of limitations and more experiments are required to improve our understanding of the transformation process induced by mitochondrial dysfunction such as FH loss.

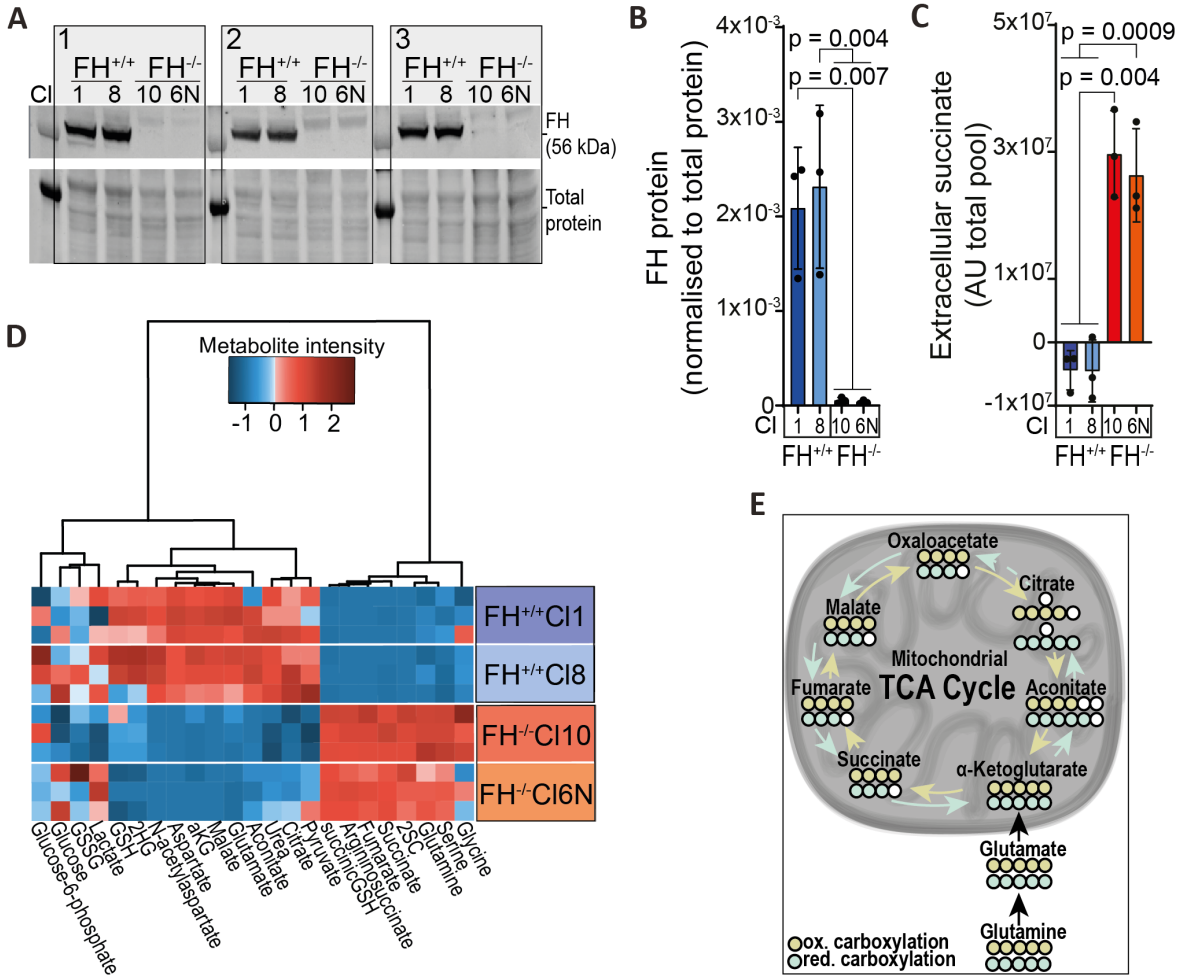
My findings also expand the general understanding of FH loss and the downstream metabolic and oncogenic rewiring. It is clear that tumour microenvironment affects the phenotype of cancer cells and that metabolic rewiring is essential for tumour cells to strive even under harsh environmental conditions. However, the impact of nutrient and oxygen deprivation on FH-deficient cells' behaviour was unknown. Here I assessed for the first time the impact of nutrient and oxygen gradients on the metabolic and oncogenic rewiring after FH loss. My work demonstrated that the metabolic signature of FH loss is robust to environmental cues, which is in line with the conservation of this signature between cellular models and *in vivo* tumours. I have also observed metabolic changes that occurred only under the combined deprivation of oxygen and nutrients. More work is needed to further understand the inderpinning molecular drivers of these changes and how to target these potential metabolic liabilities.

In conclusion, my study presents a comprehensive characterisation of a novel model of FH loss and elucidated which of the emerging phenotypes is regulated *via* hypermethylation, transcriptional rewiring, translational control and/or PTMs. Moreover, I have shown that the

main metabolic hallmarks of FH loss are independent of environmental conditions under which cells are grown. Although further investigations are needed to fully understand the impact of the different oncogenic pathways in the tumorigenesis of HLRCC, my study presents several novel approaches to dissect the metabolic and gene expression landscapes of FH loss that can be harnessed to find liabilities during tumorigenesis.

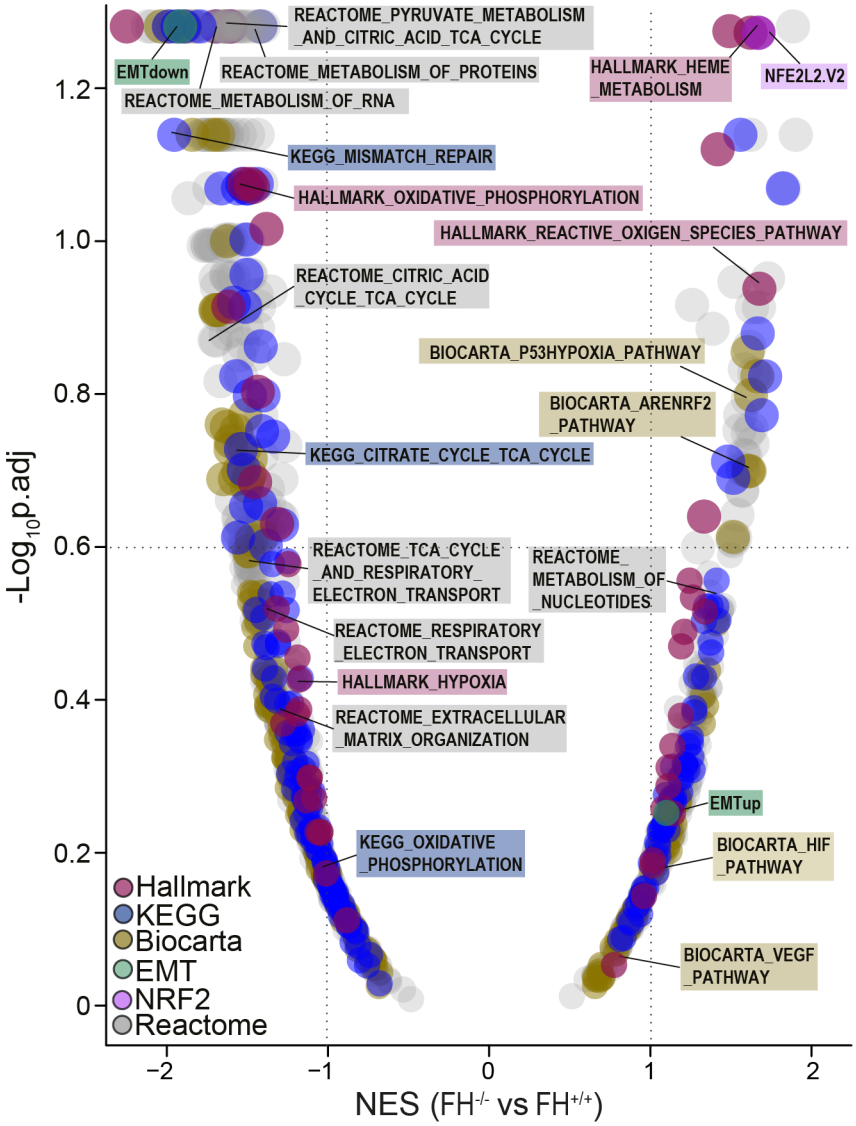
In the future, I will apply the SiRCle model also to other datasets, e.g. TCGA data, to dissect the different regulatory patterns across different tissues. At the same time I will perform further experiments validating the alternative splicing signature after FH loss. In order to further understand the role of PHGDH in serine biosynthesis and 2HG production I will use PHGDH inhibitors and U-^{13}C-Glucose labelling to further dissect this metabolic axis in light of FH loss.

