Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biology

Victoria Nilsen Gjøvaag

Exploring Different Software Workflows to Optimise Biomarker Inference for Boolean Model Calibration and Drug Synergy Predictions

Master's thesis in Biotechnology
Supervisor: Martin Kuiper
Co-supervisor: Åsmund Flobak, Eirini Tsirvouli
May 2023

NTNU
Norwegian University of
Science and Technology

Victoria Nilsen Gjøvaag

Exploring Different Software Workflows to Optimise Biomarker Inference for Boolean Model Calibration and Drug Synergy Predictions

Master's thesis in Biotechnology
Supervisor: Martin Kuiper
Co-supervisor: Åsmund Flobak, Eirini Tsirvouli
May 2023

Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biology

NTNU
Norwegian University of
Science and Technology

Abstract

Accurate inference of intracellular signalling activity from omics data poses a challenge in systems biology approaches that leverages logical modelling for personalising cancer treatments. Despite the emergence of software tools analysing data on gene expression and protein levels, ensuring the quality of activity inference remains a challenge in translational research. This master's thesis focuses on using logical modelling, software tools and the DrugLogics pipeline as a solution to this problem. By evaluating four different software tools in inferring high-quality biomarker activities from transcriptomics data, this thesis aims to enhance the accuracy and effectiveness of *in silico* biomarker inference that can be used to personalise logical models within the DrugLogics pipeline. The findings of this project offers a framework for selecting suitable software tools for high-quality inference, and lay the groundwork for future research on improving computational activity inference from omics data. Developed scripts automate aspects of the biomarker inference process, and facilitate seamless data integration into the DrugLogics pipeline. The thesis work identified CONSENSUS as the most consistent and robust software tool in inferring biomarker activities for calibrating logical models in the DrugLogics pipeline. The software tools ULM and MLM also displayed potential as alternative options, but were found to be sensitive to parameter settings and dataset characteristics. Further investigation is recommended for subsets of activities from the PROFILE software, in combination with CONSENSUS activities. Optimal tool-specific parameters for precise inferences were identified, including the recommended expression count measure 'read counts' for CONSENSUS, MLM and ULM. Utilising the gene regulatory network CollecTRI and having a relatively strict p-value threshold may also be preferred with these tools, to obtain consistent inference results. If the PROFILE software is used, normalised expression data may be preferred. However, it is acknowledged that external factors, such as data characteristics, model limitations, and statistical measures may influence the conclusions drawn from the results. By evaluating software tools in inferring accurate biomarker activities from omics data, the findings of this master's thesis highlight the potential of leveraging *in silico* methods for logical model calibration in the DrugLogics pipeline, paving the way for more effective and personalised cancer treatment strategies.

Sammendrag

Identifisering av presise biomarkør aktiviteter fra biologisk data er en utfordring i systembiologiske metoder som tar i bruk persontilpassede logiske modeller for individbasert kreftbehandling. Til tross for fremveksten av programvareverktøy som analyserer genuttrykk og protein aktiviteter, forblir identifisering av presis biomarkør aktivitet en utfordring i translasjonsforskning. Denne masteroppgaven benytter logisk modellering og programvareverktøy for å analysere omikk data til aktivitetsnivåer, kombinert med DrugLogics-algoritmen som en løsning på dette problemet. Ved å evaluere fire forskjellige programvareverktøy for identifisering av presise biomarkører fra transkripsjonsdata, har denne oppgaven som mål å forbedre effektiviteten av *in silico* identifisering av biomarkør aktiviteter som kan utnyttes for å tilpasse logiske modeller i DrugLogics-algoritmen, og legge grunnlaget for at de kan benyttes i personlig kreftbehandling. Funnene fra dette prosjektet tilbyr et rammeverk for å velge effektive programvareverktøy for presis biomarkør identifisering, og danner et grunnlag for fremtidig forskning på *in silico* biomarkør identifisering fra biologisk data. Utviklede skript automatiserer aspekter av *in silico* identifisering av biomarkører, og tilbyr sømløs dataintegrasjon i DrugLogics-algoritmen. Resultatene viser at CONSENSUS var det mest presise og robuste verktøyet for identifisering av presise aktivitetsnivåer for tilpassing av logiske modeller i DrugLogics-algoritmen. Programvareverktøyene ULM og MLM viste også potensial som gode alternativer, men demonstrerte økt sensitivitet ovenfor parameterinnstillinger og datasettkvalitet. Videre forsking anbefales for å undersøke bruken av deler av biomarkørene identifisert av verktøyet PROFILE i kombinasjon med CONSENSUS aktiviteter. Optimale verktøy-spesifikke parametere for presis identifisering ble identifisert som bruk av 'read counts' for CONSENSUS, MLM og ULM. Å bruke det regulatoriske nettverket CollecTRI, og opprettholde en relativt lav p-verditerskel kan også foretrekkes med disse verktøyene, for å oppnå konsistent biomarkør identifisering. For PROFILE kan bruk av normalisert data være fordelaktig. Eksterne faktorer som datakvalitet, begrensninger ved logiske modeller og statistiske parametere kan imidlertid ha innvirkning på konklusjonene som kan trekkes fra disse resultatene. Ved å evaluere programvareverktøy for å identifisere presise biomarkør aktiviteter, viser resultatene fra denne masteroppgaven potensialet for å utnytte *in silico* metoder for kalibrering av logiske modeller i DrugLogics-algoritmen, og åpner mulighetene for mer effektive og individuelt tilpassede kreftbehandlingsstrategier.

Preface

This master's thesis is the culmination of my academic journey over the past two years. It has been an incredible learning experience, and I am grateful for the opportunity to explore a topic that has fascinated me for years. I have learned so much during the progression of this project, and I am (now) grateful for all the challenges I encountered along the way, constantly pushing me to expand my knowledge and abilities.

This master's thesis is part of the DrugLogics Initiative, and aims to explore different software workflows to optimise biomarker activity inference, with the ultimate goal of optimising the identification of novel drug synergies for targeted and personalised cancer treatments. It is my hope that this thesis will contribute to the era of personalised medicine, and inspire further research in this area.

I would like to thank my main supervisor, Martin Kuiper, for his guidance and valuable feedback throughout the entire progression of this project. In addition, I would like to extend my appreciation to my co-supervisors, Erini Tsirvouli and Åsmund Flobak, for their expertise and thoughtful suggestions. I am grateful for the trust you have placed in me, allowing me the freedom to shape and explore the project according to my own vision. This opportunity has been a great learning experience, and I am truly thankful for your confidence in my abilities.

Finally, I want to thank my family and friends for their unwavering support throughout this journey. A special appreciation to my parents for supporting me though the toughest days. I could not have completed this journey without you. I am also extremely grateful for my flatmates for adding a sense of fun to my daily life, and making every day memorable.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AGS: Adenocarcinoma Gastric cell line

AUC: Area Under the Curve

CI: Confidence Interval

CMP: Cell Model Passports

CNA: Copy Number Alterations

CRC: Colorectal Cancer

Drabme: Drug Response Analysis to Boolean Model Ensembles

FPKM: Fragments Per Kilobase of transcript per million Mapped fragments

FPR: False Positive Rate

GC: Gastric Cancer

Gitsbe: Generic Interactions To Specific Boolean Equations

GRN: Gene Regulatory Network

HSA: Highest Single Agent

MLM: Multivariate Linear Model

MOR: Mode Of Regulation

PPI: Protein-Protein Interaction

RC: Read Count

ROC: Receiver Operating Characteristic

SD: Standard Deviation

TF: Transcription Factor

TG: Target Gene

TPM: Transcripts Per Million

TPR: True Positive Rate

ULM: Univariate Linear Model

1 Introduction

1.1 The Complexity of Cancer: A Global Health Concern

Accounting for nearly 10 million deaths in 2020, cancer is one of the leading causes of deaths worldwide [1]. Cancer imposes a huge physical, emotional and economic burden on individuals, but also on families, communities, healthcare systems and whole nations. According to the World Cancer Report, it is projected that the global number of new cancer cases will rise to over 27 million per year by 2040, which represents a 50% increase from the estimated 18.1 million cases that occurred in 2018 [2]. These increases are related to a large number of potential cancer hazards, including alcohol and tobacco, eating habits, obesity and air pollution. Despite the rising cancer incidences, the past decade has seen remarkable advances in research and treatment, leading to an overall decrease in cancer related mortality rates. In fact, cancer survival in the United States has doubled over the last 40 years, and 50% of people that are diagnosed with cancer in England and Wales survive for ten years or more [3]. Nevertheless, cancer remains one of the most complex and challenging human diseases, and there is a pressing need for new and innovative strategies to fight it effectively.

1.1.1 From Molecular Interactions to Cancer Progression

A cancer cell arises when a normal cell progresses into a tumour cell through a series of characteristic changes. These changes can be classified as the 'hallmarks of cancer', and include promoting cell growth and division, resisting cell death and stimulating angiogenesis and metastasis [4]. Although cancer is a term that covers a broad range of subtypes, most cancers arise when changes occur in genes involved in, or controlling these processes. Thus, the hallmarks are crucial to understand the underlying mechanisms that are common in most cancer types. The hallmarks of a cancer cell can arise through a series of genetic alterations such as mutations, deletions, amplifications, copy number alterations (CNA), translocations and so on. Such alterations cause *molecular interactions* in vital signalling pathways to deviate from normal, contributing to the progression of a cancer cell.

When most people think of the word 'interaction' they might think of verbal communication and exchange of information between people. Similarly to human communication, molecular interactions convey the exchange of information between molecules, and the information is encoded in a language of chemical and electrical signals. In the world of biology, these signals are often thought of as cascades, in which one signal triggers a sequence of downstream signals. Given different signals, such cascades can cause changed activity or function of proteins, or even change the expression of specific genes. Disturbances in these pathways can lead to aberrant cellular responses, which may contribute to the development and progression of complex diseases like cancer [5]. To illustrate how disrupted molecular interactions in signalling cascades can lead to the emergence of cancer, lets look at an example below.

A healthy cell relies on a complex network of signalling pathways that regulate various cellular processes, including cell growth, differentiation and death. One of the most crucial signalling pathways for maintaining healthy cell proliferation and differentiation is the Wnt/β-catenin signalling pathway [6]. This pathway is initiated when a ligand binds to its receptor, which activates a cascade that ultimately leads to the stabilisation and nuclear translocation of the transcription factor (TF) β-catenin. When stabilised, β-catenin binds to members of a TF family called the the 'TCF/LEF family', thereby regulating the expression of downstream target genes (TG). Dysregulated interactions in this pathway can have serious and fatal consequences, and is a known characteristic of many types of cancers. Aberrant activation of the pathway can occur through genetic mutations in key components, causing dysregulated interactions throughout the signalling cascade. For example, mutations in β-catenin may cause aberrant activation of downstream TGs involved in cell proliferation and differentiation, ultimately resulting in uncontrolled cell growth and tumour formation. Consequently, while the genetic mutation in itself may not cause a fatal outcome, the changed activity of the mutated entity altered the whole

dynamic of the signalling cascade, ultimately causing a healthy cell to progress into a cancer cell.

