

Word count:

Student number:

Promoters

Supervisors:

A dissertation submitted to Ghent University in partial fulfilment of the requirements for the degree of
Master of

Academic Year:

Preface

Contents

Preface	v
1 Abstract	1
2 Introduction	3
2.1 The Occurrence of Ferroptosis	3
2.2 The Prospect of Ferroptosis in Cancer Therapy	7
2.3 DNA Methylation and its Role in Cancer	8
2.4 DNA Methylation Detection Techniques	10
2.5 Predictive Modeling in Clinical Bioinformatics	13
3 Aims	15
4 Methods	17
4.1 Samples	17
4.2 Gene Whitelist	17
4.3 RNA-seq	18
4.4 ONT-seq	18
4.5 Differential Gene Expression Analysis	18
4.6 Feature Engineering	18
4.7 Feature Reduction	18

4.8	Modelling	19
4.9	Post-Analysis	19
4.10	Sensitization	19
5	Results	21
5.1	Differential Gene Expression Analysis	21
5.2	Feature Engineering	23
5.3	Feature Reduction	23
5.4	Modelling	23
5.5	Post-Analysis	23
5.6	Sensitization	23
6	Discussion	25
A	Appendix	27

List of Figures

2.1	Lipid Peroxidation Mechanism. (1) Initiation begins when R (an reactive oxygen species (ROS)) abstracts a hydrogen from the highlighted region. The carbon-centered lipid radical is stabilized by molecular rearrangement, forming a conjugated diene. (2) In propagation, lipid radicals rapidly react with oxygen to form a lipid peroxy radical. (3) Propagation continues with the lipid peroxy radical abstracting a hydrogen from another lipid molecule. This generates another unsaturated lipid radical as well as a lipid hydroperoxide. (4) Termination of the radical reaction occurs when antioxidants donate a hydrogen atom to the lipid peroxy radical species, forming nonradical products. However, termination is not possible in the absence of radical species. Therefore, propagation will continue until there is a radical sink. Figure created using BioRender and redrawn from Ayala, Muñoz, and Argüelles, 2014	5
2.2	Overview of the various regulatory pathways associated with ferroptosis (C. Zhang et al., 2022a).	6
2.3	Direct base modification detection with Oxford Nanopore Technology (ONT).	12
3.1	Overview workflow.	15
5.1	Volcano Plot	21
5.2	MA-Plot	22

5.3	PCA	22
5.4	Pval dist	23

List of Tables

4.1	Cell Line Metadata	17
-----	------------------------------	----

List of Abbreviations and Symbols

5caC 5-carboxylcytosine	9
5fC 5-formylcytosine	9
5hmC 5-hydroxymethylcytosine	9
5mC 5-methylcytosine	9
CGI CpG island	9
CoQ10 ubiquinone	6
CoQ10H2 ubiquinol	6
CpG cytosine-guanine	8
DNMT DNA methyltransferase	8
FSP1 Ferroptosis Suppressor Protein 1	5
gDNA genomic DNA	18
GPX4 glutathione peroxidase 4	4
GSH glutathione	4
I-Cys2 I-cystine	5
I-CysH I-cysteine	4
I-Glu I-glutamate	5
ONT Oxford Nanopore Technology	ix
PCR poly-chain reaction	10

PDT	photodynamic therapy	8
PUFA	poly-unsaturated fatty acid	4
RCD	regulated cell death	3
ROS	reactive oxygen species	ix
SAM	S-adenosyl-L-methionine	8
SDT	sonodynamic therapy	8
SxC-	System xc-	5
TET	ten-eleven translocation	9
TFR1	transferrin receptor 1	3
TDG	thymine DNA glycosylase	9

Chapter 1

Abstract

Cancer cells exhibit a characteristic ability to evade programmed cell death, a feature critical for their uncontrolled proliferation. An emerging strategy in cancer therapy involves targeting alternative forms of cell death, distinct from apoptosis, as evidenced by recent studies (Tait, Ichim, and Green, 2014). Non-apoptotic cell death mechanisms, including necroptosis and pyroptosis, offer promising avenues to circumvent apoptosis resistance in tumors, thus enhancing the efficacy of anti-cancer therapies (J. Ye et al., 2018).

One such non-apoptotic pathway, ferroptosis, reliant on iron and ROS, represents a novel approach in cancer treatment. However, inherent resistance to ferroptosis induction poses a challenge in its therapeutic application (Zou and Schreiber, 2020). Detecting drug resistance before treatment initiation could significantly impact treatment outcomes (Volm and Efferth, 2015).

Epigenetic modifications, such as DNA methylation, RNA methylation, and non-coding RNAs, have emerged as key regulators of cellular processes. Notably, DNA methylation alterations in genes associated with ferroptosis, including GPX4, FSP1, and NRF2, influence their expression levels, potentially modulating ferroptosis susceptibility (X. Wang et al., 2023).

Bisulfite genomic sequencing is widely utilized for DNA methylation detection, despite its limitations. Challenges such as incomplete bisulfite conversion or assay failures may compromise data accuracy (Y. Li and Tollefsbol, 2011; Heiss and Just,

2018).

Recent advancements in sequencing technologies, such as nanopore sequencing, offer improved accuracy and resolution compared to traditional methods. Nanopore episequencing, in particular, shows promise in refining genetic associations and identifying novel loci previously undetected by microarray platforms (Gombert et al., 2023; Flynn et al., 2022).

In this project, we will aim to develop a machine learning model which, based on a nanopore episequencing profile, will be able to predict the likelihood of ferroptosis resistance for a given tumor strain. The model's capacity to generalize to other (un-trained) strains will be tested. Lastly, we will identify specific methylation changes brought about by ferroptosis-inducing drugs by sequencing sensitized strains.

Chapter 2

Introduction

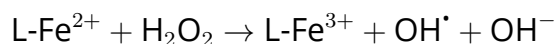
2.1 The Occurrence of Ferroptosis

The exploration of ferroptosis has unveiled a multitude of cellular factors and potential regulatory pathways. At its core, it is defined as an intracellular, non-apoptotic, iron-dependent form of regulated cell death (RCD) that is distinct from apoptosis, necrosis, and autophagy (Yan et al., 2021). As the name suggests, this phenomenon is principally driven by iron molecules absorbed by cells from the surrounding extracellular environment. One such receptor is transferrin receptor 1 (TFR1), which mediates endocytosis of diferric transferrin (S. Wang et al., 2019). This intracellular iron is often transported and stored in storage proteins such as ferritin, aiding in maintaining iron homeostasis and metabolism (Knovich et al., 2009).

Labile iron, however, which refers to loosely bound or chelatable iron, becomes a critical player in the ferroptotic process. An important aspect of iron's function is its ability to undergo cyclic oxidation and reduction, participating in a wide array of metabolic processes, playing crucial roles in oxygen transport, DNA synthesis, and electron transport (Lieu et al., 2001). However, this redox activity can generate free radicals and other strongly oxidizing species, capable of causing a wide range of biological injury. Among these unwanted activities, labile iron participates in what's known as the Fenton reaction.