Supplementary Figures



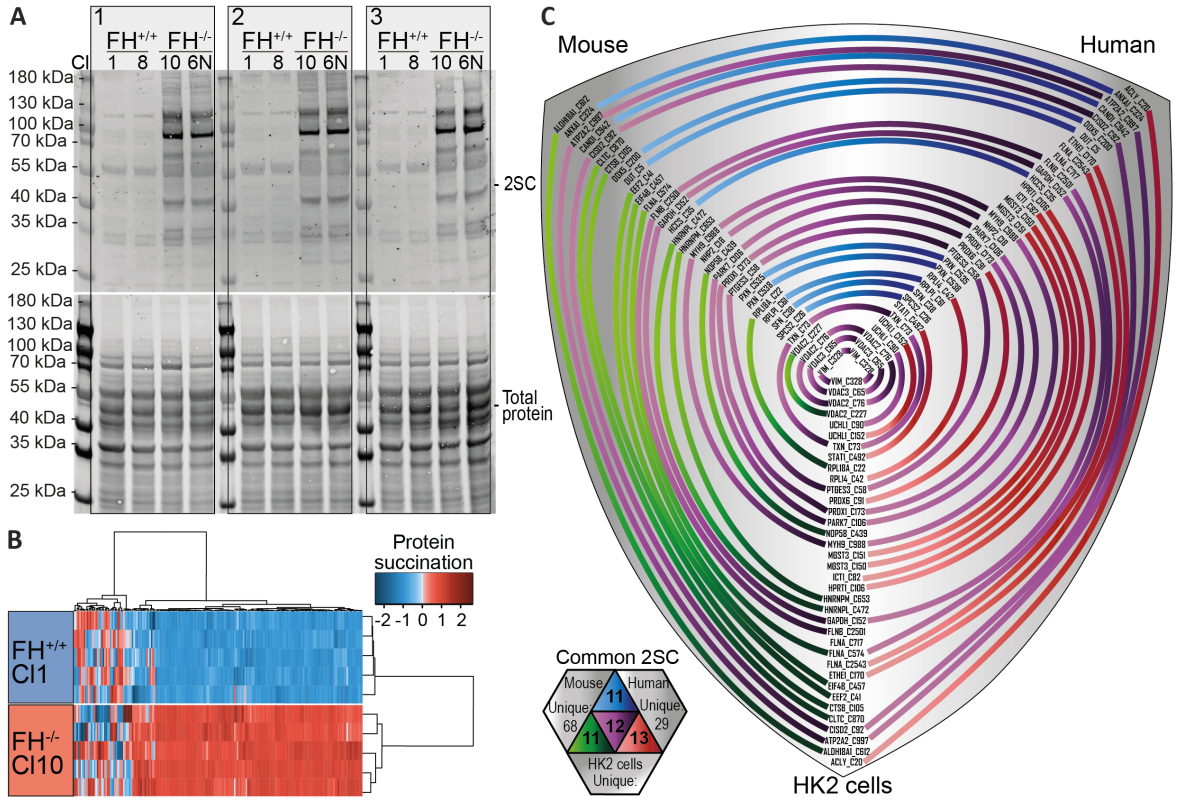
Supplementary Figure 1: Metabolic adaptation phase after FH loss.

(A + B) FH protein expression measured using western blotting (A) and quantified (B). **(C)** Extracellular succinate measured by LC-MS. **(D)** Heatmap of the key metabolite signature of FH loss measured by LC-MS. **(E)** Schematics of the U^{13}-C-Glutamine labelling distribution in the TCA cycle. Results for B and C are presented as mean ± s.d. Adjusted p-value (p) is calculated using one-way ANOVA and the Tukey's multiple comparison test.



Supplementary Figure 2: FH-deficient RNA signature.

Volcano plot illustrating the profile of the different gene sets (colour coded) after performing GSEA analysis on the pooled signatures comparing $FH^{-/-}$ versus $FH^{+/+}$. Results were obtained from five biological replicates. Adjusted p-value (p.adj) is the result of the GSEA analysis.