Dysregulation of signalling pathways is actually a common feature amongst the hallmarks of cancer [5]. As a result, cancer can be thought of as a disease with abnormal molecular interactions in pathways vital to sustain the function of hallmark characteristics. Nevertheless, the fact that cancer is not caused by a single genetic mutation, but rather by the altered activity of mutated entities that can cause dysregulated interactions throughout intricate signalling cascades underscores the complexity of cancer. Thus, there is need for a comprehensive approach to understand the underlying mechanisms of cancer, and to develop effective treatment strategies.

1.1.2 Traditional Cancer Therapies

Traditional cancer therapies has long been based on cytotoxic chemotherapy and one-size-fits-all approaches [7]. Cytotoxic chemotherapies are a class of anticancer drugs that target rapidly dividing cells in the body, including cancer cells. They work by damaging the DNA of cells, which prevents them from dividing and growing. While effective in treating cancer, cytotoxic chemotherapies can have significant side effects, as they also affect normal cells in the body, such as those in the bone marrow and hair follicles [8] [9]. Moreover, cancer tumours are also highly heterogeneous, meaning that there may be different types of cancer cells within a single tumour, each with its own genetic and molecular characteristics [7]. Even cancers that stem from the same tissue can exhibit a diverse array of alterations [2]. The complexity of cancer and the challenge of treating it effectively is further exacerbated by the growing recognition of the contribution of peoples individual genetic makeup to the tumour biology [7]. Therefore, treating all cancer patients in the same manner, without considering the differences in tumour biology of individual patients, is no longer sufficient for effective cancer treatment. Fortunately, the past decade has seen a shift from traditional cancer treatments, largely due to the emergence of so-called high-throughput technologies and systems biology.

1.2 Systems Biology: The Key to Understanding Complex Biological Diseases

Over the past decade, remarkable advancements has been made in regards to understanding the structure of the DNA, and how changes in the DNA relates to health and disease [10]. High-throughput technologies, which refer to the use of automated methods for generating large-scale biological data, have been instrumental in this progress [11]. These automated methods have significantly accelerated the speed and reduced the cost of genome sequencing, making it now possible to sequence a human genome for 1000 US dollars [10] [12]. This represents a 50 000-fold reduction in the expense of human genome sequencing since the mid-2000s. In addition, high-throughput technologies are increasingly being used to generate so-called 'omics' data.

Omics is a collective term used to describe different fields of studies that analyse biological molecules or components within a biological system. Each omics field focuses on a specific type of biological molecule or component, including genomics (the study of genes), proteomics (the study of proteins), transcriptomics (the study of RNA molecules), and metabolomics (the study of metabolites) [13]. Omics has completely transformed the field of genetics and medicine, by providing rapid and low cost information regarding genetic changes such as mutations and CNAs, that give rise to the hallmarks of cancer and other diseases [13] [14]. At a patient level, such data can highlight certain genes, mechanisms or signalling pathways that might be aberrant by comparing the patient data to control data. As a result, high-throughput technologies have provided a pool of biological data that, analysed in the right way, can inform clinical decision making, diagnosis and treatment of complex diseases.

The pool of omics data generated by high-throughput technologies, combined with novel computational methods, and an increased understanding of the biological changes during disease development has prompt the emergence of systems biology as a field. While systems biology is a term that dates back to the early 2000s [15], its relevance has increased massively in the post-omics era. Systems biology is an interdisciplinary field of

research that aims to understand the organisation and behaviour of biological systems as integrated networks of complex interactions. It seeks to develop a comprehensive understanding of biological processes at a *system-level*, by integrating information from multiple scientific fields, such as biology, informatics, computational modelling, mathematics, statistics and others [15] [16]. By utilising interdisciplinary fields, systems biology provides a way of handling vast amount of omics data in a hierarchical way.

Biological systems can be described at various hierarchical levels, including cells, tissues, organs, and organisms [17]. These systems are composed of various components, depending on the system. A cellular system may comprise of components such as DNA, RNA, proteins, and metabolites, as well as environmental factors. Although knowledge of individual components remain important, the central focus of systems biology is to comprehend the overall structure and dynamics of whole systems [15]. In the field of systems biology, biological systems are viewed as interconnected *networks* that consist of numerous interdependent components. These components interact with each other in complex ways, resulting in the emergence of the system-level properties. Therefore, a comprehensive understanding of biological systems requires an analysis of the network of interactions between its different components.

1.2.1 Biological Networks

Understanding a cell's functional organisation, its *network biology*, is key to post-omic biomedical research [18]. To describe a biological system as a network means to describe its various components in a topological way, meaning the arrangement of the components and the relationships between them [18] [19]. The components in a biological network may be genes, proteins or small molecules, and are often termed 'nodes' within the field of network biology. In biological systems, these nodes are interacting with one another, and the interactions between nodes are called 'links' or 'edges'. Edges can for example portray both activating and inhibiting interactions, binding, formation of complexes and much more, depending on the biological system. The nodes and edges together form a biological network.

Traditionally, research regarding network biology was based on model organisms such as *Escherichia coli* and *Saccharomyces cerevisiae* [19]. However, high-throughput technologies and the Human Genome Project provided the necessary data to study networks in the human body [10]. The Human Genome Project was an international research initiative that played a vital part in unravelled the chemical sequence of the entire human genetic material [20]. Some of the most commonly studied human networks include protein-protein interaction (PPI) networks, metabolic networks, signalling networks and regulatory networks (TF - TG interactions) [18] [21]. Depending of the type of network, the edges can be either directed (have a direction) or undirected (a two-way relationship). In a regulatory network, the edges may be directed, as the interactions consist of TFs activating or repressing their TGs. Undirected interactions may for example represent two proteins binding to each other. All networks display different types and levels of interactions, and each give a piece of the puzzle that is the biology of the human body. Figure 1.1 illustrates a simple directed network with five nodes and six edges.

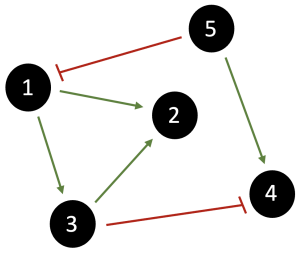


Figure 1.1: Example of a directed network consisting of five nodes and six edges. Green edges indicate an activating effects while red T-headed edges indicate inhibiting effects.

The topology of a biological network can provide insightful information in the fields of medicine and clinical research. This is due to the fact that molecular functions inside our bodies more often than not are carried out in a highly hierarchical manner [18]. As mentioned earlier, components in a given pathway can be affected by mutations in upstream components, which may result in changes throughout the pathway. By following the hierarchy of biological networks, mutation sites that may cause or contribute to the progression of a disease can be identified. Topological features of PPI networks are for example found to be very useful in identifying mutation sites and deregulated pathways in malignancy [22] [23] [24].

However, analysis of biological networks is challenging due to their complexity, size and interconnectedness. Gene regulatory networks (GRNs), for instance, can contain a large number of interactions between regulators (TFs) and their TGs, which may make it difficult to discern the meaningful patterns [25]. Additionally, GRNs are often highly dynamic and constantly changing in response to different environmental conditions, making their analysis even more challenging. Nevertheless, the study of biological networks is critical for advancing our understanding of biological processes, and developing new approaches for treating cancer and other complex diseases.

1.3 *In Silico* Modelling of Biological Networks

One systems biology approach of studying biological networks is through the use of computational *in silico* modelling and simulation. Computational modelling of a biological network involves creating mathematical models that simulate the behaviour of a biological system, based on knowledge of the interactions between the components [26]. The *in silico* biological network, which is supposed to represent the *in vivo* network, is constructed by using data from different omics. These *in silico* models can further be used to predict how the system will behave under different conditions, and to test the effects of various interventions or perturbations. This approach allows researchers to gain a deeper understanding of the underlying mechanisms of biological networks at a system-level, and predict how the system would respond to potential drug treatments. Several types of computational models can be distinguished, primarily classified as continuous or discrete [27]. Continuous and discrete models represent different levels of complexity, with discrete models being the most simple representation of a biological network.

Logical models, a sub-type of discrete models, offer a qualitative approach to describe biological networks [27]. Logical models define a set of rules to describe the behaviour (the logical state) of each component and its interactions with other elements within the system. The qualitative and discrete characteristics of logical models makes them easy to fit to biological systems, allowing researchers to focus on the most essential components of a network and model them in a simplified way. By reducing the complexity of biological systems to discrete logical states, logical modelling can be used to study and analyse the behaviour of large and interconnected biological networks. However, it is important to keep in mind that logical models are only capable of analysing networks in a qualitative way. In addition, the construction of a logical models can be a time-consuming and labour-intensive procedure, that requires a significant amount of experimental data about biological processes and signalling pathways.

1.3.1 Boolean Networks

Boolean networks are specific cases of logical models [27]. As mentioned previously, studying biological networks is challenging due to their sheer size and complexity. Boolean networks holds promise as a simplification of this complexity while still retaining dynamic properties thought to reflect vital biological information. In Boolean models each node has only two options: being active or being inactive. In the Boolean world this translates to 1's and 0's, respectively. Whether the activity of a node in the network is 0 or 1 is determined by its *logical rule*. The logical rules are node specific and are governed by the logical operators 'AND', 'OR' and 'NOT' [28]. These logical operators, together with the presence or absence of other nodes in the network,

determine a nodes activity level.

A node within the network can be regulated by multiple regulators, and the logical rule for the node is constructed by integrating the regulators with the appropriate logical operator [27] [28]. If regulators need to act together to activate a node, then they are combined with the AND operator. On the other hand, if the regulators are able to activate the target node independently, then the OR operator is used. For regulators that repress the target node, the NOT operator is used. NOT operators can also be coupled with either OR or AND operators. If only one regulator determines the activity of a node, then the logical rule is constructed using only the regulator name (with an addition of the NOT operator for repressors). For instance, in Figure 1.1, node 4 was connected to node 5 with an activating link and to node 3 with an inhibiting link. Consequently, node 4 may represented by the following logical equation:

$$4 = 5 \text{ AND NOT } 3 \tag{1}$$

From Equation 1 it can be seen that in order for node 4 to be active, node 5 needs to be active and node 3 needs to be inactive.