The Fenton reaction, originating from observations by Fenton in the 1880s, involves

iron and hydrogen peroxide to generate a strong oxidant. It has been extensively studied as a source of hydroxyl radicals and an initiator of biological damage (Im-lay, 2003). In its simplest form, when iron is complexed to ligand L, it can be represented as:



As a result of this reaction, excessive concentrations of labile iron leads to the generation of ROS, particularly hydroxyl radicals (Cabantchik, 2014). ROS, in turn, contribute to oxidative stress within the cell, resulting in oxidative damage to vital macromolecules such as DNA, proteins, and lipids (L. A. Pham-Huy, H. He, and C. Pham-Huy, 2008).

Poly-unsaturated fatty acids (PUFAs), core components of cellular membranes, are highly susceptible to oxidative damage by ROS (Ayala, Muñoz, and Argüelles, 2014). Lipid peroxides are formed as a result of the oxidation of these fatty acids. Lipid peroxidation undergoes three distinct phases: initiation, propagation, and termination (Yin, Xu, and Porter, 2011). Once initiated, lipid peroxidation sets off a cascade of chain reactions that persist until termination products are generated, which only occurs upon the intervention of an antioxidant molecule. A summary of the lipid peroxidation mechanism can be found on Fig. 2.1.

Accumulation of lipid peroxides disrupts the integrity of cellular membranes, particularly the lipid bilayer (Gaschler and Stockwell, 2017). Membrane damage leads to the loss of compartmentalization and function within the cell, including the mitochondrion (Ademowo et al., 2017). Eventually, the cell membrane will rupture, resulting in cell death.

Healthy mammalian cells stave off this threat in a number of ways. The primary cellular mechanism of protection against oxytosis/ferroptosis is mediated by glutathione peroxidase 4 (GPX4), a glutathione-dependent hydroperoxidase that converts lipid peroxides into non-toxic lipid alcohols (Xie et al., 2023). glutathione (GSH) itself, which is produced from l-cysteine (l-CysH), is able to neutralize ROS present within the cell directly, playing the role of antioxidant (Averill-Bates, 2023). For an overview

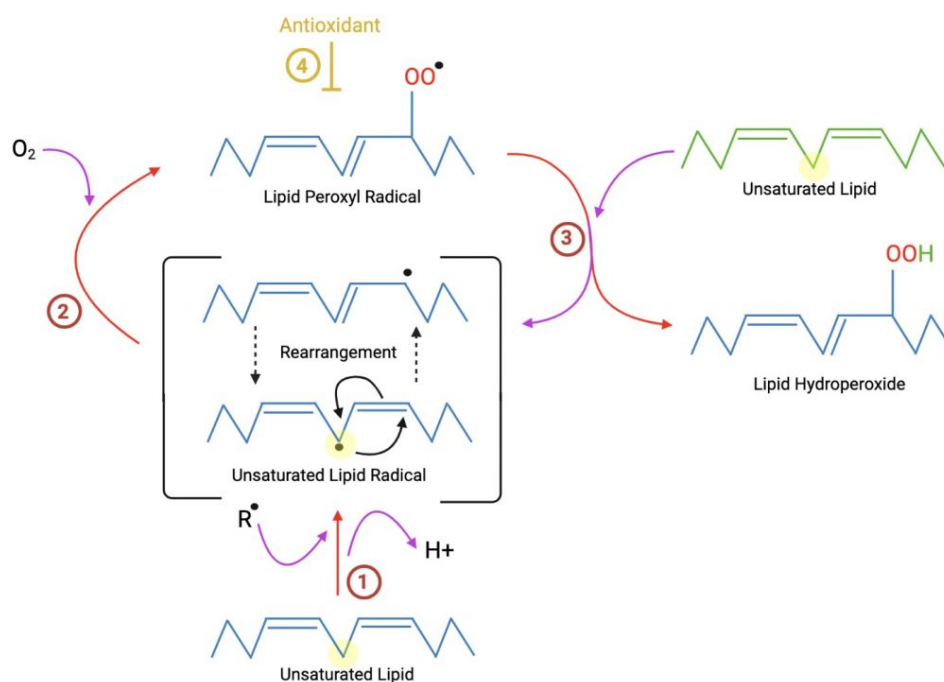


Figure 2.1: Lipid Peroxidation Mechanism. (1) Initiation begins when R (an ROS) abstracts a hydrogen from the highlighted region. The carbon-centered lipid radical is stabilized by molecular rearrangement, forming a conjugated diene. (2) In propagation, lipid radicals rapidly react with oxygen to form a lipid peroxyl radical. (3) Propagation continues with the lipid peroxyl radical abstracting a hydrogen from another lipid molecule. This generates another unsaturated lipid radical as well as a lipid hydroperoxide. (4) Termination of the radical reaction occurs when antioxidants donate a hydrogen atom to the lipid peroxyl radical species, forming nonradical products. However, termination is not possible in the absence of radical species. Therefore, propagation will continue until there is a radical sink. Figure created using BioRender and redrawn from Ayala, Muñoz, and Argüelles, 2014

of the various co-dependent pathways and effects classically associated with ferroptosis, see Fig. 2.2.

System xc⁻ (Sx⁻C⁻), an amino acid antiporter, primarily facilitates the exchange of extracellular l-cystine (l-Cys₂) and intracellular l-glutamate (l-Glu) across the cellular plasma membrane. Studies initially emphasized the transport of l-Cys₂ influx, as in numerous cells, this process acts as a rate-limiting step for supplying intracellular l-CysH essential for GSH synthesis (Bridges, Natale, and Patel, 2012).

The Xc⁻/GSH/GPX4 axis functions as a GSH-dependent defense mechanism against ferroptosis, representing the most crucial and well-studied antioxidant system for resisting this form of cell death (Y. Ye et al., 2022). However, in certain cell types or lines, inhibiting this system alone isn't sufficient to trigger ferroptosis, suggesting the existence of alternative pathways (Doll, Freitas, et al., 2019). One such mechanism involves the GSH-independent enzyme Ferroptosis Suppressor Protein

1 (FSP1), which operates alongside GPX4, serving as a key regulator of ferroptosis (M. Yang et al., 2022). Positioned at the plasma membrane with myristoylation, FSP1 utilizes NADH/NADPH as a cofactor to reduce ubiquinol (CoQ10) to its active form, ubiquinol (CoQ10H₂). While CoQ10 mainly resides within the inner mitochondrial membrane, CoQ10H₂ acts as a lipophilic antioxidant, scavenging free radicals to prevent lipid peroxidation.

Beyond these key proteins and enzymes, it has been proven that the mitochondrion plays a substantial role in ferroptosis regulation and induction, being the major compartment for cellular iron metabolism (Liu et al., 2023). Ferroptosis is characterized by significant morphological alterations in mitochondria such as fragmentation and cristae enlargement (Doll, Proneth, et al., 2017), is closely linked with cellular mechanisms. Interestingly, certain potent ferroptosis inhibitors demonstrate a remarkable specificity towards mitochondria (Krainz et al., 2016). Mitochondrial shrinkage and rupture are also common terminal observations when a cell undergoes ferroptosis, since this organelle is enveloped in a membrane (Gao et al., 2019).