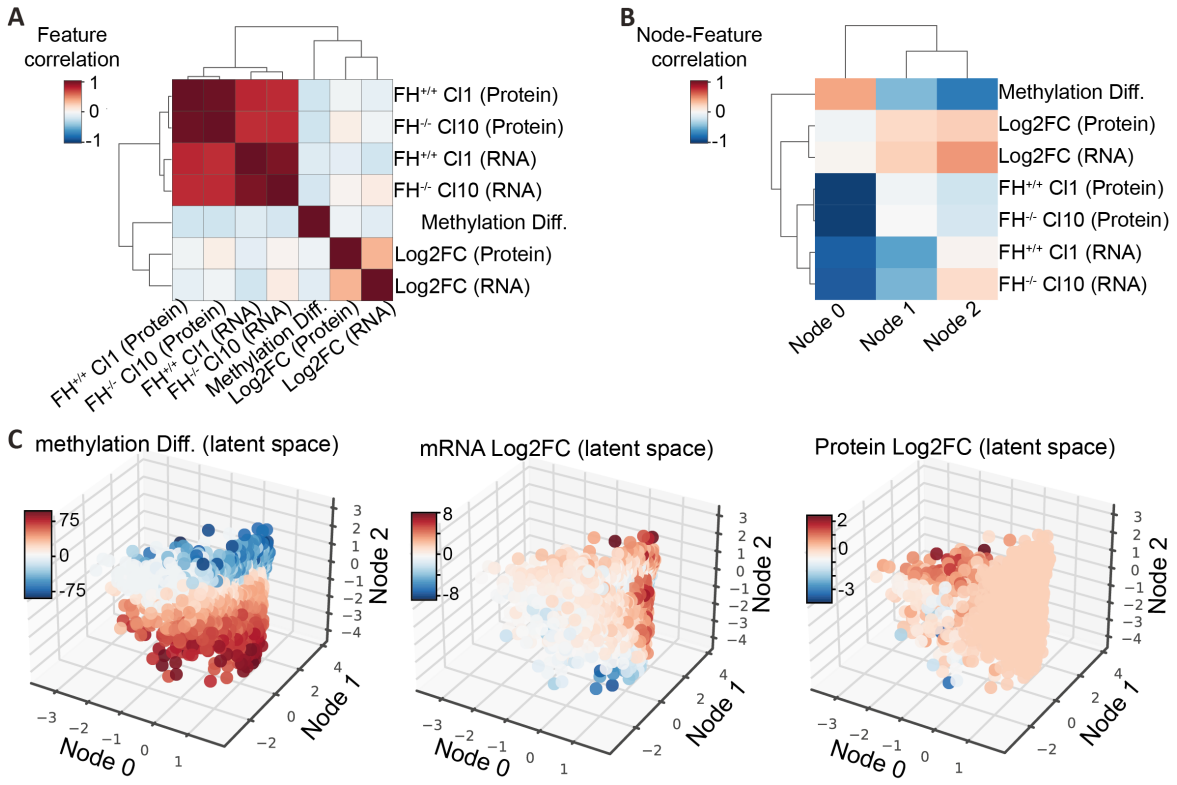


Supplementary Figure 3: Fumarate accumulation results in succination.

(A) 2SC modification on protein residues measured using western blotting **(B)** Heatmap of succinated protein residues measured *via* TMT proteomics comparing $FH^{-/-}$ to $FH^{+/+}$. **(C)** Hyphe plot showing commonly succinated proteins detected in Fh1-deficient mouse cells, human FH-deficient cells and our novel FH-deficient HK2 clones.

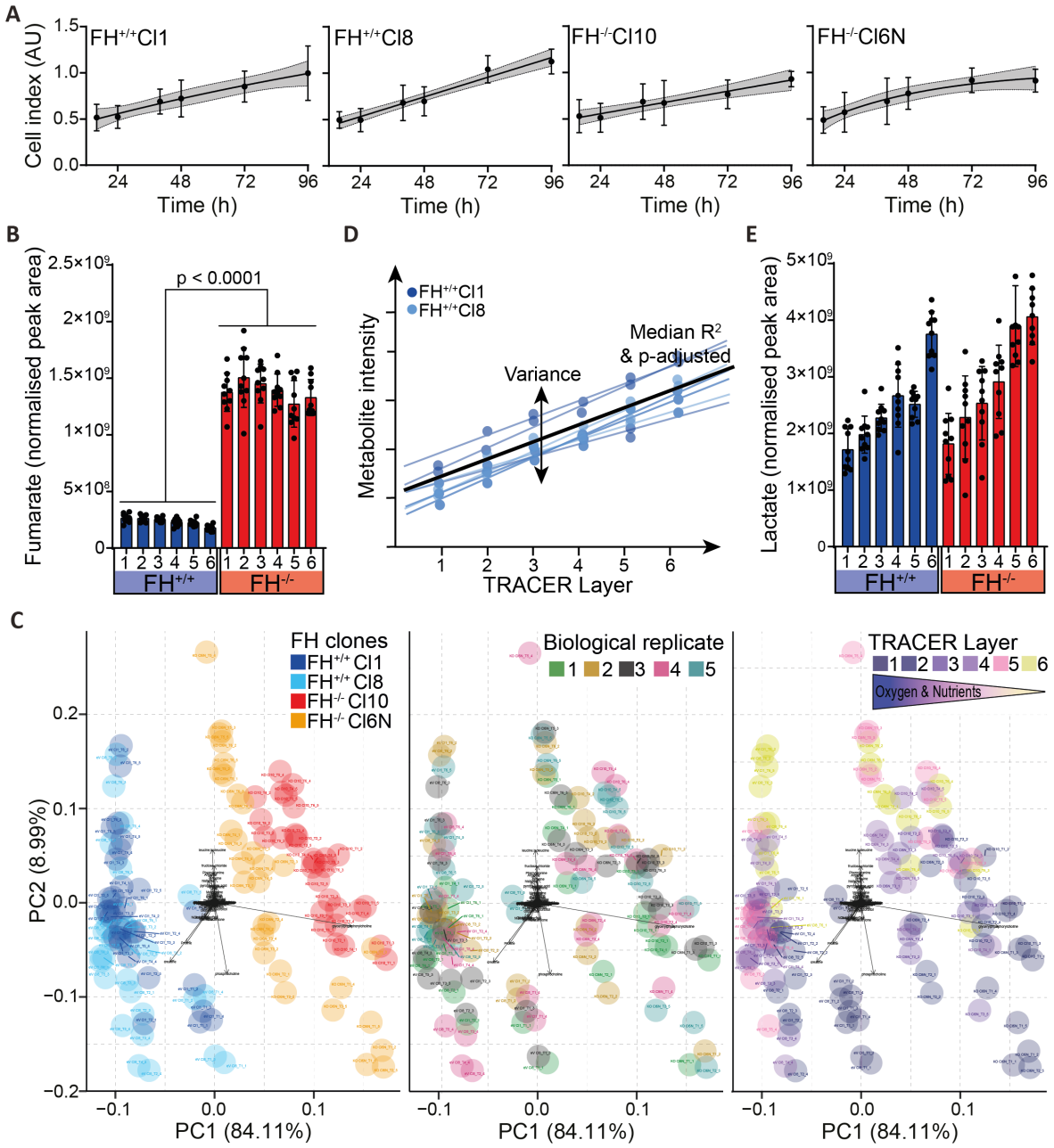
Supplementary Figure 4: Heatplot of the ORA results.

Heatplot of the ORA results for the genes in the (**A**) TPDS cluster and (**B**) TPDE cluster.



Supplementary Figure 5: The VAE model quality.

(**A**) Correlation matrix of a representative selection of the VAE input features. (**B**) Correlation matrix of the. VAE input features and the nodes of the latent space (**C**) Scatterplot of the three nodes of the VAE latent space colour coded by methylation difference (left), RNA Log2FC (middle) and protein Log2FC (right).



Supplementary Figure 6: TRACER

Growth curve of the individual clones. (B) Intracellular fumarate levels measured using LC-MS. (C) PCA plots of the LC-MS results colour coded for the clones (left), the biological replicates (middle) and the TRACER layer (right). (D) Schematics of the Pearson linear modelling. (E) Intracellular lactate levels measured using LC-MS. Results for A, B and E are presented as mean ± s.d. Adjusted p-value (p) in B and E is calculated using one-way ANOVA and the Tukey's multiple comparison test.

Acknowledgements

FIRST and foremost, I am extremely thankful to my supervisor and mentor Dr. Christian Frezza for offering me this experience that made me the scientist and person I am today. Christian was the person who got me interested in the field of cancer metabolism with his inspiring personality and in the past years Christian has been an invaluable mentor who taught me to navigate my way through different obstacles. I am extremely grateful for Christian being enormously understanding and patient with me not just in scientific manners but also in life decisions. He gave me all the scientific freedom I needed to grow as a scientist, whilst at the same time guiding me when necessary. His life lessons where often embedded in quotes or comparisons. Indeed, as Christian would say: Similar to a caterpillar that learns to walk using its many legs I learnt to juggle all my different responsibilities.