Despite the reduction of biological networks to 0's and 1's, Boolean models have proven to explain biological reality surprisingly well. For instance, Davidich & Bornholdt [29] used a Boolean network model to accurately predict the phenotype resulting from a knockout mutation in yeast. In fact, both logical and Boolean models combined with systems biology approaches, can be utilised to develop effective cancer treatments strategies. The advantages of utilising logical models for drug effect prediction are becoming more widely acknowledged [30], and that is not without reason. In the world of systems medicine, logical models provide as an economic alternative by lowering the bottleneck of experimental testing of drugs. In addition, they are relatively simple to use and easy to understand. They can be constructed based on existing biological knowledge, and their structure can be easily modified as new information becomes available. Logical models are also computationally efficient, making them suitable for large-scale simulations. And last but not least, logical models can provide accurate predictions of drug effects, especially for diseases where the underlying biological mechanisms are well understood. By integrating multiple data sources and experimental results, logical models can be used to identify drug targets and predict their effects on biological systems *in silico*.

1.4 *In Silico* Prediction of Drug Synergies: A New Approach to Personalised Cancer Treatments

Personalised medicine represents a transition from conventional therapeutic approaches that are designed based on the average patient. Instead, it focuses on developing tailored treatment strategies to small groups of patients, or even to individuals. Hood & Friend [31] states that *"This revolution is being fueled by several factors: first, an appreciation that medicine is an information science; second, systems or holistic approaches to studying the enormous complexities of disease; third, emerging technologies that will let us explore new dimensions of patient data space; and fourth, powerful new analytical technologies—both mathematical and computational—that will let us decipher the billions of data points associated with each individual."*

Targeted therapy lies in the heart of personalised medicine, and relates to drugs that are designed to target drivers of a specific disease [7] [31] [32]. These drugs target specific pathways or processes that are known to be related to the disease, for instance by inhibiting or regulating their central genes or proteins. Targeted drugs may be adapted to the individual patients needs, with respect to their unique genetic makeup and tumour biology. In addition, they may eliminate unwanted side effects of cytotoxic chemotherapies as they do not damage proliferating healthy cells, and overall generate smaller levels of toxicity. The add-on from cytotoxic chemotherapies to targeted therapies has largely been adopted by the field of medicine in the post-omics era [7].

Furthermore, it has been suggested that anticancer treatments in the future will adopt *combinatorial* drug therapies, where two or more anti-cancer drugs target a variety of robustness features or weaknesses of a specific tumour [32]. When two targeted drugs are combined to a combinatorial drug therapy, they can have *synergistic effects*. A synergistic interaction can be defined as one that performs better than the expected additive effect of two drugs [33] [34]. Identification of effective drug synergies is a vital step towards fighting cancer efficiently, as they have proven to be more potent at killing tumour cells, while decreasing drug resistance and allowing for reduced drug dosages, thus decreasing the likelihood of overlapping toxicity [32]. However, to clinically test all possible drug combinations would be costly and take a long time, to say the least. To illustrate this, the National Cancer Institute (NCI) in the United States lists 676 drugs for cancer and conditions related to cancer [35]. To pairwise test all of these drugs would result in 228 150 possible combinations. Taking into account that drug synergy could involve three or even four drugs, the number rises to over 51 million and 8 billion combinations, respectively. In addition, if dosage optimisation of each individual drug are taken into account one can only imagine the resulting number of combinations.

In silico methods can be suitable for the development, assessment, and prioritisation of drug combinations for clinical testing, by simulating the impact of drugs on biological systems [36]. In fact, recent studies have shown that logical models hold great promise for predicting novel and targeted drug combinations [30] [36] [37] [38] [39]. In addition, logical cancer models can be modified to represent the unique tumour biology of individual patients. A generic logical model can be customised to represent the genetic variation in cancer patients by customising it to fit to the characteristics of their individual tumour biology. In simple, this is done by changing the logical formalism of the model to fit to the biological behaviour of the patients cancer state. In order to do so, patient data from the individuals cancer cells need to be obtained, and used to *personalise the logical model*.

1.4.1 Personalising Logical Models with Biomarkers

The term biomarker, a fusion of 'biological marker', refers to the quantifiable indication of the state of a cell or an organism [40]. In principle, anything that can be used to describe or measure features of a biological system can be called a biomarker. However, in the context of bio-medicine, biomarkers often relate to different biological molecules, for instance DNA, RNA and proteins, which can be used in clinical research as indicators of diseases.

Clinical practice has a long-standing tradition of relying on biomarkers, and they are still recognised as the most reliable and objectively measurable medical indicators of disease [40]. However, the use of biomarkers, particularly laboratory-measured ones, is relatively new in clinical *research*, and the most effective methods for implementing this practice are still being developed and improved. One challenge is to establish the association between a specific biomarker and a relevant clinical outcome. In reality, this will be how a specific biomarker is related to a given state of being, for example a state of cancer. Every cancer patient may have their own unique set of personal biomarkers indicating the unique characteristics and behaviour of their tumour biology.

As mentioned earlier, personal biomarkers may be used to customise a generic cancer model to represent a persons specific cancer state. However, in order to determine that an entity is a biomarker for a persons cancer, there needs to be strong linking of that entity to the unique characteristics of the persons tumour biology. This typically refers to biological entities that exhibit a strong association to the observed activity in the patient's data, and may potentially drive the disease progression in the patient. Such patient data may for instance be transcriptomics or proteomics. As mentioned earlier, the real bottleneck is no longer the availability of relevant biological data, it is to interpret the large amount of data in a meaningful way. In recent years, several software tools have been developed to infer biological activity levels of proteins and genes from omics data. Such tools are able to analyse gene expression levels to estimate gene activity, assess protein abundance or activity, measure phosphoproteomics (the level of protein phosphorylation), and use matrices denoting the relation between regulators and their targets to estimate regulatory activities. Biological regulators like TFs and kinases

may for instance be inferred from downstream gene expression levels and proteomics, respectively. However, the quality of *in silico* biomarker inference remains a challenge in systems biology approaches and clinical research. In this context, quality biomarkers refers to biomarker activities that are accurate and representative of the underlying biology of the sample in which they are inferred from. Given the complexity and interconnected of biological networks, finding tools that are able to infer high-quality biomarker activities is challenging, but crucially important.

1.5 Project Objectives

The primary objective of this master's thesis was to enhance the accuracy and effectiveness of *in silico* inference of biomarker activities for personalising logical models in the DrugLogics pipeline. The goal was to be achieved through evaluating the performance of different software tools at inferring high-quality biomarkers from omics data, by testing them in configuring logical models that accurately represent and predict cellular behaviour. The term 'biomarker activities' is in this project related to biological *activity levels* of entities inferred from omics data by software tools. By using high-quality biomarker activities to personalise logical models, the overarching aim of this project is to pave the way towards more targeted and personalised cancer treatments.

The objectives related to this project are:

- To identify software tools capable of inferring high-quality biomarker activities from omics data, and to utilise these biomarker activities to calibrate logical models that may be used to predict accurate drug synergies, for personalised and targeted cancer treatment.
- To identify optimal combinations of tool-specific parameters, to ensure accurate biomarker activity inference. This involves identifying the most optimal combinations of parameters and settings for each software tool, such as input data types, regulatory networks, and statistical thresholds, to maximise the accuracy and robustness of high-quality biomarker activity inference.

The following specific questions are also addressed:

- Is the inference performance of the software tools generalisable across different cancer models and cell lines?
- Do optimal tool-specific parameter settings persist across models and cell lines?
- Can the work in this thesis support recommendations for use of specific new tools and approaches in the DrugLogics pipeline?

2 Materials and Methods

This project was carried out utilising a systems biology approach. A brief overview of the workflow conducted throughout this project is outlined here, while the succeeding sections provide a comprehensive explanation of the tools and methods utilised. Initially, three cancer models were gathered from different research groups. Next, a number of cell line datasets were obtained for each cancer model, and four software tools were utilised to infer biomarker activities from the datasets. The inferred biomarker activities were fitted to the format of the synergy prediction pipeline of the DrugLogics Initiative. The fitted data were subsequently used to calibrate logical models in the DrugLogics pipeline, resulting in drug synergy predictions. To asses the inference accuracy of the software tools, AUC ROC values were calculated for each of the pipeline-generated synergy predictions. Then, to evaluate the significance of the obtained synergy results, a statistical analysis using bootstrapping was carried out. This involved re-running the pipeline to generate new synergy predictions using random activity data. By comparing the synergy results generated by using biomarker activities inferred by the software tools to the random results, the statistical significance of the predictions could be discussed. Finally, the results of this analysis were evaluated to determine the most effective and precise tool for identifying biomarker activities from omics data to calibrate logical models in the DrugLogics pipeline. Figure 2.1 displays a brief overview of the workflow conducted throughout this project. Overall, this workflow provided a systematic approach to analysing software inference tools, using rigorous data processing techniques and ensuring the reproducibility of the results obtained.

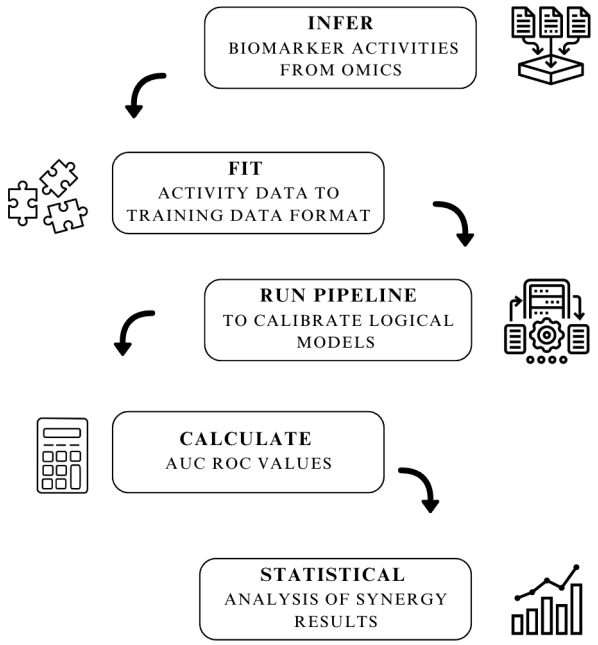


Figure 2.1: Overview of the workflow conducted to analyse the performance of software tools in inferring biomarker activities from omics data. First, biomarker activities were inferred from cell line transcriptomics data, and the activities were converted to the training data format DrugLogics synergy prediction pipeline. Next, the pipeline was executed with the biomarker activities to calibrate logical models and generate drug synergy predictions. Finally, AUC ROC values were calculated, and a statistical analysis was performed to determine the significance of the obtained results.