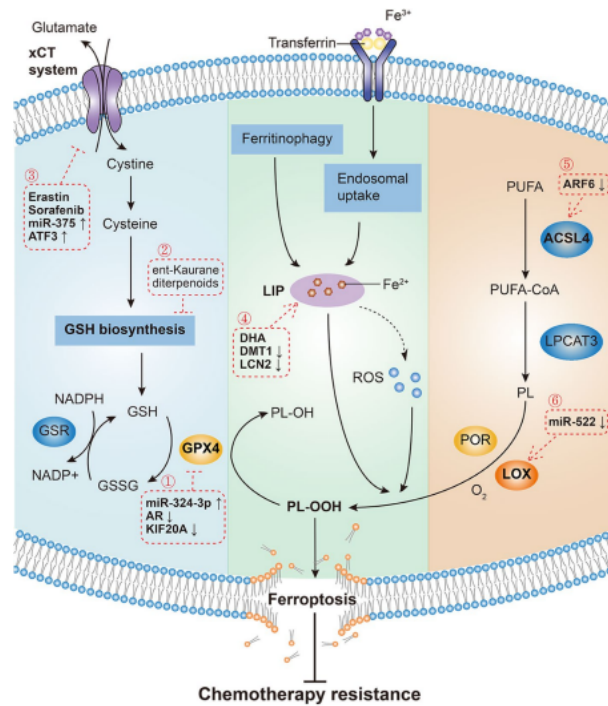


Figure 2.2: Overview of the various regulatory pathways associated with ferroptosis (C. Zhang et al., 2022a).

2.2 The Prospect of Ferroptosis in Cancer Therapy

Cancer is the second leading cause of death worldwide, behind cardiovascular disease (H. Wang et al., 2016). Ongoing global research efforts have led to notable enhancements in cancer survival rates over the years. Despite these significant advancements in cancer treatment, including surgery, radiation therapy, chemotherapy, hormone therapy, targeted therapy, and immunotherapy (Debela et al., 2021), resistance to these therapies remains a prevalent challenge in clinical practice (Bukowski, Kciuk, and Kontek, 2020).

The phenomenon of drug resistance, where diseases develop tolerance to pharmaceutical treatments, initially emerged with antibiotic-resistant bacteria. However, this concept extends beyond bacteria to encompass various diseases, including cancer. While some resistance mechanisms are disease-specific, others, like drug efflux, occur across microbes and drug-resistant cancers, indicating evolutionary conservation. Despite initial susceptibility to chemotherapy, many cancers eventually develop resistance through mechanisms such as DNA mutations and metabolic changes, which hinder drug inhibition and promote degradation (Housman et al., 2014).

The prevalence of acquired resistance to targeted therapy underscores the urgent need for effective strategies to combat drug resistance. Recent findings highlight the significant role of ferroptosis in cancer therapy, disrupting target molecules and influencing cancer progression. Furthermore, studies have indicated that leveraging ferroptosis may offer a promising avenue to overcome resistance to targeted therapy (C. Zhang et al., 2022b). Typically in cancer, the apoptotic pathway faces inhibition through various mechanisms, such as the overexpression of anti-apoptotic proteins and the under-expression of pro-apoptotic proteins. These alterations often lead to intrinsic resistance to chemotherapy, a prevalent anticancer therapy. It is primarily due to ferroptosis' profile as a non-apoptotic form of RCD that makes it an incredibly promising avenue for new therapy strategies, allowing it to circumvent many tumors' natural resilience to apoptotic induction.

In the field, ferroptosis is indeed commonly acknowledged as a tumor suppressor

mechanism. Research has demonstrated that the inactivation of the p53 tumor suppression pathway, implicated in the genesis of most human cancers, is correlated with the natural suppression of ferroptosis (Jiang et al., 2015). Activation of p53 obstructs cystine uptake via the cystine/glutamate antiporter, thereby reducing intracellular GSH levels and shielding tumor cells from ferroptosis.

The synthesis and validation of ferroptosis inducers through various *in vivo* experiments underscore their potential as a viable anticancer therapy for clinical use pending extensive population validation. Consequently, numerous investigations have delved into the development of diverse inducers to trigger ferroptosis, examining its efficacy in various cancer treatment modalities including chemotherapy, immunotherapy, radiotherapy, nanotherapy, sonodynamic therapy (SDT), and photodynamic therapy (PDT).

Tumor drug resistance arises from a multitude of mechanisms, with the disruption of redox homeostasis emerging as a pivotal factor. Tumor cells bolster their resilience against oxidative stress by curtailing their own production of ROS, fostering the development of acquired drug resistance (C. Yang et al., 2015). Being able to predict which tumors are more inclined than others to respond well to ferroptosis induction would aid significantly in lowering the bar for targeted research and treatment.

2.3 DNA Methylation and its Role in Cancer

In mammalian cells, the process of DNA methylation predominantly involves the addition of a methyl group to the carbon-5 atom of cytosine within cytosine-guanine (CpG) dinucleotides. This biochemical reaction is facilitated by three main enzymes known as DNA methyltransferases (DNMTs), including DNMT1, DNMT3A, DNMT3B (Jin and Robertson, 2013). Essential for mammalian development, these enzymes facilitate the transfer of a methyl group from the universal methyl donor, S-adenosyl-L-methionine (SAM), to the 5-position of cytosine residues in DNA (Robertson, 2005). Following DNA replication, the newly synthesized DNA strand lacks methylation. DNMT1 predominantly methylates the unmethylated strand of hemimethylated

DNA, with DNMT3a/3b assisting in ensuring efficient propagation of DNA methylation, especially at sites overlooked by DNMT1 (Jin and Robertson, 2013).

While DNMTs are responsible for adding methyl groups, the ten-eleven translocation (ten-eleven translocation (TET)) family of dioxygenases has emerged as key players in DNA demethylation. Through successive enzymatic processes, TET enzymes can oxidize 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), and further to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Pastor, Aravind, and Rao, 2013). This oxidation of 5mC contributes to the gradual loss of DNA methylation during cell replication. Additionally, the oxidized intermediates can be converted back to cytosine through iterative oxidation followed by base excision repair facilitated by thymine DNA glycosylase (TDG) (Kohli and Y. Zhang, 2013). Together with DNMTs, these enzymes orchestrate the dynamic regulation of DNA methylation.

Methylated cytosine residues are prone to spontaneous deamination, leading to the conversion of cytosine to thymine, a transition that is often poorly repaired. Consequently, a significant portion of familial mutations and single-nucleotide polymorphisms associated with diseases are found within methylated CpG sites. Moreover, CpG residues within gene bodies or coding regions in somatic cells frequently serve as mutational hotspots, exemplified by inactivating C to T transitions in the tumor suppressor gene p53 (Peter A Jones and Baylin, 2002).

This phenomenon also results in a diminished presence of the CpG palindrome throughout the human genome, except within regions identified as CpG islands (CGIs). CGIs, originally characterized by Gardiner-Garden and Frommer, 1987, are defined as 200-bp DNA segments with a 50% C + G content and an observed CpG frequency exceeding the expected value of 0.6. While most CpGs undergo methylation, those within CGIs typically remain unmethylated (Meissner et al., 2008). CGIs are predominantly located within approximately half of all gene promoters (Mikkelsen et al., 2007). Conversely, non-CGI promoters are usually methylated and transcriptionally silent. These genes are more likely to exhibit tissue-specific expression, resulting in only a fraction of non-CGI promoters remaining unmethylated and accessible to transcription factors in each tissue type (Eckhardt et al.,

2006).