I am deeply grateful for the extensive support of Dr. Lorea Valcarcel during my PhD. Lorea has always supported me by sparing her time to help with experiments or discussions inside as well as outside of the lab. Her constant understanding and advice has been invaluable and I am glad I met such a great scientist I can not only call a colleague but also a friend.

I want to thank Dr. Dylan Ryan and Dr. Connor Rogerson for the inspiring scientific discussion spanning all fields of science. Sharing the excitement about science and our projects over lunch, coffee or dinner led to fruitful ideas. I am grateful that they always made me remember why I started working in science in first place: The excitement of novel discoveries. Thanks to both for sharing their excitement with me.

Furthermore, I am very thankful to Dr. Laura Tronci, Efterpi Nikitopoulou Dr. Ana Sofia Costa, and Dr. Ming Yang for their continuous input and the countless metabolomics samples they processed for me. It was amazing, how patiently they always addressed all the questions I had.

I am deeply grateful to Dr. Vincent Zecchini and Dr. Marco Sciacovelli for being my steady constant in the lab. With their extensive experience they helped me a lot to find my way around the lab when I started in 2017.

I want to thank all the people that helped me collecting and processing the data for this thesis:

A special thanks goes to Ariane Mora for her help with the multi-omics model and for teaching me a lot of new bioinformatics tools. I truly enjoyed our lively discussions, which made the project lots of fun and took it onto another level. I am thankful for her patience with me and for the dedication to meet at really early or late hours of the day due to the time difference between Australia and the UK.

I also thank Lucas Maddalena, who not just helped me perform my experiments under hypoxia but supported me during my PhD journey.

I am thankful to Tim Young, who made sure that the lab environment was always in a great shape to support my work and helped me with my experiments.

I am grateful for the fruitful collaboration with Dr. Alexander von Kriegsheim from the University of Edinburgh. He was incredibly knowledgeable and ensured great proteomics measurements.

I am also thankful for the many great collaborators across Cambridge, who made this work possible. Especially Sabrina Rossie and Dr. Charly Massie for performing the bisulfide sequencing experiments. Moreover, to the many facilities of the Cancer Research UK.

I am also very thankful for the many experiences I made in different laboratories across the globe:

I am grateful for the enormously welcoming time in Toronto in the laboratory of Dr. Alison McGuigan and for teaching me how to build the TRACER model. Especially to all the students in her lab, who ensured I have a great experience inside the lab, but also exploring Toronto.

I am thankful to the laboratory of Dr. Alasdair Russell, who accommodated me during my first years. There I learnt a great deal about CRISPR and genome editing techniques, but they also made me feel part of their team including me in all lab activities. Special thanks go to Amy Smith, for taking the time to teach me all the magic of cloning and sequence analysis.

I also thank the Marie Curie Network TRANSMIT (Translating the role of Mitochondria in Tumorigenesis) for bringing together a group of curious students in numerous courses across Europe including Italy, Belgium, Scotland and Austria. Attending all the activities and training events has also taught me to become more efficient with my time in the lab. I am also thankful for the funding they provided that made my PhD possible: The European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 722605.

I would like to offer my special thanks to all current members and alumni of the Frezza Lab: Edo, Lorea, Dylan, Ariane, Connor, Marco, Vinny, Effie, Laura, Ming, Sofia, Marc, Lucas, Annie, Cissy, Tim, Ming, Elsa, Irene, Clement, Isaac and Christian. I especially enjoyed the

discussions in lab meetings and the many shared lunches. Thanks for teaching me about new cultures, food, books and music.

My warmest thanks are going to Dr. Markus Hecht for supporting me outside of the lab in the journey through my PhD. I would like to express my sincere gratitude that Markus was always there to discuss my scientific findings or struggles and patiently listened to my complaints and concerns. His positive attitude and care made me enjoy my PhD journey even at difficult times.

I would like to express my special thanks to my family, my brothers and especially to my parents Renate and Heymo: Vielen Dank dass ihr immer an mich geglaubt habt und mich in jeder Situation unterstützt habt.

Last but not least I would like to thank all my friends across the globe for always being supportive. Many thanks to my housemates, Maria, Karel, Dani, Marco and Paula who had my back in any difficult situation. I am also very thankful to Hafsa, Vanessa, Elsa, Steffie, Leonie, Pierre, Andi, Adi, Peter, Cordula, Aiske, Thuylinh and Marvin.

Bibliography

1. Hanahan, D. & Weinberg, R. A. The Hallmarks of Cancer. *Cell* **100**, 57–70 (2000).
2. Vogelstein, B. *et al.* Cancer Genome Landscapes. *Science (80-.).* **339**, 1546–1558 (2013).
3. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).
4. Boveri Theodor. Zur Frage der Entstehung maligner Tumoren. *Verlag von Gustav Fischer* 29–32 (1914).
5. Brücher, B. L. D. M. & Jamall, I. S. Somatic Mutation Theory - Why it's Wrong for Most Cancers. *Cell. Physiol. Biochem.* **38**, 1663–1680 (2016).
6. Knudson, A. G. Mutation and Cancer: Statistical Study of Retinoblastoma. *Proc. Natl. Acad. Sci.* **68**, 820–823 (1971).
7. Wishart, D. S. Is Cancer a Genetic Disease or a Metabolic Disease? *EBioMedicine* **2**, 478–479 (2015).
8. Boroughs, L. K. & DeBerardinis, R. J. Metabolic pathways promoting cancer cell survival and growth. *Nat. Cell Biol.* **17**, 351–359 (2015).
9. Pavlova, N. N. & Thompson, C. B. The Emerging Hallmarks of Cancer Metabolism. *Cell Metab* **23**, 27–47
10. E. Freund. Zur Diagnose des Carcinoms. *Wiener Medizinische Blätter* **9**, 268–269 (1885).
11. Wassermann, A. V., Keysser, F., & Wassermann, M. Beiträge zum problem: Geschwülste von der blutbahn aus therapeutisch zu beeinflussen. *Dtsch. Medizinische Wochenzeitschrift* **37**, 2389–2391 (1911).
12. Van Alstyne, E. V & Beebe, S. P. Diet Studies in transplantable Tumors : I. The Effect of non-carbohydrate Diet upon the Growth of transplantable Sarcoma in Rats. *J. Med. Res.* **29**, 217–232 (1913).
13. Woglom, W. H. DIET AND TUMOR GROWTH. *J. Exp. Med.* **22**, 766–779 (1915).
14. Warburg and Negelein, E., O. P. K. Über den Stoffwechsel der Carcinomzelle. *Biochem Zeitschr* **152**, 309–344 (1924).
15. Frezza, C. Mitochondrial metabolites: undercover signalling molecules. *Interface Focus* **7**, 20160100 (2017).
16. Warburg, O. On the Origin of Cancer Cells. *Science (80-.).* **123**, 309–314 (1956).
17. Sciacovelli, M., Gaude, E., Hilvo, M. & Frezza, C. *The Metabolic Alterations of Cancer Cells*. *Methods in enzymology* **542**, (2014).
18. Joshi, S. *et al.* The Genomic Landscape of Renal Oncocytoma Identifies a Metabolic Barrier to Tumorigenesis. *Cell Rep.* **13**, 1895–1908 (2015).
19. Bajzikova, M. *et al.* Reactivation of Dihydroorotate Dehydrogenase-Driven Pyrimidine Biosynthesis Restores Tumor Growth of Respiration-Deficient Cancer Cells. *Cell Metab.* **29**, 399-416.e10 (2019).
20. Martínez-Reyes, I. *et al.* Mitochondrial ubiquinol oxidation is necessary for tumour growth. *Nature* **585**, 288–292 (2020).
21. Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science (80-.).* **324**, 1029–1033 (2009).
22. Romani, P., Valcarcel-Jimenez, L., Frezza, C. & Dupont, S. Crosstalk between mechanotransduction and metabolism. *Nat. Rev. Mol. Cell Biol.* **22**, 22–38 (2021).
23. Fendt, S.-M., Frezza, C. & Erez, A. Targeting Metabolic Plasticity and Flexibility Dynamics for Cancer Therapy. *Cancer Discov.* **10**, 1797 LP – 1807 (2020).
24. Faubert, B., Solmonson, A. & DeBerardinis, R. J. Metabolic reprogramming and cancer