An overview of the different versions of software packages, databases and tools utilised in this project can be found in Appendix A. In addition, all related scripts and files necessary to redo the analysis have been made available on GitHub: https://github.com/victoriagjovaag/Master-Thesis.

2.1 The DrugLogics Pipeline

This master thesis has been carried out as part of the DrugLogics Initiative. The DrugLogics Initiative comprise of multiple sub-projects, all aiming to utilise multi-omics data combined with logical models towards the development of personalised and targeted medicine [41]. DrugLogics operates on a pipeline-based architecture, which relates to a series of data processing steps, where the output of one element is the input in the next element. In this way, pipelines are able to facilitate the management and processing of large and complex datasets.

The DrugLogics Initiative is composed of several modular components, which can be viewed from the related *GitHub repository*. Each modular component performs a specific data processing or analysis task. This project aims to improve the predictive results of one of these software modules: the 'druglogics-synergy' module. The druglogics-synergy module, referred to as the 'DrugLogics pipeline' from this point, is a software pipeline module that provides an *in silico* foundation that guides the generation of disease specific logical models, and predicts the response of drug combinations on those models [42]. The pipeline is also designed to identify novel drug synergies for a given cancer model. The goal is that the DrugLogics pipeline will be an automated modelling pipeline capable of generating logical models for any cancer network that may be customised using baseline biomarker data from patient tumours, and provide combinations of drug treatments that should be further examined in pre-clinical and clinical settings (Figure 2.2)

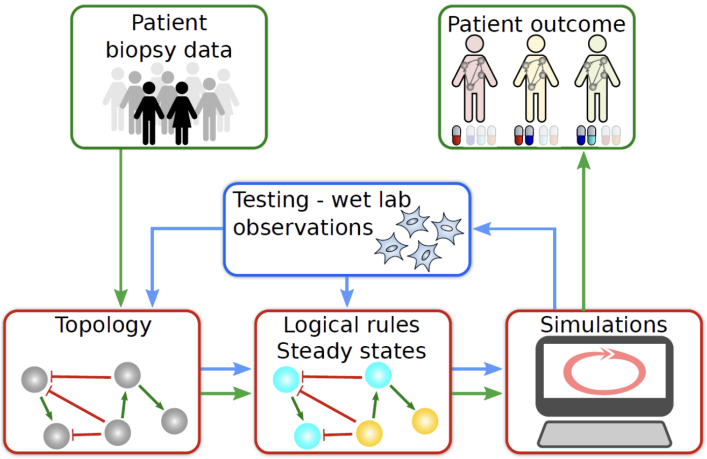


Figure 2.2: Simplified overview of the DrugLogics pipeline workflow. The pipeline is designed to be an automated modelling pipeline capable of generating logical models for any cancer network that may be customised using baseline biomarker data from patient tumours, and provide combinations of drug treatments that should be further examined in pre-clinical and clinical settings. The figure is retrieved from https://druglogics.eu/projects/colosys/

.

The DrugLogics pipeline subsequently runs two software modules: 'Gitsbe' and 'Drabme'. Initially, Gitsbe is utilised to generate a collection of Boolean models that are calibrated to specific biological activity levels, by using a genetic parameterisation algorithm. Subsequently, Drabme uses the calibrated models from Gitsbe to conduct an analysis of drug responses to these models, and predicts scores representing the synergistic effect of the drugs. The full documentation for the Gitsbe and Drabme modules can be accessed at the *DrugLogics Software Documentation*.

2.1.1 Gitsbe

Gitsbe is an acronym for 'Generic Interactions To Specific Boolean Equations', which highlights its function of converting generic interactions into Boolean formulas. In order to run the Gitsbe module, there are some required inputs:

- *A network:* A file displaying a cancer network. The network needs to be in Cytoscape's .sif format, and its interaction statements must be directed and binary signed.
- *A training data file:* A file that contains a condition-response pair, where the condition refers to a specific biological state (for instance a steady cancer state), and the response relates to observed activity levels (for instance inferred biomarker activities) at this state.
- *A model outputs file:* A file with biological entities that directly influence the model outputs (proliferation or cell death).
- *A configuration file:* A file where common options are specified.

In addition, there are some optional inputs:

- *A project name:* If specified, it will be the title of the folder storing the outputs of Gitsbe.
- *A drug panel file:* A file with information about the drugs that will be tested. This file is required when the training data file specifies a single or double drug perturbation condition.

The Training Data File
The condition-response pairs specified in the training data file are used to calibrate logical models in the Gitsbe module. There are several different condition - response options that are supported in the DrugLogics pipeline:

- *Knockout/overexpression condition - globaloutput response:* This is a global output response to a knockout and/or an overexpression. The globaloutput response refers to an overall cell fate which must be specified in the [0,1] interval (ranging from cell death to a cell proliferation state). This option allows to calibrate models to fit to a knockout/overexpression of one or several biological entities in the model network.
- *Single drug perturbation - globaloutput response:* This is a global output response to a drug perturbation. This option allows to calibrate models to fit to the effect a single drug has on the network model.
- *Double drug perturbation - globaloutput response:* This is a global output response to a double drug perturbation. This option allows to calibrate models to fit to the synergistic effects a drug combination has on the network model.
- *Unperturbed condition - globaloutput response:* This is a global output response to an unperturbed system.
- *Unperturbed condition - steady state response:* This is a steady state response to an unperturbed system. A steady state means that the condition of the system does not change over time. The nodes specified as a response will be biological entities and activity levels related to the steady state.

The 'unperturbed condition - steady state response' option is the most commonly used option in the DrugLogics pipeline, and also the option used throughout this project. In this project, the steady state response relates to the inferred biomarker activities from cell line data.

The content of the training data file needs to be formatted in a specific way:

Condition
[data]
Response
[data]
Weight: [number]

The training data file, together with the other inputs, are used by the Gitsbe module to calibrate logical models to fit to the biological observations described in the training data. This process is illustrated in Figure 2.3, and can can be summed up by three steps.

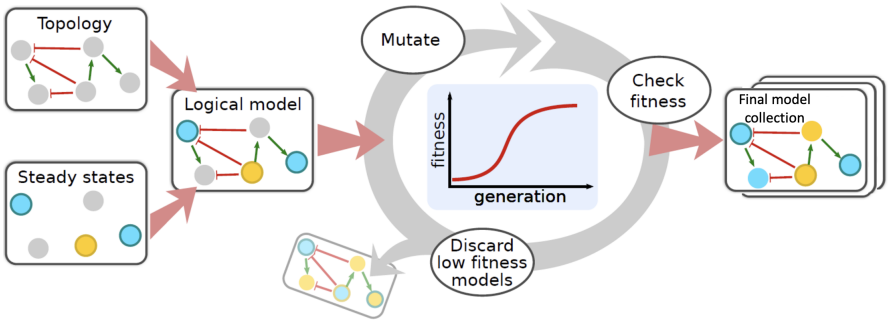


Figure 2.3: Gitsbe cycle for utilising training data to calibrate logical models in the DrugLogics pipeline. Using prior knowledge of molecular interactions (signalling topology), along with steady state data (biomarker activities), logical models are constructed with predefined rules. A genetic algorithm iteratively mutate the logical logical rules of the models and calculate their fitness score. This process is performed simultaneously for multiple models, resulting in an optimised ensemble that closely matches the observed biomarker activities. The figure is modified from Flobak et al. [42].

Step 1:
The model interactions defined in the network .sif file are converted to logical equations. This is accomplished by relating the nodes to their regulators in the network, based on the characteristics of the following equation:

Target = (A OR B) AND NOT (C OR D)

In this equation, A and B are activating regulators, and C and D are inhibitory regulators. The logical operator 'OR' links the A and B, and C and D together. Similarly, the 'AND NOT' operator links to two pairs of regulators together. Essentially, the expression states that either regulator A or B must be present, and in addition both C and D must be absent for the target to be active. If one or more of these criteria are not met, the target node will be inactive. Each node in the network will end up with its own logical equation, based on its interactions with the other entities in the network.

Step 2:
Gitsbe uses a genetic algorithm to create logical models that fit to the biomarker activity data that is stated in the training data file. The creation of these logical models follows a series of repeated steps:

- An initial generation of logical models are randomly generated from the original model (the logical model created from the network .sif file) in a mutation phase. This first collection of models are generated by randomly introducing mutations in the logical equations of the original model (for example switching an 'AND NOT' operator to an 'OR NOT', or vice versa).
- A fitness score is calculated for each model. This score relates to how well a given model represents the biomarker activity observations in the training data. The fitness score is a continuous value between 0 (no fitness at all) and 1 (perfect fitness).

- The models with the highest fitness scores go on to the next phase.
- The mutation phase is repeated, where the logical equations of the highest ranking models are exchanged in a crossover phase. This phase will produce the next generation of models, which then again will have a fitness score calculated.

Step 3:
After repeating the series of steps for a specified number of times (as specified in the configuration file), a final collection of output models are obtained, that best represent the observations in the training data file.

2.1.2 Drabme

After logical models are calibrated by Gitsbe, Drabme (Drug Response Analysis to Boolean Model Ensembles) uses them together with additional input files to identify and evaluate their responses to drug combinations. Like Gitsbe, Drabme has some required inputs:

- *A models directory:* A directory of logical models that was generated by Gitsbe.
- *A drug panel file:* A file storing information about the different drugs that will be tested on the models.
- *A model outputs file:* The same model outputs file that was an input in Gitsbe.
- *A configuration file:* The configuration file that also was an input in Gitsbe.

Drabme also has some optional inputs:

- *A project name:* If specified, it will be title of the folder storing the outputs of Drabme.
- *A perturbations file:* A file listing the single and combined drug perturbations to be used.

Drabme will use these inputs to predict drug synergies following these three steps:

1. For each model from Gitsbe, all perturbations specified in the perturbations file are simulated. These perturbations results in a new logical model for every combination of models (in the models directory) and drugs (in the perturbation file). The new models are generated by permanently changing the logical equations of the perturbed target nodes in the models to either false (when the drug inhibits its target) or true (when the drug causes the expression of its target).
2. Next, a global output response parameter referred to as the 'growth value' is calculated for each model. Essentially, a growth value of 0 denotes zero growth (cell death state), whereas a growth value of 1 represents a model with maximum growth (a proliferating cell).
3. Lastly, the growth values of the perturbed models are utilised to assess which drug combinations are synergistic.