Recent advancements in next-generation sequencing techniques applied to cancer genomes have unveiled a prevalent pattern of mutations recurring in diverse epigenetic modulators. These modulators encompass mediators of DNA methylation, covalent histone modifiers, and genes encoding subunits of chromatin remodelers (You and Peter A. Jones, 2012). The aberrant functioning of these pivotal epigenetic regulators leads to the dysregulation of gene expression and is implicated in a myriad of malignancies, spanning various cancer types (Hanahan and Weinberg, 2011).

2.4 DNA Methylation Detection Techniques

Bisulfite genomic sequencing stands as the benchmark technology for detecting DNA methylation due to its ability to offer qualitative, quantitative, and highly efficient insights into identifying 5-methylcytosine at single base-pair resolution. Initially pioneered by Frommer et al., 1992, this method capitalizes on the discovery that the amination reactions of cytosine and 5mC yield markedly distinct outcomes following treatment with sodium bisulfite.

In this context, cytosines present in single-stranded DNA undergo conversion into uracil residues, which are subsequently recognized as thymine during poly-chain reaction (PCR) amplification and sequencing. However, 5mCs remain unaffected by this conversion, retaining their cytosine identity and enabling the distinction between methylated and unmethylated cytosines (Y. Li and Tollefsbol, 2011).

Following bisulfite treatment, a subsequent PCR step becomes imperative to ascertain the methylation status at specific loci, employing methylation-specific primers. The actual methylation status can then be determined through either direct sequencing of PCR products (for the detection of average methylation status) or sub-cloning sequencing (for the detection of the distribution of methylation patterns at the single-molecule level) (Y. Li and Tollefsbol, 2011).

Despite being the current gold standard, bisulfite sequencing is not without its flaws. Understanding bisulfite-treated DNA data requires careful consideration

of two types of conversion errors: failed conversion and inappropriate conversion. Failed conversion, extensively studied, arises when an unmethylated cytosine fails to deaminate, resulting in its appearance as methylated in sequencing data. In mammalian somatic cells, where 5-methylcytosine primarily occurs at CpG cytosines, the frequency of failed conversion is indicated by the proportion of non-CpG cytosines appearing as cytosines in sequence data (Bird, 1992). Ignoring failed conversion in data analysis can inflate methylation density estimates and impede the determination of DNA methyltransferase sequence motif preferences. Inappropriate conversion, the second type of error, occurs when a methylated cytosine undergoes deamination, yielding thymine (Shiraishi and Hayatsu, 2004). Similar to uracils resulting from cytosine deamination, thymine pairs with adenine during PCR. Consequently, 5-methylcytosines subjected to inappropriate conversion are misinterpreted as unmethylated. Failure to account for inappropriate conversion leads to underestimation of genomic methylation densities. However, if the frequency of inappropriate conversion is known, it can be integrated as a parameter in data analysis. Therefore, information on both failed and inappropriate conversion frequencies is crucial for drawing meaningful insights from detailed DNA methylation patterns.

In addressing the limitations of bisulfite sequencing, new techniques have emerged despite its status as the gold standard for detecting methylation marks. These limitations encompass the aforementioned confounding effects of bisulfite treatment on DNA methylation, alongside the challenge of distinguishing between DNA methylation and hydroxylation (López-Catalina et al., 2024). In light of these considerations, we advocate for the utilization of Nanopore sequencing as a promising alternative for identifying methylation marks.

Nanopore sequencing is a next-generation sequencing technology that enables the real-time analysis of DNA or RNA molecules as they pass through a nanoscale pore. This technique is based on the principle that as a DNA strand translocates through a nanopore, changes in the electrical current can be measured and used to deduce the sequence of the DNA molecule Y. Wang et al., 2021.

The raw electrical signal is translated into DNA sequence information through a

process called basecalling. DNA methylation involves the addition of a methyl group to a cytosine base, commonly referred to as a 5-methylcytosine. Nanopore sequencing can detect these modifications as changes in the electrical current during translocation using specialized algorithms and basecallers (Fig. 2.3).

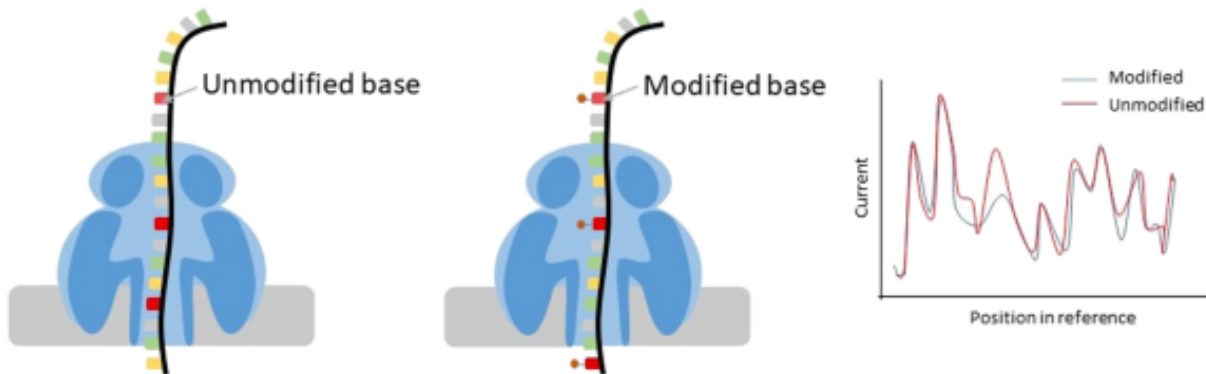


Figure 2.3: Direct base modification detection with ONT.

One of the key advantages of nanopore sequencing is its ability to generate long reads, providing information about entire genomic regions in a single continuous sequence. Short-read sequencing offers a cost-effective and accurate solution, bolstered by a diverse array of analysis tools and pipelines (Heather and Chain, 2016). However, the vast range of lengths spanning natural nucleic acid polymers, extending over eight orders of magnitude, presents a challenge when sequencing them in short amplified fragments, hindering the reconstruction and quantification of the original molecules.

Long reads, on the other hand, hold the potential to enhance de novo assembly, ensure mapping accuracy, facilitate transcript isoform identification, and enable the detection of structural variants. Moreover, the sequencing of native molecules, both DNA and RNA, using long reads, eliminates amplification bias while retaining base modifications (Depledge et al., 2019). These capabilities, coupled with ongoing advancements in accuracy, throughput, and cost reduction, are establishing long-read sequencing as a viable option for a wide spectrum of genomic applications in both model and non-model organisms (Yuan et al., 2017).

Adaptive sampling is a feature that helps optimize the sequencing process by adjusting data collection parameters in real-time based on the characteristics of the DNA being sequenced. The system monitors various parameters such as the qual-

ity of the signal, the speed at which the DNA is translocating through the nanopore, and other relevant metrics (Martin et al., 2022). The software requires a user to upload a file of whitelisted reference sequences, and the system can be set to either deplete or enrich for these on a specified set of channels.

In order to achieve this, the software basecalls the first few hundred bases of each read and compares it with the target reference sequences. Matching or unmatching sequences are rejected, depending on whether the software is set to enrich or deplete. By using this, coverage of our genes of interest is enriched, greatly reducing noise.