progression. *Science (80-.).* **368**, eaaw5473 (2020).
25. Valcarcel-Jimenez, L., Gaude, E., Torrano, V., Frezza, C. & Carracedo, A. Mitochondrial Metabolism: Yin and Yang for Tumor Progression. *Trends Endocrinol. Metab.* **28**, 748–757 (2017).
26. Sheth, R. A., Hesketh, R., Kong, D. S., Wicky, S. & Oklu, R. Barriers to Drug Delivery in Interventional Oncology. *J. Vasc. Interv. Radiol.* **24**, 1201–1207 (2013).
27. Russell, S., Wojtkowiak, J., Neilson, A. & Gillies, R. J. Metabolic Profiling of healthy and cancerous tissues in 2D and 3D. *Sci Rep* **7**, 15285 (2017).
28. Gaude, E. & Frezza, C. Tissue-specific and convergent metabolic transformation of cancer correlates with metastatic potential and patient survival. *Nat. Commun.* **7**, 13041 (2016).
29. Hu, J. *et al.* Heterogeneity of tumor-induced gene expression changes in the human metabolic network. *Nat. Biotechnol.* **31**, 522–529 (2013).
30. Peng, X. *et al.* Molecular Characterization and Clinical Relevance of Metabolic Expression Subtypes in Human Cancers. *Cell Rep.* **23**, 255-269.e4 (2018).
31. Rosario, S. R. *et al.* Pan-cancer analysis of transcriptional metabolic dysregulation using The Cancer Genome Atlas. *Nat. Commun.* **9**, 5330 (2018).
32. Sciacovelli, M. & Frezza, C. Oncometabolites: Unconventional triggers of oncogenic signalling cascades. *Free Radic. Biol. Med.* **100**, 175–181 (2016).
33. Yang, M., Soga, T. & Pollard, P. J. Oncometabolites: linking altered metabolism with cancer. *J. Clin. Invest.* **123**, 3652–3658 (2013).
34. Ryan, D. G. *et al.* Coupling Krebs cycle metabolites to signalling in immunity and cancer. *Nat. Metab.* **1**, 16–33 (2019).
35. Arts, R. J. W. *et al.* Glutaminolysis and Fumarate Accumulation Integrate Immunometabolic and Epigenetic Programs in Trained Immunity. *Cell Metab.* **24**, 807–819 (2016).
36. Picaud, S. *et al.* Structural basis of fumarate hydratase deficiency. *J Inherit Metab Dis* **34**, 671–676 (2011).
37. Schmidt, L. S. & Linehan, W. M. Hereditary leiomyomatosis and renal cell carcinoma. *Int J Nephrol Renov. Dis* **7**, 253–260 (2014).
38. Tomlinson IP1, Alam NA, Rowan AJ, Barclay E, Jaeger EE, Kelsell D, Leigh I, Gorman P, Lamlum H, Rahman S, Roylance RR, Olpin S, Bevan S, Barker K, Hearle N, Houlston RS, Kiuru M, Lehtonen R, Karhu A, Vilkki S, Laiho P, Eklund C, Vierimaa O, Aittomäki K, H, A. L. M. L. C. Germline mutations in FH predispose to dominantly inherited uterine fibroids, skin leiomyomata and papillary renal cell cancer. *Nat. Genet.* **30**, 406 (2002).
39. Menko, F. H. *et al.* Hereditary leiomyomatosis and renal cell cancer (HLRCC): renal cancer risk, surveillance and treatment. *Fam Cancer* **13**, 637–644 (2014).
40. Castro-Vega, L. J. *et al.* Germline mutations in FH confer predisposition to malignant pheochromocytomas and paragangliomas. *Hum. Mol. Genet.* **23**, 2440–2446 (2014).
41. Clark, G. R. *et al.* Germline FH mutations presenting with pheochromocytoma. *J Clin Endocrinol Metab* **99**, E2046-50 (2014).
42. Zhang, J. *et al.* Germline Mutations in Predisposition Genes in Pediatric Cancer. *N. Engl. J. Med.* **373**, 2336–2346 (2015).
43. Fieuw, A. *et al.* Identification of a novel recurrent 1q42.2-1qter deletion in high risk MYCN single copy 11q deleted neuroblastomas. *Int. J. Cancer* **130**, 2599–2606 (2011).
44. Ha, Y.-S. *et al.* Downregulation of Fumarate Hydratase Is Related to Tumorigenesis in Sporadic Renal Cell Cancer. *Urol. Int.* **90**, 233–239 (2013).
45. Hu, J. *et al.* Heterogeneity of tumor-induced gene expression changes in the human metabolic network. *Nat. Biotechnol.* **31**, 522 (2013).
46. Frezza, C., Pollard, P. J. & Gottlieb, E. Inborn and acquired metabolic defects in cancer. *J Mol Med* **89**, 213–220 (2011).
47. Schmidt, C., Sciacovelli, M. & Frezza, C. Fumarate hydratase in cancer: A multifaceted tumour suppressor. *Semin. Cell Dev. Biol.* (2019). doi:https://doi.org/10.1016/j.semcdb.2019.05.002
48. Akram, M. Citric Acid Cycle and Role of its Intermediates in Metabolism. *Cell Biochem. Biophys.* **68**, 475–478 (2014).
49. Frezza, C. *et al.* Haem oxygenase is synthetically lethal with the tumour suppressor fumarate hydratase. *Nature* **477**, 225–228 (2011).
50. Yang, Y. *et al.* Metabolic reprogramming for producing energy and reducing power in fumarate