In the configuration file one of two mathematical models can be selected to evaluate if two drugs are synergistic, either HSA (Highest Single Agent) or Bliss Independence. The HSA mathematical model was used exclusively in this project as it is a less conservative model in identifying synergies, which enables more potential synergies to be classified. These mathematical models work by comparing the expected additive response of two drugs, to the observed or predicted combined response [43]. If the combined response is lower than the additive response, the drugs are considered to be synergistic, whereas if it is higher, they are considered antagonistic (meaning that the combined effect of the drugs is less than what would be expected based on the sum of their individual effect) [44]. In the case of the HSA model, the minimum of the two single-drug perturbed

model responses is used as the additive response in the pipeline. Hence, if the output response of the two-drug perturbed model is lower than the minimum of the two single-drug perturbed models, the combination of the two drugs is considered to act *synergistically* on that particular model.

There are two ways to assess synergies in Drabme: model-wise and ensemble-wise. In the model-wise approach, Bliss or HSA is used to compare the number of models predicting a drug combination as synergistic or antagonistic. In the ensemble-wise approach, the average single-drug responses and average combination response are calculated across available models, and Bliss or HSA is used on these ensemble-wise values. The ensemble-wise approach was utilised in this project.

The so-called 'HSA-exceed' value is one of the main outputs of Drabme and relates to the the difference between the average combination response and the minimum of the two average single-drug responses. The more negative exceed value, the stronger is the indication of the synergy. A positive value indicates a more antagonistic effect.

2.2 Collecting Models and Cell Lines

2.2.1 Model Selection

Specific cancer models needed to be selected in order to to run the DrugLogics pipeline. The cancer models used in this analysis were selected based on availability of necessary background information in order to run the pipeline. The model-specific inputs that are needed to run the DrugLogics pipeline are:

1. *Logical rules* of the components in the cancer network. This information is vital in order to create the model network file (the .sif file) and the model outputs file that is needed as Gitsbe inputs.
2. *Observed synergies* for appropriate cell lines of the cancer model. This information is needed in order to be able to compare the resulting synergies to known synergies. Drug synergies are cell line specific, so when selecting a model there needed to exist information about drug synergies for appropriate cell lines for the respective cancer model.

Three logical cancer models were selected based these criteria: a gastric adenocarcinoma model, a colitis-associated colon cancer (CAC) model and a colorectal cancer (CRC) model. All necessary files used in this project to run these models with the DrugLogics pipeline can be found in the related *GitHub folder*.

The CASCADE 1.0 Model
CASCADE 1.0 is a logical model representing the cell fate decision network in the human gastric adenocarcinoma cell line (AGS) [30]. The main reason why CASCADE 1.0 was chosen for this project was that it was specifically developed by the DrugLogics group for use in the DrugLogics pipeline. As a result, the model fulfilled the necessary criteria for running the pipeline. In addition, CASCADE 1.0 is a relatively small model in terms of number of nodes and edges, providing a fast compilation time when running the pipeline.

CASCADE 1.0 was constructed to contain key signalling pathways in AGS [30]. The construction of the model was based on knowledge of the mutations in the AGS cell line from scientific papers and databases. Amongst others, this comprised of the MAPK, PI3K, Wnt/β-catenin and NF-κB pathways. These pathways, as well as the interacting components between them, were included in the CASCADE 1.0 model to represent the cell fate decision network of AGS cells. How these entities are connected can be seen from the interaction network of CASCADE 1.0 in Figure 2.4.

Both transcriptomics and proteomics data from the AGS cell line was used to curate CASCADE 1.0, as well as cancer and cell line databases, like the Cancer Genome Atlas and the Cancer Cell Line Encyclopedia. The final model consists of 75 nodes and 149 edges. Each component in the network is associated with a logical

rule that defines its activity level based on the presence or absence of its regulators. The logical formalism of the model is defined to recapitulate unperturbed AGS cells in their baseline proliferating state, and can be found in Appendix B. The network does not have any external input components, but contains two phenotypic output nodes, Prosurvival and Antisurvival, representing the cell fates of the model. Figure 2.4 displays that Prosurvival is directly influenced by active MYC and CCND1, whereas Antisurvival is directly influenced by CASP3 and FOXO_f. These nodes are again affected by other components in the network. See Flobak et al. [30] for the full documentation of the model construction.

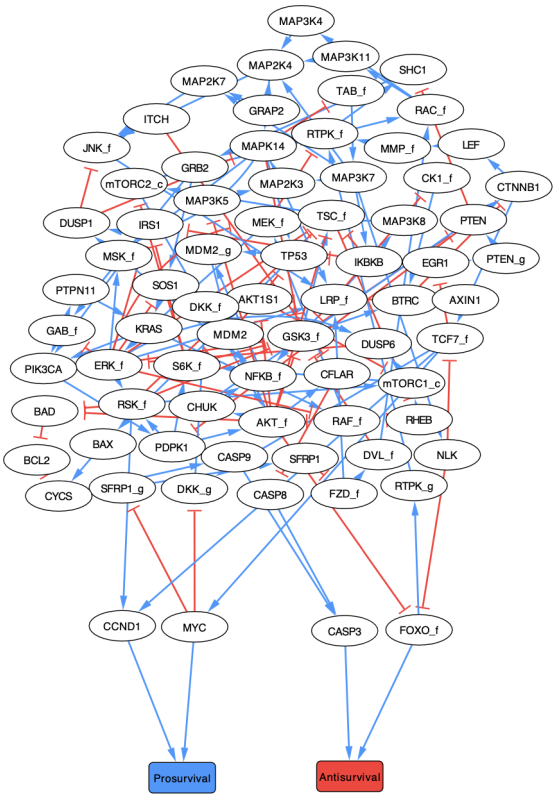


Figure 2.4: Cell fate decision network of AGS gastric adenocarcinoma cell line model (CASCADE 1.0) derived from Flobak et al. [30]. The network does not receive any external inputs but consists of two outputs: Antisurvival and Prosurvival (coloured in red and blue, respectively, as phenotypic indicators). The signalling components of the network are proteins, protein complexes or genes, each connected to a logical rule that determines its Boolean activity level (0 or 1). Blue arrows indicate an activating effects while red T-headed lines indicate inhibiting effects.

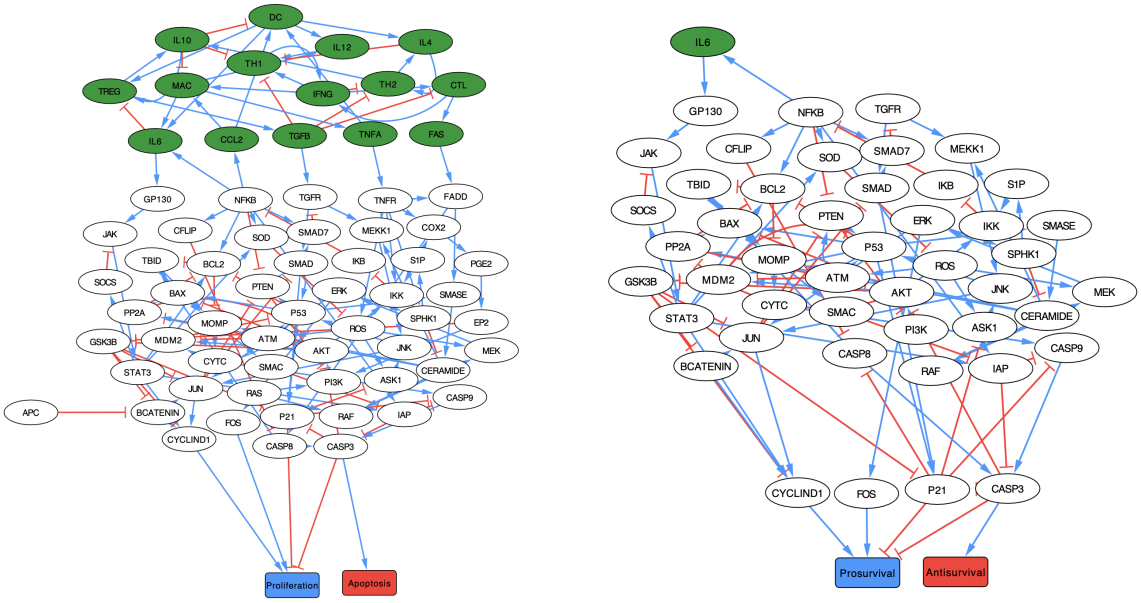
By utilising the CASCADE 1.0 model Flobak et al. [30] were able to identify four drug synergies that had a synergistic effect in inhibiting gastric cancer (GC) cell growth in laboratory testing.

The Lu Model

The CAC network model used in this project was constructed by Lu et al. [45] to investigate the mechanism underlying CAC. This model will be referred to as the 'Lu model' from this point onward. The Lu model is a Boolean logical model incorporating multiple layers of information, including gene expression data, protein-protein interactions (PPI), and known signalling pathways in intestinal epithelial cells. As a result, the model reveals key modules that are dysregulated in CAC, including modules involved in inflammation, cell proliferation and apoptosis. The model was used by Lu et al. [45] to identify potential therapeutic targets. The research group found that several key nodes in the network, such as IL-6, STAT3, and NF-κB, were highly connected and could serve as effective targets for anti-cancer therapy. Additionally, they identified several novel targets that had not been previously implicated in CAC, including the protein kinase PAK1 and the transcription factor (TF) SRF. By combining multiple targets, the authors were able to construct combinatorial therapies that

displayed enhanced efficacy in reducing tumour growth in mouse models of CAC.

The original Lu model network consists of 70 nodes and 153 edges. However, in this project it was necessary to focus solely on the intracellular signalling pathways, as these pathways often are the ones that are dysregulated in cancer cells, and the ones that can be targeted by drug therapy. In addition, by focusing of the intracellular pathways, it is not necessary to specify inputs, thereby building on the endogenous proliferative nature of cancer cells. To accommodate this, the nodes and edges related to the extracellular immune microenvironment were removed. The goal was that reducing the network to include only intracellular components would simplify the model while still capturing the key signalling pathways involved in CAC. The reduction resulted in removing 14 nodes and all their corresponding edges. The original and modified Lu models are displayed in Figure 2.5.



(a) Original Lu model network. **(b) Modified Lu model network.**

Figure 2.5: Cell fate decision network of a colitis-associated colon cancer model derived from Lu et al. [45]. Panel (a) displays the original CAC model and Panel (b) displays a modified CAC model where 14 nodes and edges related to the extracellular immune microenvironment are removed. The model have two phenotypic output nodes: Antisurvival and Prosurvival (coloured in red and blue, respectively). The signalling components of the network are proteins, protein complexes or genes, each connected to a logical rule that determines its Boolean activity level (0 or 1). Green nodes are related to the extracellular immune microenvironment. Blue arrows indicate an activating effects while red T-headed lines indicate inhibiting effects.