2.5 Predictive Modeling in Clinical Bioinformatics

Clinicians recognize the significance of accessing patients' genetic information across genomes, transcriptomes, proteomes, and metabolomes for enhancing disease diagnostics and prognostics (Kholodenko, Yaffe, and Kolch, 2012a). This approach, often termed precision medicine, holds the promise of enabling more personalized and effective treatments, including in emerging fields like ferroptosis research (J. He, X. Li, and Yu, 2021). Recent advancements in Next-Generation Sequencing (NGS) and Mass Spectrometry (MS) technologies have facilitated genome characterization and proteome quantification from patients' biological samples with reasonable accuracy and scalability, making them compatible with clinical point-of-care applications (McDermott, 2015). However, the complexity of data generated by these technologies surpasses human capacity for manual handling and interpretation.

In contemporary medical practice, bioinformatics serves as a crucial bridge between clinicians and 'omics' technology, offering the potential for automated data analysis and facilitating "big data" processing (Pais, Iles, and Zmuidinaite, 2021). For instance, leveraging bioinformatics tools could potentially allow for the design of an mRNA vaccine capable of targeting all circulating variants and lineages of the coronavirus (Oladipo et al., 2022), a feat challenging to achieve through exhaustive experimental methods alone.

While bioinformatics traditionally emerged from evolutionary biology and its close association with genomics, a rapidly evolving clinically focused branch, often referred to as clinical bioinformatics, is diverging from traditional bioinformatics (Belmont and Shaw, 2016). Clinical bioinformatics aims to derive diagnostics, prognostics, or therapy assessments from individual patient data, playing a central role in identifying genetic diseases, discovering novel biomarkers, characterizing pathogens, and guiding therapeutic strategies (Simon, 2011).

Predictive modeling, a longstanding specialty in data science, computational biology, and systems biology, has been instrumental for over six decades (Kholodenko, Yaffe, and Kolch, 2012b). It has proven advantageous in harnessing the full potential of high-throughput technologies for large-scale disorder screening and precision medicine. This is particularly evident in high-throughput technologies like MS and NGS, which generate vast and complex datasets requiring high computational processing power and advanced mathematical modeling approaches to translate data complexity into clinically relevant predictions. Thus, predictive modeling is poised to play a pivotal role in clinical bioinformatics, becoming an integral part of standard clinical bioinformatic pipelines as a downstream analysis step following traditional bioinformatics processes.

Chapter 3

Aims

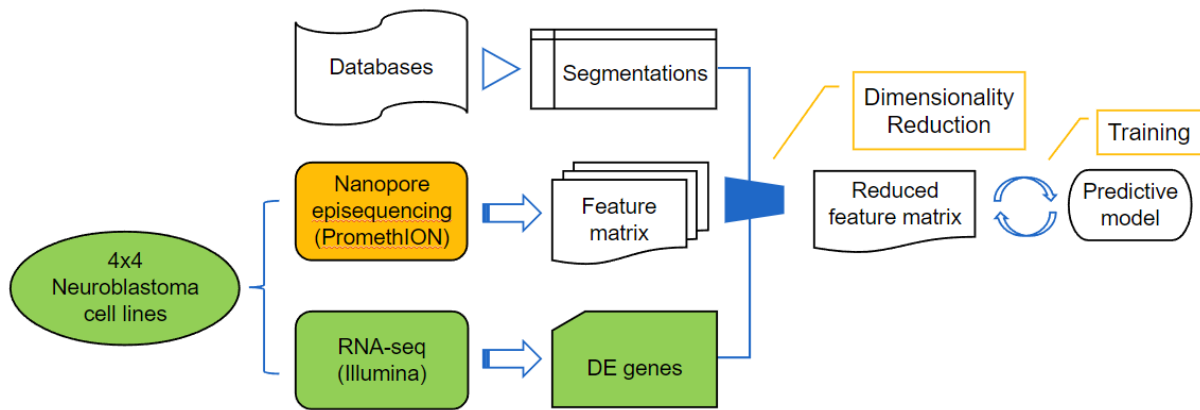


Figure 3.1: Overview workflow.

The aim of this project is, first and foremost, to develop a machine learning model capable of accurately predicting ferroptosis sensitivity for a given tumor sample. To achieve the best possible performance, we will have to compare and determine the best feature reduction and modelling methodologies.

In doing this, emphasis will be put on working annotation-agnostic whenever possible, to reduce reliance on prior knowledge and enhance confidence in the information extracted from the samples themselves. In light of these considerations, we will attempt multiple feature reduction strategies, and compare them as alternatives. Modelling performance will also be benchmarked for the various algorithms on offer.

In post-analysis, we will reverse engineer the extracted features back to annotation,

in hopes of discovering novel ferroptosis biomarkers. Finally, we will execute a comparative study of the methylation patterns between drug-sensitized tumors and wild type, non-sensitized cells from the same cell line.

Chapter 4

Methods

4.1 Samples

To start off, expression data from 4 cell lines was gathered, namely 2 resistant and 2 sensitive neuroblastoma cell lines in quadruplicate (16 samples in total). These cell lines were characterized in terms of sensitivity towards ferroptosis beforehand experimentally. Furthermore, neuroblastoma is highly iron addicted which makes it exceptionally vulnerable to ferroptosis (Floros et al., 2021). Metadata for each cell line can be found in Table 4.1.

Cell Line	Tumor Type	Ferroptosis Sensitivity	Gender	Age (months)
IMR-32	Neuroblastoma	Sensitive	Male	13
SH-EP	Neuroblastoma	Sensitive	Female	48
SH-SY5Y	Neuroblastoma	Resistant	Female	48
SK-N-BE(2)-C	Neuroblastoma	Resistant	Male	26

Table 4.1: Cell Line Metadata

4.2 Gene Whitelist

Genes thought to be associated with ferroptosis were compiled into a custom gene whitelist. The full list can be found in Appendix X.

4.3 RNA-seq

4.4 ONT-seq

For the episequencing, a PromethION nanopore sequencing device based at OHMX.bio was used, containing 24 channels per flow cell. PromethION offers a significantly higher number of channels compared to other ONT platforms like MinION and GridION. This high channel count enhances the platform's capacity for parallel sequencing.

From the whitelist, a BED file was constructed. Additionally, each gene was extended by 15kB, after which overlapping genes were joined together.

Library prep involved following protocol SQK-NBD114.24, describing how to carry out native barcoding of genomic DNA (gDNA) using the Native Barcoding Kit 24 V14. Raw signal data in pod5 format was basecalled with a SUP model and 5(h)mC calling in CG context using GPU hardware on a GridION unit installed at OHMX.bio, and then subsequently demultiplexed. The basecalling model was set to fast, with the intent of making the decision of whether to reverse the voltage of the pore as fast as possible. The resulting unmapped bam files were then aligned against a reference genome, and finally sorted and indexed using SAMtools (Danecek et al., 2021).

4.5 Differential Gene Expression Analysis

4.6 Feature Engineering

4.7 Feature Reduction

Since the episequencing returns many thousands of individual detected CpGs by locus, the number of features in our model will be staggering if used in its raw form. The first reduction of features will be actualized by bundling CpG sites into regions

(e.g. CpG islands). A second dimensionality reduction will also be performed based on drivers derived from expression data, to remove sites or regions without genes that are differentially expressed by sensitivity.