hydratase null cells from hereditary leiomyomatosis renal cell carcinoma. *PLoS One* **8**, e72179–e72179 (2013).
51. Zheng, L. *et al.* Reversed argininosuccinate lyase activity in fumarate hydratase-deficient cancer cells. *Cancer Metab* **1**, 12 (2013).
52. Mullen, A. R. *et al.* Reductive carboxylation supports growth in tumour cells with defective mitochondria. *Nature* **481**, 385–388 (2011).
53. Ternette, N. *et al.* Inhibition of Mitochondrial Aconitase by Succination in Fumarate Hydratase Deficiency. *Cell Rep.* **3**, 689–700 (2013).
54. Adam, J. *et al.* A Role for Cytosolic Fumarate Hydratase in Urea Cycle Metabolism and Renal Neoplasia. *Cell Rep.* **3**, 1440–1448 (2013).
55. Tyrakis, P. A. *et al.* Fumarate Hydratase Loss Causes Combined Respiratory Chain Defects. *Cell Rep* **21**, 1036–1047 (2017).
56. Gonçalves, E. *et al.* Post-translational regulation of metabolism in fumarate hydratase deficient cancer cells. *Metab. Eng.* **45**, 149–157 (2018).
57. Islam, M. S., Leissing, T. M., Chowdhury, R., Hopkinson, R. J. & Schofield, C. J. 2-Oxoglutarate-Dependent Oxygenases. *Annu. Rev. Biochem.* **87**, 585–620 (2018).
58. Zheng, L. *et al.* Fumarate induces redox-dependent senescence by modifying glutathione metabolism. *Nat Commun* **6**, 6001 (2015).
59. Jin, L. *et al.* Glutamate Dehydrogenase 1 Signals through Antioxidant Glutathione Peroxidase 1 to Regulate Redox Homeostasis and Tumor Growth. *Cancer Cell* **27**, 257–270 (2015).
60. Nowicki, S. & Gottlieb, E. Oncometabolites: tailoring our genes. *FEBS J* **282**, 2796–2805 (2015).
61. Isaacs, J. S. *et al.* HIF overexpression correlates with biallelic loss of fumarate hydratase in renal cancer: novel role of fumarate in regulation of HIF stability. *Cancer Cell* **8**, 143–153 (2005).
62. Kim, J., Tchernyshyov, I., Semenza, G. L. & Dang, C. V. HIF-1-mediated expression of pyruvate dehydrogenase kinase: A metabolic switch required for cellular adaptation to hypoxia. *Cell Metab.* **3**, 177–185 (2006).
63. Papandreou, I., Cairns, R. A., Fontana, L., Lim, A. L. & Denko, N. C. HIF-1 mediates adaptation to hypoxia by actively downregulating mitochondrial oxygen consumption. *Cell Metab.* **3**, 187–197 (2006).
64. Xie, H. *et al.* LDH-A inhibition, a therapeutic strategy for treatment of hereditary leiomyomatosis and renal cell cancer. *Mol Cancer Ther* **8**, 626–635 (2009).
65. Semenza, G. L. Targeting HIF-1 for cancer therapy. *Nat. Rev. Cancer* **3**, 721 (2003).
66. Adam, J. *et al.* Renal Cyst Formation in Fh1-Deficient Mice Is Independent of the Hif/Phd Pathway: Roles for Fumarate in KEAP1 Succination and Nrf2 Signaling. *Cancer Cell* **20**, 524–537 (2011).
67. Rasmussen, K. D. & Helin, K. Role of TET enzymes in DNA methylation, development, and cancer. *Genes Dev* **30**, 733–750 (2016).
68. Park, S. Y., Park, J.-W. & Chun, Y.-S. Jumonji histone demethylases as emerging therapeutic targets. *Pharmacol. Res.* **105**, 146–151 (2016).
69. Sciacovelli, M. *et al.* Fumarate is an epigenetic modifier that elicits epithelial-to-mesenchymal transition. *Nature* **537**, 544–547 (2016).
70. Linehan, W. M. *et al.* Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. *N Engl J Med* **374**, 135–145 (2015).
71. Ricketts, C. J. *et al.* The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. *Cell Rep.* **23**, 313-326.e5 (2018).
72. Brabletz, T., Kalluri, R., Nieto, M. A. & Weinberg, R. A. EMT in cancer. *Nat. Rev. Cancer* (2018). doi:10.1038/nrc.2017.118
73. Rayess, H., Wang, M. B. & Srivatsan, E. S. Cellular senescence and tumor suppressor gene p16. *Int. J. Cancer* **130**, 1715–1725 (2012).
74. Alderson, N. L. *et al.* S-(2-Succinyl)cysteine: a novel chemical modification of tissue proteins by a Krebs cycle intermediate. *Arch Biochem Biophys* **450**, 1–8 (2006).
75. Blatnik, M., Frizzell, N., Thorpe, S. R. & Baynes, J. W. Inactivation of glyceraldehyde-3-phosphate dehydrogenase by fumarate in diabetes: formation of S-(2-succinyl)cysteine, a novel chemical modification of protein and possible biomarker of mitochondrial stress. *Diabetes* **57**,

41–49 (2008).
76. Bardella, C. *et al.* Aberrant succination of proteins in fumarate hydratase-deficient mice and HLRCC patients is a robust biomarker of mutation status. *J Pathol* **225**, 4–11 (2011).
77. Adam, J. *et al.* Renal cyst formation in Fh1-deficient mice is independent of the Hif/Phd pathway: roles for fumarate in KEAP1 succination and Nrf2 signaling. *Cancer Cell* **20**, 524–537 (2011).
78. Ooi, A. *et al.* An Antioxidant Response Phenotype Shared between Hereditary and Sporadic Type 2 Papillary Renal Cell Carcinoma. *Cancer Cell* **20**, 511–523 (2011).
79. Kerins, M. J. *et al.* Fumarate Mediates a Chronic Proliferative Signal in Fumarate Hydratase-Inactivated Cancer Cells by Increasing Transcription and Translation of Ferritin Genes. *Mol Cell Biol* **37**, (2017).
80. Tomczak, K., Czerwińska, P. & Wiznerowicz, M. Review The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Współczesna Onkol.* **1A**, 68–77 (2015).
81. Creixell, P. *et al.* Pathway and network analysis of cancer genomes. *Nat. Methods* **12**, 615–621 (2015).
82. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545 LP – 15550 (2005).
83. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
84. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
85. Kanehisa, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
86. Huang, S., Chaudhary, K. & Garmire, L. X. More Is Better: Recent Progress in Multi-Omics Data Integration Methods. *Front. Genet.* **8**, (2017).
87. Simidjievski, N. *et al.* Variational Autoencoders for Cancer Data Integration: Design Principles and Computational Practice. *Front. Genet.* **10**, 1205 (2019).
88. Diederik P Kingma, M. W. Auto-Encoding Variational Bayes. *Proc. 2nd Int. Conf. Learn. Represent.* (2013).
89. Ghojogh, B., Ghodsi, A., Karray, F. & Crowley, M. Factor Analysis, Probabilistic Principal Component Analysis, Variational Inference, and Variational Autoencoder: Tutorial and Survey. (2021).
90. Way, G. P. & Greene, C. S. Extracting a biologically relevant latent space from cancer transcriptomes with variational autoencoders. in *Biocomputing 2018* 80–91 (WORLD SCIENTIFIC, 2018). doi:10.1142/9789813235533_0008
91. Truitt, M. L. & Ruggero, D. New frontiers in translational control of the cancer genome. *Nat. Rev. Cancer* **16**, 288–304 (2016).
92. Cunningham, J. T., Moreno, M. V., Lodi, A., Ronen, S. M. & Ruggero, D. Protein and Nucleotide Biosynthesis Are Coupled by a Single Rate-Limiting Enzyme, PRPS2, to Drive Cancer. *Cell* **157**, 1088–1103 (2014).
93. Truitt, M. L. *et al.* Differential Requirements for eIF4E Dose in Normal Development and Cancer. *Cell* **162**, 59–71 (2015).
94. Xu, Y. & Ruggero, D. The Role of Translation Control in Tumorigenesis and Its Therapeutic Implications. *Annu. Rev. Cancer Biol.* **4**, 437–457 (2020).
95. Vaklavas, C., Blume, S. W. & Grizzle, W. E. Translational Dysregulation in Cancer: Molecular Insights and Potential Clinical Applications in Biomarker Development. *Front. Oncol.* **7**, 158 (2017).
96. Eales, K. L., Hollinshead, K. E. R. & Tennant, D. A. Hypoxia and metabolic adaptation of cancer cells. *Oncogenesis* **5**, e190 (2016).
97. Denko, N. C. Hypoxia, HIF1 and glucose metabolism in the solid tumour. *Nat. Rev. Cancer* **8**, 705–713 (2008).
98. Gottfried, E. *et al.* Tumor-derived lactic acid modulates dendritic cell activation and antigen expression. *Blood* **107**, 2013 LP – 2021 (2006).
99. Sullivan, M. R. *et al.* Quantification of microenvironmental metabolites in murine cancers reveals determinants of tumor nutrient availability. *Elife* **8**, (2019).