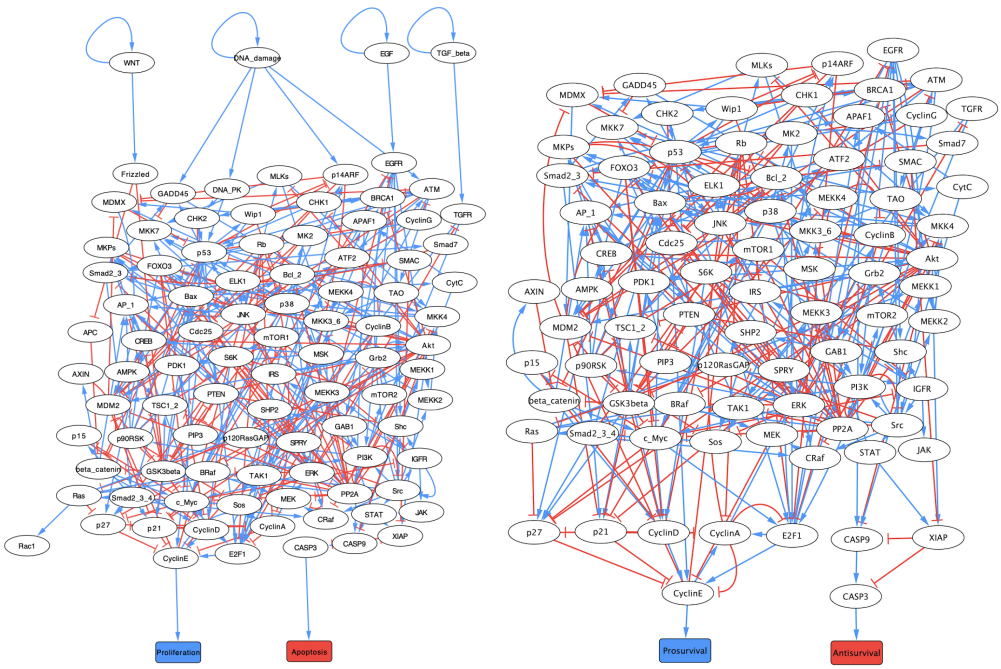
Same as for the CASCADE 1.0 model, there are two phenotypic output nodes in the Lu model. Each node in the Lu network is also associated with a corresponding logical rule, and thus it was possible to create the necessary input files for running the DrugLogics pipeline with this model. The network files for the Lu model was created in another DrugLogics master's thesis by Thea Hettasch (*'Effect of calibration data subsetting on Boolean model calibration and drug synergy predictions'*, May 2023). The network file was created by manually converting the logical formalism of the Lu model to tab-delimited, directed and binary signed single interactions. The model outputs file was created from the topological network by annotating all nodes contributing to Apoptosis as -1, and all nodes contributing to Proliferation as 1. The nodes that contributed to either Apoptosis or Proliferation were assumed to be nodes linked to a node directly connected to the phenotypic output (Figure 2.5). In the case of Apoptosis, this involved annotating all nodes with an *activating* link to CASP3 with a value of -1, and all nodes with an *inhibiting* link to CASP3 with a value of 1. Similarly for Proliferation, all nodes with a direct link to either P21, CASP3, FOS or CYCLIND1 were annotated as either 1 or -1 depending on their activating or inhibiting contribution, respectively. The logical formalism of the original and reduced Lu models can be found in Appendix B. The observed synergies of appropriate cell lines for the Lu

model was derived from the findings of Jaaks et al. [46].

The Park Model

The CRC model used in this project was developed by Park et al. [47], and will be referred to as the 'Park model' from this point forward. The Park model is a Boolean network model representing the large-scale signalling events in the HCT-116 CRC cell line (Figure 2.6). The Park model was based on integrating PPI data, gene expression data, and information on signalling pathways and feedback loops. The model incorporates several key components of signalling pathways involved in CRC progression, including signalling events such as the MAPK/ERK, PI3K/AKT and Wnt/β-catenin pathways. The model also includes feedback loops that can modulate the activity of these pathways in response to external stimuli, such as drug treatments. Integration of PPI data and feedback loops in this way allows for a more comprehensive understanding of the complex signalling networks involved in CRC.

Using the CRC model, Park et al. [47] found that the combination of inhibiting MEK and JNK pathways, was particularly effective in disrupting feedback loops and promoting cell death in CRC cells. As a result, inhibition of the MEK and JNK pathways could enhance the efficacy of targeted therapy in CRC by suppressing compensatory signalling pathways that can lead to drug resistance. The model analysis also revealed that the combination treatment was able to overcome adaptive resistance to multiple targeted therapies, suggesting that this approach could be applicable for the treatment of CRC.



(a) Original Park model network. **(b) Modified Park model network.**

Figure 2.6: Cell fate decision network of a colorectal cancer model derived form Park et. al [47]. Panel (a) displays the original CRC model and Panel (b) displays a modified CRC model where input nodes and related edges are removed. The network have of two output nodes representing phenotypic outputs: Antisurvival and Prosurvival (coloured in red and blue, respectively). The signalling components of the network are proteins, protein complexes or genes, each connected to a logical rule that determines its Boolean activity level (0 or 1). Blue arrows indicate an activating effects while red T-headed lines indicate inhibiting effects.

As shown in Figure 2.6, the original Park model consists of 95 nodes and 341 links, with Apoptosis and Proliferation representing the two phenotypes produced by the model. The original model also contains four input nodes: EGD, DNA damage, WNT and TGFβ, which represent the external stimuli that can activate the signalling pathways involved in the cancer progression. Similar to the Lu model, these external input nodes

(and the ones that were soley dependent on them) were removed from the Park model prior to using it in the DrugLogics pipeline. In the same manner as the other cancer models, there are some signalling pathways (like the MAPK/ERK and PI3K/AKT pathways) that can promote cell proliferation, survival, and invasion of CRC. Understanding the mechanisms underlying the activation of these pathways is therefore important for developing effective cancer therapies of CRC. The logical formalism underlying the original and reduced Park models can be found in Appendix B.

The necessary files used in this project to run the DrugLogics pipeline with the Park model were created in a similar way as for the Lu model in the master's project of Thea Hettach, by utilising the models logical formalism and network topology. The observed synergies of appropriate cell lines for the Park model was also derived from the findings of Jaaks et al. [46].

2.2.2 Cell Line Selection

In order to utilise software tools to infer biomarker activities, it was necessary to select appropriate cell line datasets to infer the activities from. Cell lines are cell populations derived from single cell sources, like tissues, organs, and cancerous tumours [48]. Human cancer cell lines have been extensively studied and are widely available, providing a valuable resource for researchers to investigate cancer-related processes. During this project, cell lines were selected based on two criteria:

1. That the cell line is derived from the cancer type of one of the respective models.
2. That known drug synergies for the cell line is available. As stated previously this information is necessary to be able to compare the resulting synergies to known synergies of the cancer model.

Cell line transcriptomics data was collected from Cell Model Passports (CMP) in this project. In the world of genomics and cancer research, CMP is a hub for functional datasets and preclinical cancer models [49]. CMP currently holds over 2000 cancer cell lines, and include detailed annotations like tissue and cancer type. The platform is designed to be constantly evolving by allowing for incorporation of new models and datasets. The selected cell lines and their identifiers are displayed in Table 2.1.

Table 2.1: Identifiers of the cell lines utilised in this project [49].

Cell line	Cancer type	Tissue status	Sample site	Sanger ID
AGS	Gastric carcinoma	Tumour	Stomach	SIDM00850
HCT-116	Colorectal carcinoma	Tumour	Unknown	SIDM00783
COLO-205	Colorectal carcinoma	Metastasis	Ascites	SIDM00826
SW48	Colorectal carcinoma	Tumour	Colon	SIDM00837
SW620	Colorectal carcinoma	Metastasis	Colon	SIDM00841

All models used in this project have their own specific node annotations. In order to have a unified framework to work with, it was necessary to convert the CPM datasets from HGNC (HUGO Gene Nomenclature Committee) annotations to the specific annotations of the logical models. The *HGNC multi-symbol checker* was used as a resource for this purpose. The multi-symbol checker compares search items against all HGNC approved symbols as well as all previous, withdrawn and alias symbols. In this project, the HUGO database was used by uploading a list of model specific node names to the multi symbol checker tool. The checker then returns a list of matches for each search item, which was used to create scripts for model specific annotation conversions. Scripts for converting from HGNC annotations to model specific annotations can be accessed at the related *GitHub folder*.

Ethical Concerns of Using Biological Data

There are some ethical concerns regarding the use of biological data, such as cell lines. Amongst others, these

concerns include legal and ethical compliance, consent, and questions related to storing and using biological material [50]. Using human cell line data for research generally requires informed consent, and balancing the benefits of research with individual rights is important. An ethics statement is not necessary here, since all cell lines used in this study are publicly available at the open-access database CPM.

2.3 Part I: Creating Training Data for the DrugLogics Pipeline

This section presents the workflow from inferring biomarker activities from omics data with software tools, to a fitted training data file that can be used in the DrugLogics pipeline to calibrate logical models.

2.3.1 Selecting Inference Tools

There are many software tools that can infer biomarker activities from omics data, and their applicability can vary depending on the research question and datatype being analysed. As this project aims to identify an optimal tool for use in the DrugLogics pipeline, multiple software tools needed to be analysed. Because of this, a review of the current available tools was performed. Two approaches were selected based on this review: the 'decoupleR' software suite and the 'PROFILE' framework.

There are several reasons why these two approaches were selected. Both decoupleR and PROFILE have been extensively tested and validated in different research contexts, and are supported by software communities and ongoing development efforts. DecoupleR software suite has for instance been used in a variety of research contexts, including cancer biology [51]. The decoupleR methods was shown to be effective in inferring activities of molecular entities associated with various disease states, and has been validated using experimental data. Similarly, PROFILE has also been used in a variety of research contexts, particularly in cancer research. It was shown to be effective at identifying prognostic biomarkers and predicting treatment outcomes, and has been validated using both experimental and clinical data [52]. While there is always room for improvement and further validation, these tools are generally considered to be robust and reliable, and are likely to remain important tools for omics data analysis.