4.8 Modelling

4.9 Post-Analysis

4.10 Sensitization

Chapter 5

Results

5.1 Differential Gene Expression Analysis

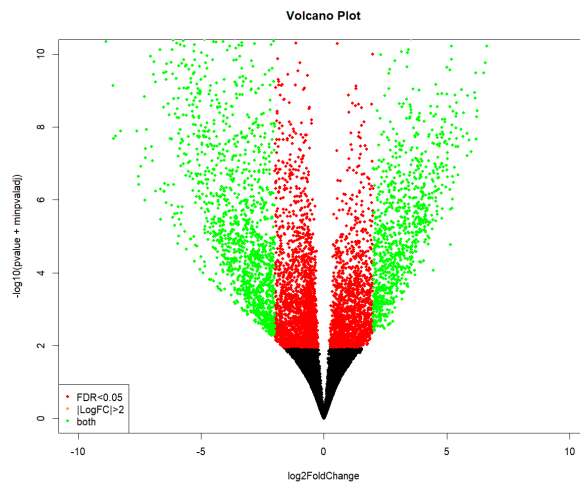


Figure 5.1: Volcano Plot

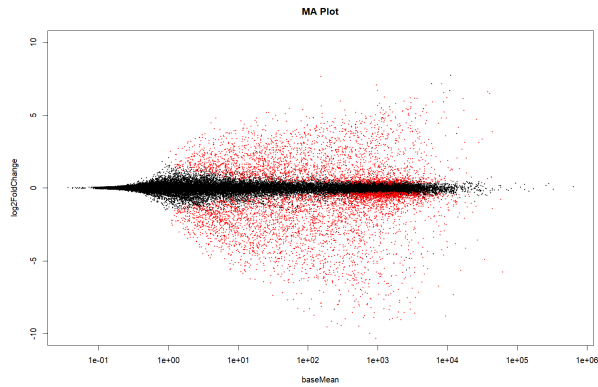


Figure 5.2: MA-Plot

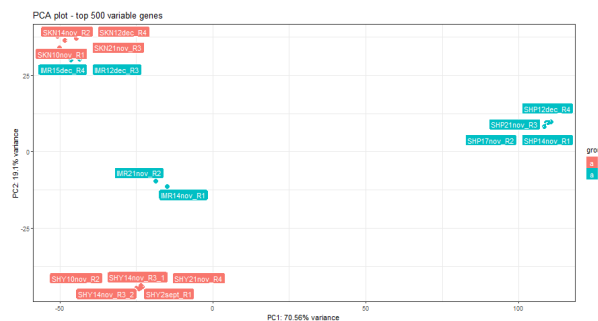


Figure 5.3: PCA

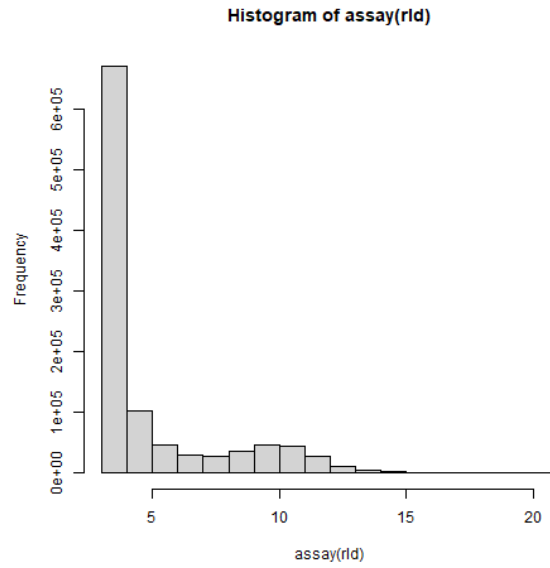


Figure 5.4: Pval dist

5.2 Feature Engineering

5.3 Feature Reduction

5.4 Modelling

5.5 Post-Analysis

5.6 Sensitization

Chapter 6

Discussion

Appendix A

Appendix

Bibliography

- Zhang, Chen et al. (Feb. 2022a). "Ferroptosis in cancer therapy: a novel approach to reversing drug resistance". en. In: *Mol. Cancer* 21.1, p. 47.
- Tait, Stephen W. G., Gabriel Ichim, and Douglas R. Green (May 2014). "Die another way – non-apoptotic mechanisms of cell death". In: *Journal of Cell Science* 127.10, pp. 2135–2144. ISSN: 0021-9533. DOI: 10.1242/jcs.093575. URL: <http://dx.doi.org/10.1242/jcs.093575>.
- Ye, Jing et al. (2018). "Non-apoptotic cell death in malignant tumor cells and natural compounds". In: *Cancer Letters* 420, pp. 210–227. ISSN: 0304-3835. DOI: <https://doi.org/10.1016/j.canlet.2018.01.061>. URL: <https://www.sciencedirect.com/science/article/pii/S0304383518300946>.
- Zou, Yilong and Stuart L. Schreiber (2020). "Progress in Understanding Ferroptosis and Challenges in Its Targeting for Therapeutic Benefit". In: *Cell Chemical Biology* 27.4. Special Issue on Chemical Biology of Ferroptosis, pp. 463–471. ISSN: 2451-9456. DOI: <https://doi.org/10.1016/j.chembiol.2020.03.015>. URL: <https://www.sciencedirect.com/science/article/pii/S2451945620301124>.
- Volm, Manfred and Thomas Efferth (Dec. 2015). "Prediction of Cancer Drug Resistance and Implications for Personalized Medicine". In: *Frontiers in Oncology* 5. ISSN: 2234-943X. DOI: 10.3389/fonc.2015.00282. URL: <http://dx.doi.org/10.3389/fonc.2015.00282>.
- Wang, Xiancan et al. (2023). "Effects of DNA, RNA, and Protein Methylation on the Regulation of Ferroptosis". In: *International Journal of Biological Sciences* 19.11, pp. 3558–3575. ISSN: 1449-2288. DOI: 10.7150/ijbs.85454. URL: <http://dx.doi.org/10.7150/ijbs.85454>.
- Li, Yuanyuan and Trygve O. Tollefsbol (2011). "DNA Methylation Detection: Bisulfite Genomic Sequencing Analysis". In: *Epigenetics Protocols*. Humana Press, pp. 11–

21. ISBN: 9781617793165. DOI: 10.1007/978-1-61779-316-5_2. URL: http://dx.doi.org/10.1007/978-1-61779-316-5_2.
- Heiss, Jonathan A. and Allan C. Just (June 2018). "Identifying mislabeled and contaminated DNA methylation microarray data: an extended quality control toolset with examples from GEO". In: *Clinical Epigenetics* 10.1. ISSN: 1868-7083. DOI: 10.1186/s13148-018-0504-1. URL: <http://dx.doi.org/10.1186/s13148-018-0504-1>.
- Gombert, Sara et al. (Oct. 2023). "Comparison of methylation estimates obtained via MinION nanopore sequencing and sanger bisulfite sequencing in the TRPA1 promoter region". In: *BMC Medical Genomics* 16.1. ISSN: 1755-8794. DOI: 10.1186/s12920-023-01694-6. URL: <http://dx.doi.org/10.1186/s12920-023-01694-6>.
- Flynn, Robert et al. (May 2022). "Evaluation of nanopore sequencing for epigenetic epidemiology: a comparison with DNA methylation microarrays". In: *Human Molecular Genetics* 31.18, pp. 3181–3190. ISSN: 1460-2083. DOI: 10.1093/hmg/ddac112. URL: <http://dx.doi.org/10.1093/hmg/ddac112>.
- Yan, Hong-fa et al. (Feb. 2021). "Ferroptosis: mechanisms and links with diseases". In: *Signal Transduction and Targeted Therapy* 6.1. ISSN: 2059-3635. DOI: 10.1038/s41392-020-00428-9. URL: <http://dx.doi.org/10.1038/s41392-020-00428-9>.
- Wang, Shufen et al. (Oct. 2019). "Transferrin receptor 1-mediated iron uptake plays an essential role in hematopoiesis". In: *Haematologica* 105.8, pp. 2071–2082. ISSN: 1592-8721. DOI: 10.3324/haematol.2019.224899. URL: <http://dx.doi.org/10.3324/haematol.2019.224899>.
- Knovich, Mary Ann et al. (May 2009). "Ferritin for the clinician". In: *Blood Reviews* 23.3, pp. 95–104. ISSN: 0268-960X. DOI: 10.1016/j.blre.2008.08.001. URL: <http://dx.doi.org/10.1016/j.blre.2008.08.001>.
- Lieu, P T et al. (Feb. 2001). "The roles of iron in health and disease". en. In: *Mol. Aspects Med.* 22.1-2, pp. 1–87.
- Imlay, James A (2003). "Pathways of oxidative damage". en. In: *Annu. Rev. Microbiol.* 57.1, pp. 395–418.
- Cabantchik, Zvi loav (Mar. 2014). "Labile iron in cells and body fluids: physiology, pathology, and pharmacology". In: *Frontiers in Pharmacology* 5. ISSN: 1663-9812.