100. Garcia-Bermudez, J. *et al.* Aspartate is a limiting metabolite for cancer cell proliferation under hypoxia and in tumours. *Nat. Cell Biol.* **20**, 775–781 (2018).
101. Cantor, J. R. *et al.* Physiologic Medium Rewires Cellular Metabolism and Reveals Uric Acid as an Endogenous Inhibitor of UMP Synthase. *Cell* **169**, 258-272 e17 (2017).
102. Cantor, J. R. *et al.* Physiologic Medium Rewires Cellular Metabolism and Reveals Uric Acid as an Endogenous Inhibitor of UMP Synthase. *Cell* **169**, 258-272.e17 (2017).
103. Vande Voorde, J. *et al.* Improving the metabolic fidelity of cancer models with a physiological cell culture medium. *Sci. Adv.* **5**, eaau7314 (2019).
104. Onodera, Y., Nam, J.-M. & Bissell, M. J. Increased sugar uptake promotes oncogenesis via EPAC/RAP1 and O-GlcNAc pathways. *J. Clin. Invest.* **124**, 367–384 (2014).
105. Sonveaux, P. *et al.* Targeting lactate-fueled respiration selectively kills hypoxic tumor cells in mice. *J. Clin. Invest.* **118**, 3930–3942 (2008).
106. Allen, E. *et al.* Metabolic Symbiosis Enables Adaptive Resistance to Anti-angiogenic Therapy that Is Dependent on mTOR Signaling. *Cell Rep.* **15**, 1144–1160 (2016).
107. Fennema, E., Rivron, N., Rouwkema, J., van Blitterswijk, C. & de Boer, J. Spheroid culture as a tool for creating 3D complex tissues. *Trends Biotechnol* **31**, 108–115 (2013).
108. Takasato, M. *et al.* Kidney organoids from human iPS cells contain multiple lineages and model human nephrogenesis. *Nature* **526**, 564–568 (2015).
109. de Souza, N. Organoids. *Nat. Methods* **15**, 23 (2018).
110. Rios, A. C. & Clevers, H. Imaging organoids: a bright future ahead. *Nat. Methods* **15**, 24–26 (2018).
111. Freedman, B. S. *et al.* Modelling kidney disease with CRISPR-mutant kidney organoids derived from human pluripotent epiblast spheroids. *Nat Commun* **6**, 8715 (2015).
112. Lu, Y. *et al.* A novel 3D liver organoid system for elucidation of hepatic glucose metabolism. *Biotechnol Bioeng* **109**, 595–604 (2012).
113. Lukovac, S. *et al.* Differential modulation by Akkermansia muciniphila and Faecalibacterium prausnitzii of host peripheral lipid metabolism and histone acetylation in mouse gut organoids. *MBio* **5**, (2014).
114. Xu, Q. *et al.* Renal carcinoma/kidney progenitor cell chimera organoid as a novel tumourigenesis gene discovery model. *Dis Model Mech* (2017). doi:10.1242/dmm.028332
115. Rodenhizer, D. *et al.* A three-dimensional engineered tumour for spatial snapshot analysis of cell metabolism and phenotype in hypoxic gradients. *Nat. Mater.* **15**, 227 (2015).
116. Rodenhizer, D., Dean, T., Xu, B., Cojocari, D. & McGuigan, A. P. A three-dimensional engineered heterogeneous tumor model for assessing cellular environment and response. *Nat. Protoc.* **13**, 1917–1957 (2018).
117. Ghelli, A. *et al.* The cytochrome b p.278Y>C mutation causative of a multisystem disorder enhances superoxide production and alters supramolecular interactions of respiratory chain complexes. *Hum. Mol. Genet.* **22**, 2141–2151 (2013).
118. Cheng, Q.-Y. *et al.* Sensitive Determination of Onco-metabolites of D- and L-2-hydroxyglutarate Enantiomers by Chiral Derivatization Combined with Liquid Chromatography/Mass Spectrometry Analysis. *Sci. Rep.* **5**, 15217 (2015).
119. Su, X., Lu, W. & Rabinowitz, J. D. Metabolite Spectral Accuracy on Orbitraps. *Anal. Chem.* **89**, 5940–5948 (2017).
120. Gaude, E. *et al.* muma, An R Package for Metabolomics Univariate and Multivariate Statistical Analysis. *Current Metabolomics* **1**, 180–189 (2013).
121. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and Go Extraction Tips for Matrix-Assisted Laser Desorption/Ionization, Nanoelectrospray, and LC/MS Sample Pretreatment in Proteomics. *Anal. Chem.* **75**, 663–670 (2003).
122. Aftab, W. wasimaftab/LIMMA-pipeline-proteomics: LIMMA-pipeline-proteomics. (2020). doi:10.5281/ZENODO.3731967
123. Sergushichev, A. A. An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *bioRxiv* 60012 (2016). doi:10.1101/060012
124. Morgan M, Falcon S, G. R. Gene set enrichment data structures and methods. *R Packag. version 1.50.0* doi:10.18129/B9.bioc.GSEABase
125. Taube, J. H. *et al.* Core epithelial-to-mesenchymal transition interactome gene-expression