DecoupleR
DecoupleR is a an open source Bioconductor package containing a collection of statistical tools that extract biological activities by employing prior knowledge [53]. DecoupleR can be utilised to extract information from omics data, such as transcriptomics or proteomics, given that the data can be associated with a biological process that is built from known phenomena. According to Badia-I-Mompel et al. [53] such biological processes can for example be *"transcriptomics gene sets regulated by a transcription factor or in phosphoproteomics phosphosites that are targeted by a kinase."*. In that way, *regulators of the data*, such as TFs and kinases, can be deduced from their downstream transcripts and phosphosite targets, respectively. As of now, the decoupleR software suite contains 11 different statistics, including Area Under the Curve (AUCell), Fast Gene Set Enrichment Analysis (FGSEA), Gene Set Variation Analysis (GSVA), Univariate Linear Model (ULM), Multivariate Linear Model (MLM) and Weighted Sum (WSUM) [53]. In addition, the decoupleR ensemble includes a so-called CONSENSUS tool that generates a combined score when multiple methods are used. A complete list of the decoupleR tools can be found in Appendix C.

All decoupleR tools have a standard input format:

- *A matrix* of molecular readouts (for example gene expression data).
- *A network* that relates target features (for example genes and proteins) to 'source' biological entities (for example pathways, TFs and molecular processes).

In this project, cell line data was utilised as the molecular matrix and two different gene regulatory networks

(GRNs) were used to relate TFs to target genes (TGs). In addition, a 'minsize' argument can be specified according to the desired minimum target of features per biological entity. The minsize argument was sat to 5, meaning that biological entities with less target features were removed. By specifying the minsize argument potential noise from TFs with few corresponding TGs in the molecular matrix were removed. In addition, a MOR (mode of regulation) argument is available for methods that can leverage weights, and denotes whether an interaction is activating (1) or inactivating (-1).

The various tools offered by decoupleR all aim to identify activities that are driving the observed patterns in the input data. This is accomplished by exploiting the combination of prior knowledge resources and different statistical methods. By doing so, the decoupleR tools are able to infer entities (for instance TFs) that regulate the behaviour in the data, and to assign scores that reflect the degree to which they exhibit those activities. Figure 2.7 displays the general workflow of decoupleR tools.

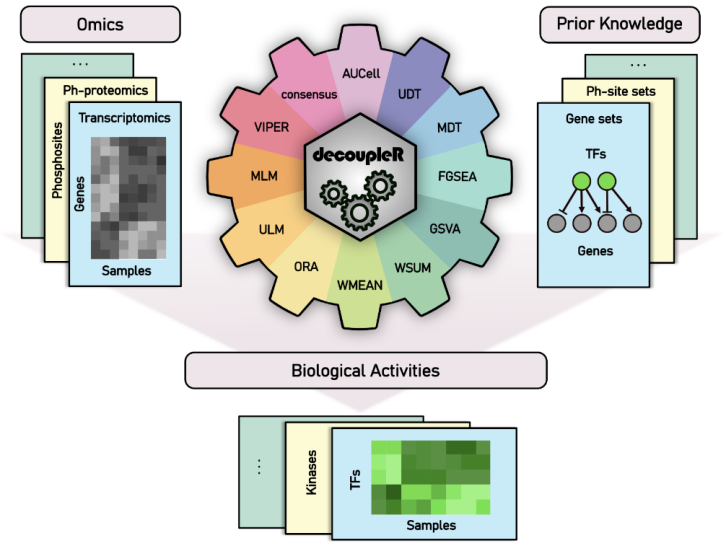


Figure 2.7: General workflow of decoupleR tools. The decoupleR software suite comprises various computational methods that, when combined with relevant prior knowledge resources, are capable of estimating the biological processes and functions reflected in omics data. The figure is retrieved from Badia-i-Mompel et al. [53].

Recent benchmarking of the decoupleR tools showed that some tools performed consistently better than others in regards to inferring biomarker activities from omics data [53]. The benchmark was designed to evaluate the accuracy, robustness, and computational efficiency of the tools. The documentation for the decoupleR benchmark is described by Badia-i-Mompel et al. [53] in *'decoupleR: Ensemble of computational methods to infer biological activities from omics data'*. According to the benchmark, most of the tools return acceptable estimates. However, the combination into a CONSENSUS score, and the use of linear models consistently outperformed the other methods. In addition, they found that methods that utilise weights perform better when those are taken into consideration. Badia-I-Mompel et al. [53] also investigated the methods speed and found that the top three performing tools ran at a satisfactory speed, allowing their use with large datasets. A list of decoupleR methods ranked by their performance in the benchmarking can be found in Appendix C. As a result of the benchmarking analysis of decoupleR tools estimate and speed performances, the overall top three performing tools, namely ULM, MLM and CONSENSUS, were selected for further analysis in this project.

Univariate Linear Model

ULM (previously known as SCIRA) is a statistical method that estimate the correlation between a *single* regulator (like a TF) and molecular features (like its TGs) [54]. This strategy enables ULM to identify the unique contributions of each TF to their TGs. To achieve this, ULM uses the prior knowledge given in the regulatory

network as input. The regulatory network contains information about TF-TG interactions, which can be either activating or inhibiting. This information is used to identify which TFs potentially regulate the expression of each gene in a data (the molecular matrix). A simplified step-by-step workflow of ULM is displayed in Figure 2.8.

For each gene in the molecular matrix, ULM looks up the corresponding information in the regulatory network to identify which TFs potentially regulate its expression (Step 1 in Figure 2.8). Once ULM has identified the TFs that potentially regulate the expression of each gene in the data (Step 2 in Figure 2.8), it needs to estimate the activity levels of these TFs. ULM uses a linear model to estimate the activity levels of each TF, given by the associated gene expressions in the molecular matrix, and the MOR of the interactions in the regulatory network. Target features with no MOR are set to zero. A simplified mathematical representation of this can be the following equation:

$$\text{Molecular Readout} = \text{Mode of Regulation} \times \text{Regulator Activity} \tag{2}$$

A positive MOR indicates that the molecular readout (expression level) of the gene is positively associated with the activity level of the regulator, while a negative coefficient indicates that the expression level of the gene is negatively associated with the activity level of the regulator. The obtained t-values from the fitted linear model is the estimated activity score of the regulator, based on the expression levels of its TGs. In this way, ULM assumes that the activity of each TF is independent of the activities of other TFs in the network, and does not take into account potential interactions between TFs. The output of ULM, and the other decoupleR tools used in this analysis, is a list of the estimated activity levels of each TF in the sample, along with their corresponding p-values (Step 3 in Figure 2.8).

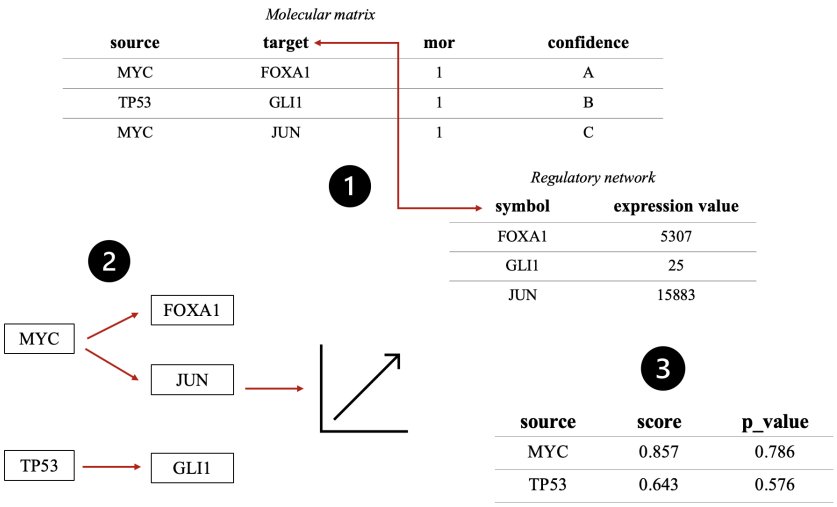


Figure 2.8: Simplified step-by-step workflow of decoupleR tools. The first step includes utilising information about regulon interactions from network resources to identify the entities that potentially regulate the expressions observed in the input data. The second step consists of estimate the activity levels of these regulators by using a statistical algorithm. This results in step three, a list of the estimated activity levels of each regulator in the sample, along with their corresponding p-values.

In summary, ULM uses a prior knowledge resource in form of a regulatory network, to identify the TFs that potentially regulate the expression of genes in a specific data sample. It then estimates the activity levels of these TFs using a linear model that takes into account the strength of the regulatory interaction between each TF and its TGs, as well as the gene expression data.

Multivariate Linear Model

Similar to ULM, MLM is also a statistical method used to identify a subset of omics features that are associated with the biological activity of a given data sample. Unlike ULM, MLM considers all regulators simultaneously in a *multivariate* linear model. The MLM workflow still follow the same steps as indicated in Figure 2.8, only that the linear regression is based on multiple TF's. By utilising multiple regression, the activity levels of multiple TFs are estimated simultaneously, by taking into account the possible interactions between the regulators and their joint influence on TGs, as well as the expression levels of the genes. Equation 2 displays a simplified mathematical representation of MLM:

$$\text{Molecular Readout} = \beta 1 \times \text{Activity of R1} \ + \beta 2 \times \text{Activity of R2} \ + \ ... \ + \beta n \times \ \text{Activity of Rn} \quad (3)$$

In this equation R represent regulators, and β denotes the the MOR's related to the relationship between the molecular readout and the regulator. MLM may provide a more comprehensive analysis of TF activities than ULM, as it considers the potential interdependence between multiple TFs and their TGs. This allows for a more comprehensive analysis of interconnected regulatory networks. This is particularly important in cases when one TF may activate another TF, or two TFs may cooperate to regulate a set of TGs. By considering all regulators jointly, MLM can provide a more accurate representation of complex regulatory networks.

CONSENSUS
CONSENSUS is an ensemble method that combines the results of the top-performing statistics in decoupleR (MLM, ULM, and WSUM) to generate a final list of potential biomarkers [53]. The idea behind CONSENSUS is to obtain a more robust set of biomarker activities by using the strengths of several methods.

To calculate a 'consensus score' for each biomarker, the activity scores obtained from the top-performing methods are first standardised using a double-tailed z-score transformation. A z-score is a statistical measure that represents the number of standard deviations (SDs) an individual data point is away from the mean of a population [55]. The first step of a double-tailed z-score transformation is to calculate the z-score for each activity score by subtracting the population mean from the raw score, and dividing the result by the population SD. This gives a standardised z-score representing how many SDs the raw score is away from the population mean. Next, the z-scores are transformed using a new mean and SD. This is done by calculating a new z-score for each standardised score using the formula:

$$\text{New z-score} = (\text{old z-score} - \text{current mean}\ )/\ \text{current SD} \times \text{new SD} + \text{new mean} \quad (4)$$

In this equation, the 'old z-score' refers to the standardised score calculated in the previous step, and the 'current mean' and 'current SD' refer to the mean and SD of the old z-score. The new mean and SD refer to the desired mean and SD of the transformed scores. In a double-tailed z-score transformation, the new mean and SD are chosen such that the resulting z-scores are symmetrical around zero. This means that both positive and negative z-scores will have the same magnitude, and the distribution will be centred around the new mean. The purpose of this transformation is to standardise the data and make it easier to compare scores across different variables or populations. After the double-tailed z-score transformation is performed, the CONSENSUS score for each biomarker is calculated as the mean of the transformed activity scores.