- DOI: 10.3389/fphar.2014.00045. URL: <http://dx.doi.org/10.3389/fphar.2014.00045>.
- Pham-Huy, Lien Ai, Hua He, and Chuong Pham-Huy (2008). "Free radicals, antioxidants in disease and health". In: *International journal of biomedical science: IJBS* 4.2, p. 89.
- Ayala, Antonio, Mario F. Muñoz, and Sandro Argüelles (2014). "Lipid Peroxidation: Production, Metabolism, and Signaling Mechanisms of Malondialdehyde and 4-Hydroxy-2-Nonenal". In: *Oxidative Medicine and Cellular Longevity* 2014, pp. 1–31. ISSN: 1942-0994. DOI: 10.1155/2014/360438. URL: <http://dx.doi.org/10.1155/2014/360438>.
- Yin, Huiyong, Libin Xu, and Ned A Porter (Oct. 2011). "Free radical lipid peroxidation: mechanisms and analysis". en. In: *Chem. Rev.* 111.10, pp. 5944–5972.
- Gaschler, Michael M. and Brent R. Stockwell (Jan. 2017). "Lipid peroxidation in cell death". In: *Biochemical and Biophysical Research Communications* 482.3, pp. 419–425. ISSN: 0006-291X. DOI: 10.1016/j.bbrc.2016.10.086. URL: <http://dx.doi.org/10.1016/j.bbrc.2016.10.086>.
- Ademowo, O. S. et al. (May 2017). "Lipid (per) oxidation in mitochondria: an emerging target in the ageing process?" In: *Biogerontology* 18.6, pp. 859–879. ISSN: 1573-6768. DOI: 10.1007/s10522-017-9710-z. URL: <http://dx.doi.org/10.1007/s10522-017-9710-z>.
- Xie, Yangchun et al. (June 2023). "GPX4 in cell death, autophagy, and disease". In: *Autophagy* 19.10, pp. 2621–2638. ISSN: 1554-8635. DOI: 10.1080/15548627.2023.2218764. URL: <http://dx.doi.org/10.1080/15548627.2023.2218764>.
- Averill-Bates, Diana A. (2023). "The antioxidant glutathione". In: *Antioxidants*. Elsevier, pp. 109–141. ISBN: 9780443157684. DOI: 10.1016/bs.vh.2022.09.002. URL: <http://dx.doi.org/10.1016/bs.vh.2022.09.002>.
- Bridges, Richard J, Nicholas R Natale, and Sarjubhai A Patel (Jan. 2012). "System xc⁻ cystine/glutamate antiporter: an update on molecular pharmacology and roles within the CNS". en. In: *Br. J. Pharmacol.* 165.1, pp. 20–34.
- Ye, Yuanzhi et al. (Dec. 2022). "Repression of the antiporter SLC7A11/glutathione/glutathione peroxidase 4 axis drives ferroptosis of vascular smooth muscle cells to facilitate vascular calcification". en. In: *Kidney Int.* 102.6, pp. 1259–1275.

- Doll, Sebastian, Florencio Porto Freitas, et al. (Nov. 2019). "FSP1 is a glutathione-independent ferroptosis suppressor". en. In: *Nature* 575.7784, pp. 693–698.
- Yang, Ming et al. (May 2022). "Involvement of FSP1-CoQ10-NADH and GSH-GPx-4 pathways in retinal pigment epithelium ferroptosis". en. In: *Cell Death Dis.* 13.5, p. 468.
- Liu, Yu'e et al. (Aug. 2023). "The diversified role of mitochondria in ferroptosis in cancer". In: *Cell Death & Disease* 14.8. ISSN: 2041-4889. DOI: 10.1038/s41419-023-06045-y. URL: <http://dx.doi.org/10.1038/s41419-023-06045-y>.
- Doll, Sebastian, Bettina Proneth, et al. (Jan. 2017). "ACSL4 dictates ferroptosis sensitivity by shaping cellular lipid composition". en. In: *Nat. Chem. Biol.* 13.1, pp. 91–98.
- Krainz, Tanja et al. (Sept. 2016). "A mitochondrial-targeted nitroxide is a potent inhibitor of ferroptosis". en. In: *ACS Cent. Sci.* 2.9, pp. 653–659.
- Gao, Minghui et al. (Jan. 2019). "Role of Mitochondria in Ferroptosis". In: *Molecular Cell* 73.2, 354–363.e3. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2018.10.042. URL: <http://dx.doi.org/10.1016/j.molcel.2018.10.042>.
- Wang, Haidong et al. (Oct. 2016). "Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015". In: *The Lancet* 388.10053, pp. 1459–1544. ISSN: 0140-6736. DOI: 10.1016/S0140-6736(16)31012-1. URL: [http://dx.doi.org/10.1016/S0140-6736\(16\)31012-1](http://dx.doi.org/10.1016/S0140-6736(16)31012-1).
- Debela, Dejene Tolossa et al. (Aug. 2021). "New approaches and procedures for cancer treatment: Current perspectives". en. In: *SAGE Open Med.* 9, p. 20503121211034366.
- Bukowski, Karol, Mateusz Kciuk, and Renata Kontek (May 2020). "Mechanisms of multidrug resistance in cancer chemotherapy". en. In: *Int. J. Mol. Sci.* 21.9, p. 3233.
- Housman, Genevieve et al. (Sept. 2014). "Drug resistance in cancer: an overview". en. In: *Cancers (Basel)* 6.3, pp. 1769–1792.
- Zhang, Chen et al. (Feb. 2022b). "Ferroptosis in cancer therapy: a novel approach to reversing drug resistance". In: *Molecular Cancer* 21.1. ISSN: 1476-4598. DOI: 10.1186/s12943-022-01530-y. URL: <http://dx.doi.org/10.1186/s12943-022-01530-y>.