signature is associated with claudin-low and metaplastic breast cancer subtypes. *Proc. Natl. Acad. Sci.* **107**, 15449 LP – 15454 (2010).
126. Blighe K, Rana S, L. M. (2020). EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. *R Packag. version 1.6.0* doi:10.18129/B9.bioc.EnhancedVolcano
127. Papachristou, E. K. *et al.* A quantitative mass spectrometry-based approach to monitor the dynamics of endogenous chromatin-associated protein complexes. *Nat. Commun.* **9**, 2311 (2018).
128. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).
129. Kulkarni, R. A. *et al.* A chemoproteomic portrait of the oncometabolite fumarate. *Nat. Chem. Biol.* (2019). doi:10.1038/s41589-018-0217-y
130. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
131. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10 (2011).
132. 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).
133. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
134. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
135. 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).
136. Mora, A. scibiomart. (2020). doi:https://doi.org/10.5281/zenodo.4099048
137. Kovaka, S. *et al.* Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* **20**, 278 (2019).
138. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
139. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**, 1650–1667 (2016).
140. Vitting-Seerup, K. & Sandelin, A. The Landscape of Isoform Switches in Human Cancers. *Mol. Cancer Res.* **15**, 1206–1220 (2017).
141. Vitting-Seerup, K. & Sandelin, A. IsoformSwitchAnalyzeR: analysis of changes in genome-wide patterns of alternative splicing and its functional consequences. *Bioinformatics* **35**, 4469–4471 (2019).
142. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research* **4**, 1521 (2016).
143. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47–e47 (2015).
144. Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-seq data. *Genome Res.* **22**, 2008–2017 (2012).
145. Chen, D.-P., Lin, Y.-C. & Fann, C. S. J. Methods for identifying differentially methylated regions for sequence- and array-based data: Table 1. *Brief. Funct. Genomics* elw018 (2016). doi:10.1093/bfgp/elw018
146. Park, Y., Figueroa, M. E., Rozek, L. S. & Sartor, M. A. MethylSig: a whole genome DNA methylation analysis pipeline. *Bioinformatics* **30**, 2414–2422 (2014).
147. Korthauer, K., Chakraborty, S., Benjamini, Y. & Irizarry, R. A. Detection and accurate false discovery rate control of differentially methylated regions from whole genome bisulfite sequencing. *Biostatistics* **20**, 367–383 (2019).
148. Hansen, K. D., Langmead, B. & Irizarry, R. A. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol.* **13**, R83 (2012).
149. Akalin, A. *et al.* methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* **13**, R87 (2012).
150. Mora, A. sciepi2gene. (2020). doi:https://doi.org/10.5281/zenodo.4294215

151. Mora, A. scidmg. (2021). doi:10.5281/zenodo.4543954
152. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *Omi. A J. Integr. Biol.* **16**, 284–287 (2012).
153. Mora, A. scimotf. (2021). doi:10.5281/zenodo.4543970
154. Dewey, M. metap: meta-analysis of significance values. (2020).
155. Yang, Y. *et al.* UOK 262 cell line, fumarate hydratase deficient (FH-/FH-) hereditary leiomyomatosis renal cell carcinoma: in vitro and in vivo model of an aberrant energy metabolic pathway in human cancer. *Cancer Genet Cytogenet* **196**, 45–55 (2010).
156. Yang, Y. *et al.* A novel fumarate hydratase-deficient HLRCC kidney cancer cell line, UOK268: a model of the Warburg effect in cancer. *Cancer Genet* **205**, 377–390 (2012).
157. Ryan, M. J. *et al.* HK-2: An immortalized proximal tubule epithelial cell line from normal adult human kidney. *Kidney Int.* **45**, 48–57 (1994).
158. Buescher, J. M. *et al.* A roadmap for interpreting 13C metabolite labeling patterns from cells. *Curr. Opin. Biotechnol.* **34**, 189–201 (2015).
159. Matthew, B., R., T. S. & W., B. J. Succination of Proteins by Fumarate. *Ann. N. Y. Acad. Sci.* **1126**, 272–275 (2008).
160. Malhotra, D. *et al.* Global mapping of binding sites for Nrf2 identifies novel targets in cell survival response through ChIP-Seq profiling and network analysis. *Nucleic Acids Res.* **38**, 5718–5734 (2010).
161. de la Cruz López, K. G., Toledo Guzmán, M. E., Sánchez, E. O. & García Carrancá, A. mTORC1 as a Regulator of Mitochondrial Functions and a Therapeutic Target in Cancer. *Front. Oncol.* **9**, (2019).
162. Geuens, T., Bouhy, D. & Timmerman, V. The hnRNP family: insights into their role in health and disease. *Hum. Genet.* **135**, 851–867 (2016).
163. Szalai, B. & Saez-Rodriguez, J. Why do pathway methods work better than they should? *FEBS Lett.* **594**, 4189–4200 (2020).
164. Yoshimura, S. H., Otsuka, S., Kumeta, M., Taga, M. & Takeyasu, K. Intermolecular disulfide bonds between nucleoporins regulate karyopherin-dependent nuclear transport. *J. Cell Sci.* **126**, 3141 LP – 3150 (2013).
165. Zhang, W. *et al.* Redox-Sensitive Cysteines Confer Proximal Control of the Molecular Crowding Barrier in the Nuclear Pore. *Cell Rep.* **33**, 108484 (2020).
166. Duan, Y. & Zhang, D. Identification of novel prognostic alternative splicing signature in papillary renal cell carcinoma. *J. Cell. Biochem.* **121**, 672–689 (2020).
167. Li, H. *et al.* Pan-cancer analysis of alternative splicing regulator heterogeneous nuclear ribonucleoproteins (hnRNPs) family and their prognostic potential. *J. Cell. Mol. Med.* **24**, 11111–11119 (2020).
168. Siegfried, Z., Bonomi, S., Ghigna, C. & Karni, R. Regulation of the Ras-MAPK and PI3K-mTOR Signalling Pathways by Alternative Splicing in Cancer. *Int. J. Cell Biol.* **2013**, 1–9 (2013).
169. Panasyuk, G. *et al.* mTORβ Splicing Isoform Promotes Cell Proliferation and Tumorigenesis. *J. Biol. Chem.* **284**, 30807–30814 (2009).
170. Rodenhizer, D. *et al.* A three-dimensional engineered tumour for spatial snapshot analysis of cell metabolism and phenotype in hypoxic gradients. *Nat. Mater.* **15**, 227 (2015).
171. Ye, D., Guan, K.-L. & Xiong, Y. Metabolism, Activity, and Targeting of D- and L-2-Hydroxyglutarates. *Trends in Cancer* **4**, 151–165 (2018).
172. Frezza, C. *et al.* Metabolic Profiling of Hypoxic Cells Revealed a Catabolic Signature Required for Cell Survival. *PLoS One* **6**, e24411 (2011).
173. Zhang, H. *et al.* Mitochondrial Autophagy Is an HIF-1-dependent Adaptive Metabolic Response to Hypoxia * . *J. Biol. Chem.* **283**, 10892–10903 (2008).
174. Rossi, M. *et al.* Heterogeneity in PHGDH protein expression potentiates cancer cell dissemination and metastasis. *bioRxiv* 2021.01.24.427949 (2021). doi:10.1101/2021.01.24.427949
175. Amelio, I., Cutruzzola, F., Antonov, A., Agostini, M. & Melino, G. Serine and glycine metabolism in cancer. *Trends Biochem Sci* **39**, 191–198 (2014).
176. Pacold, M. E. *et al.* A PHGDH inhibitor reveals coordination of serine synthesis and one-carbon

unit fate. *Nat. Chem. Biol.* **12**, 452–458 (2016).