- Jiang, Le et al. (Mar. 2015). "Ferroptosis as a p53-mediated activity during tumour suppression". In: *Nature* 520.7545, pp. 57–62. ISSN: 1476-4687. DOI: 10.1038/nature14344. URL: <http://dx.doi.org/10.1038/nature14344>.
- Yang, Cuixia et al. (Jan. 2015). "Increased drug resistance in breast cancer by tumor-associated macrophages through IL-10/STAT3/bcl-2 signaling pathway". In: *Medical Oncology* 32.2. ISSN: 1559-131X. DOI: 10.1007/s12032-014-0352-6. URL: <http://dx.doi.org/10.1007/s12032-014-0352-6>.
- Jin, Bilian and Keith D Robertson (2013). "DNA methyltransferases, DNA damage repair, and cancer". en. In: *Adv. Exp. Med. Biol.* 754, pp. 3–29.
- Robertson, Keith D (Aug. 2005). "DNA methylation and human disease". en. In: *Nat. Rev. Genet.* 6.8, pp. 597–610.
- Pastor, William A, L Aravind, and Anjana Rao (June 2013). "TETonic shift: biological roles of TET proteins in DNA demethylation and transcription". en. In: *Nat. Rev. Mol. Cell Biol.* 14.6, pp. 341–356.
- Kohli, Rahul M and Yi Zhang (Oct. 2013). "TET enzymes, TDG and the dynamics of DNA demethylation". en. In: *Nature* 502.7472, pp. 472–479.
- Jones, Peter A and Stephen B Baylin (June 2002). "The fundamental role of epigenetic events in cancer". en. In: *Nat. Rev. Genet.* 3.6, pp. 415–428.
- Gardiner-Garden, M and M Frommer (July 1987). "CpG islands in vertebrate genomes". en. In: *J. Mol. Biol.* 196.2, pp. 261–282.
- Meissner, Alexander et al. (Aug. 2008). "Genome-scale DNA methylation maps of pluripotent and differentiated cells". en. In: *Nature* 454.7205, pp. 766–770.
- Mikkelsen, Tarjei S et al. (Aug. 2007). "Genome-wide maps of chromatin state in pluripotent and lineage-committed cells". en. In: *Nature* 448.7153, pp. 553–560.
- Eckhardt, Florian et al. (Dec. 2006). "DNA methylation profiling of human chromosomes 6, 20 and 22". en. In: *Nat. Genet.* 38.12, pp. 1378–1385.
- You, Jueng Soo and Peter A. Jones (July 2012). "Cancer Genetics and Epigenetics: Two Sides of the Same Coin?" In: *Cancer Cell* 22.1, pp. 9–20. ISSN: 1535-6108. DOI: 10.1016/j.ccr.2012.06.008. URL: <http://dx.doi.org/10.1016/j.ccr.2012.06.008>.
- Hanahan, Douglas and Robert A Weinberg (Mar. 2011). "Hallmarks of cancer: the next generation". en. In: *Cell* 144.5, pp. 646–674.

- Frommer, M et al. (Mar. 1992). "A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 89.5, pp. 1827–1831.
- Bird, A (July 1992). "The essentials of DNA methylation". en. In: *Cell* 70.1, pp. 5–8.
- Shiraishi, Masahiko and Hikoya Hayatsu (Dec. 2004). "High-speed conversion of cytosine to uracil in bisulfite genomic sequencing analysis of DNA methylation". en. In: *DNA Res.* 11.6, pp. 409–415.
- López-Catalina, Adrián et al. (Jan. 2024). "Oxford nanopore sequencing as an alternative to reduced representation bisulphite sequencing for the identification of CpGs of interest in livestock populations". en. In: *Livest. Sci.* 279.105377, p. 105377.
- Wang, Yunhao et al. (Nov. 2021). "Nanopore sequencing technology, bioinformatics and applications". In: *Nature Biotechnology* 39.11, pp. 1348–1365. ISSN: 1546-1696. DOI: 10.1038/s41587-021-01108-x. URL: <http://dx.doi.org/10.1038/s41587-021-01108-x>.
- Heather, James M and Benjamin Chain (Jan. 2016). "The sequence of sequencers: The history of sequencing DNA". en. In: *Genomics* 107.1, pp. 1–8.
- Depledge, Daniel P et al. (Feb. 2019). "Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen". en. In: *Nat. Commun.* 10.1, p. 754.
- Yuan, Yuxuan et al. (June 2017). "Improvements in genomic technologies: Application to crop genomics". en. In: *Trends Biotechnol.* 35.6, pp. 547–558.
- Martin, Samuel et al. (Jan. 2022). "Nanopore adaptive sampling: a tool for enrichment of low abundance species in metagenomic samples". In: *Genome Biology* 23.1. ISSN: 1474-760X. DOI: 10.1186/s13059-021-02582-x. URL: <http://dx.doi.org/10.1186/s13059-021-02582-x>.
- Kholodenko, Boris, Michael B Yaffe, and Walter Kolch (Apr. 2012a). "Computational approaches for analyzing information flow in biological networks". en. In: *Sci. Signal.* 5.220, re1–re1.
- He, Jie, Xiaoyan Li, and Mi Yu (2021). "Bioinformatics analysis identifies potential ferroptosis key genes in the pathogenesis of pulmonary fibrosis". en. In: *Front. Genet.* 12, p. 788417.

- McDermott, Ultan (Dec. 2015). "Next-generation sequencing and empowering personalised cancer medicine". en. In: *Drug Discov. Today* 20.12, pp. 1470–1475.
- Pais, Ricardo J, Ray K Iles, and Raminta Zmuidinaite (2021). "MALDI-ToF mass spectra phenomic analysis for human disease diagnosis enabled by cutting-edge data processing pipelines and bioinformatics tools". en. In: *Curr. Med. Chem.* 28.32, pp. 6532–6547.
- Oladipo, Elijah Kolawole et al. (Dec. 2022). "Bioinformatics designing and molecular modelling of a universal mRNA vaccine for SARS-CoV-2 infection". en. In: *Vaccines (Basel)* 10.12, p. 2107.
- Belmont, John W and Chad A Shaw (Nov. 2016). "Clinical bioinformatics: emergence of a new laboratory discipline". en. In: *Expert Rev. Mol. Diagn.* 16.11, pp. 1139–1141.
- Simon, Richard (Aug. 2011). "Genomic biomarkers in predictive medicine: an interim analysis". en. In: *EMBO Mol. Med.* 3.8, pp. 429–435.
- Kholodenko, Boris, Michael B Yaffe, and Walter Kolch (Apr. 2012b). "Computational approaches for analyzing information flow in biological networks". en. In: *Sci. Signal.* 5.220, re1–re1.
- Floros, Konstantinos V. et al. (Jan. 2021). "MYCN-Amplified Neuroblastoma Is Addicted to Iron and Vulnerable to Inhibition of the System Xc-/Glutathione Axis". In: *Cancer Research* 81.7, pp. 1896–1908. ISSN: 1538-7445. DOI: 10.1158/0008-5472.can-20-1641. URL: <http://dx.doi.org/10.1158/0008-5472.CAN-20-1641>.
- Danecek, Petr et al. (Feb. 2021). "Twelve years of SAMtools and BCFtools". In: *GigaScience* 10.2. giab008. ISSN: 2047-217X. DOI: 10.1093/gigascience/giab008. eprint: <https://academic.oup.com/gigascience/article-pdf/10/2/giab008/36332246/giab008.pdf>. URL: <https://doi.org/10.1093/gigascience/giab008>.