PROFILE
PROFILE is a pipeline that tailors logical models to biological measurements, like a patient tumour [52]. Similar to the DrugLogics pipeline, PROFILE aims to use logical modelling as a tool for precision medicine, and facilitate better clinical choices for patient-specific drug treatments. PROFILE is developed to handle high-dimensional, heterogeneous omics data that is often encountered in clinical studies. Béal et al. [52] demonstrates the effectiveness of PROFILE at identifying biomarkers for patients with acute myeloid leukemia. They identified a set of biomarkers that is associated with poor overall survival, including genes involved cell cycle regulation. In addition, they identified a set of biomarkers that is associated with response to specific

therapies, such as genes involved in DNA repair.

PROFILE is a comprehensive tool that provides all the necessary steps from inferring biomarker activities to personalising logical models. Since the goal of this project was to utilise the DrugLogics pipeline for personalising logical models and predicting synergies, only the part of PROFILE that relates to inferring patient profiles (biomarker activities) was employed.

PROFILE can infer activity levels from various data types, such as mutation data, copy number alterations (CNA), transcriptomics and proteomics from the biological system of interest (for example patient data or cell lines). In this project, PROFILE was only utilised to infer biomarkers from RNA sequencing data taken from cancer cell lines. In PROFILE, activity levels are obtained by either binarising or normalising the data. To achieve this, the gene expression data is initially divided into three categories based on their distribution across the samples: bimodal, unimodal, or zero-inflated distribution. Genes are treated differently according to the distribution they have. In short, a unimodal distribution means that the data has only one peak, while a biomodal distribution has two peaks. On the other hand, zero-inflated refers to a distribution that allows for frequent zero valued observations (the data has excess of zero counts).

Figure 2.9 displays the process of classifying genes as bimodal, unimodal or zero-inflated in PROFILE. To classify the genes, an admissibility test is first conducted [52]. This test checks if the gene expression values are sufficiently variable, and contain enough non-zero values. If a gene fails the test, it is filtered out. The remaining genes are classified as bimodal, unimodal, or zero-inflated based on their distribution patterns using three statistical techniques: the dip test, bimodality index, and kurtosis criteria. In this context the term 'distribution pattern' refers to the genes expression value in the cancer data compared to a reference expression dataset. The reference expression dataset was in this project a cohort of other cell lines originating from the same cancer type. A gene is only classified as bimodal if it meets the criteria for all three tests. The full documentation for the classification of genes based on their distribution pattern is described by Béal et al. [52].

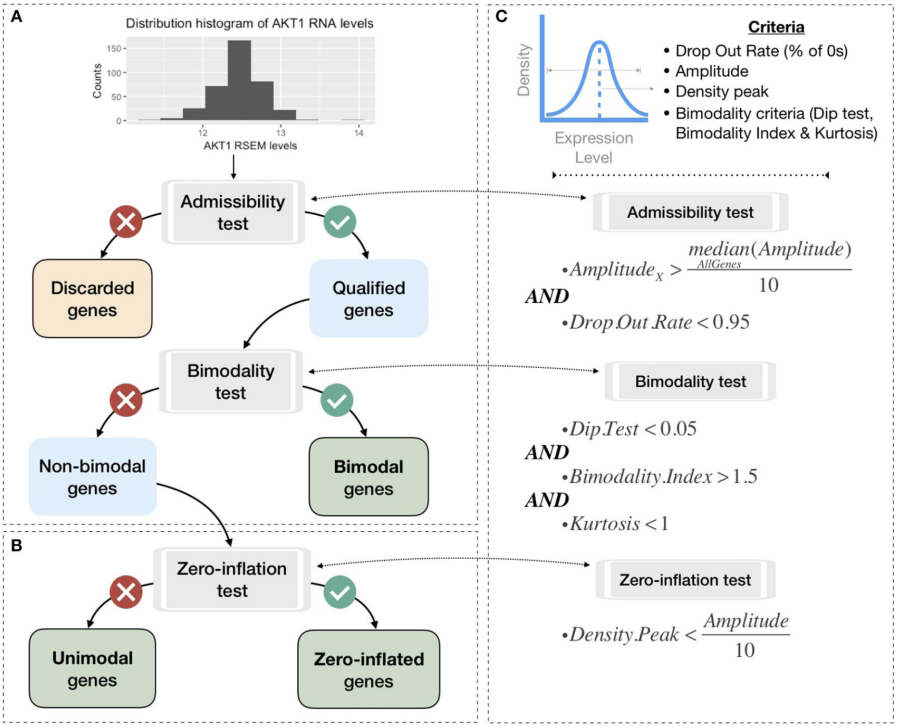


Figure 2.9: Overview of the gene classification process in PROFILE. Panel A displays the workflow of the admissibly test the bimodality test to classify genes as either bimodal or non-bimodal. Panel C displays the zero-inflation test to classify genes as either unimodal or zero-inflated. The figure is retrieved from Béal et al. [52].

Genes with bimodal distributions are binarised, while genes with unimodal or zero-inflated distributions are normalised, as displayed in Figure 2.10. PROFILE can also provide binarised values for more genes that those with a bimodal distribution. In that case, bimodal genes are processed using a Gaussian mixture model and binarised based on the probability of belonging to one of the modes. Unimodal genes are normalised using a sigmoid function, while zero-inflated genes are linearly transformed to maintain the asymmetric original pattern. The resulting binarised and normalised gene expressions can consequently be utilised as biomarker activity levels in downstream analysis.

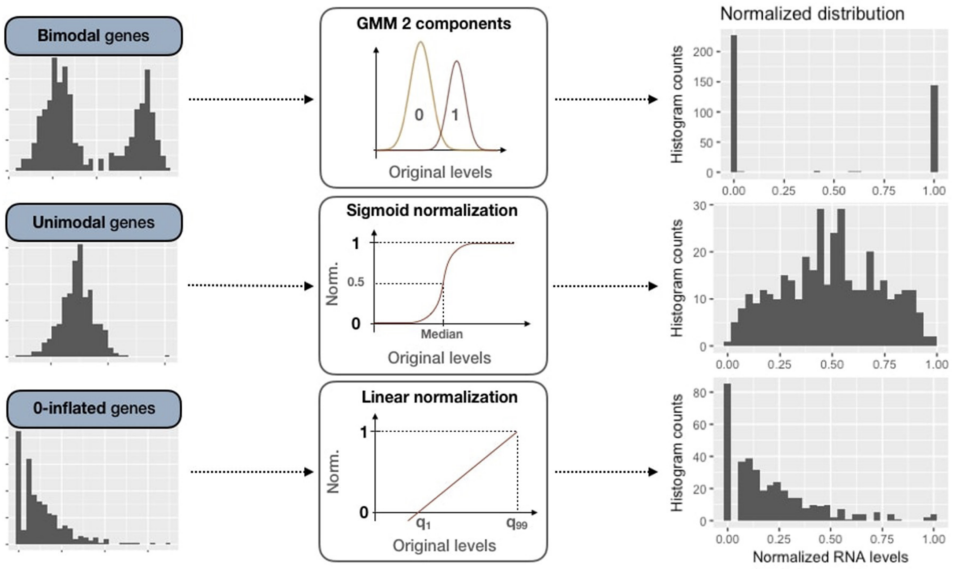


Figure 2.10: Normalisation techniques in PROFILE applied to expression data of genes belonging to three distinct categories (bimodal, unimodal, and zero-inflated). The initial column panels display original gene distributions representing each category. The subsequent column demonstrate the normalisation methods employed for each distribution, and the third column shows the resulting normalised distributions. The figure is retrieved from Béal et al. [52].

2.3.2 Selecting the Tool Parameters

Within each software tool, there are a variety of parameters that can be specified. In this project, multiple combinations of parameters were analysed for each tool, aiming to identify some optimal parameters to ensure accurate biomarker inference.

Gene Regulatory Networks

For all decoupleR tools it was necessary to use a GRN. Two GRN networks were used in this project, namely DoRothEA and an early version of CollecTRI.

DoRothEA

DoRothEA is a comprehensive GRN that is constructed based on prior knowledge of regulatory interactions between TFs and their TGs [56]. TF-TG interactions (regulons) are the fundamental building blocks of GRNs, and they play a critical role in determining the activity levels of genes and the biological processes that they control.

DoRothEA was constructed using a systematic approach that integrated diverse types of genomics data, including chromatin accessibility, TF binding, and gene expression data to infer regulatory interactions. The resulting network contains over 400 000 regulatory interactions among 1395 TFs and over 20 000 TGs in humans, as can bee seen from the *package documentation*. DoRothEA provides a valuable resource for investigating the regulatory mechanisms underlying complex biological processes, such as cancer. An important feature of

DoRothEA is that it provides a confidence score for each regulatory interaction, which reflects the strength of evidence supporting the interaction. The confidence score is based on a statistical model that accounts for the reliability and consistency of the various types of functional genomics data used to infer the interaction. The confidence score can be used to prioritise regulatory interactions for further experimental validation, or to filter out less reliable once. The confidence levels are categorised in five levels from A (highest confidence) to E (lowest confidence). Interactions that are supported by at least four lines of evidence, including literature-curated resources, ChIP-seq interactions, gene expression, or TF binding motifs, are considered highly reliable and are assigned an A level. Interactions curated by experts in specific reviews or that are supported by at least two curated resources are also assigned an A level. Level B-D are reserved for curated and/or ChIP-seq interactions with different levels of additional evidence. Finally, level E is used for interactions that are uniquely supported by computational predictions.

CollecTRI
CollecTRI was used as an alternative to DoRothEA in this project to see if using different GRNs with the decoupleR tools affected the biomarker inference results, and subsequently the synergy results in the DrugLogics pipeline. An early version of CollecTRI was utilised in this project, as the final version of CollecTRI was not established of the point of the analysis. The main difference of the early and the released version of CollecTRI is the assigned signs (MORs) of the different interactions.

The final version of CollecTRI is a GRN containing signed TF-TG interactions for 1183 TFs [57]. CollecTRI was created to improve the accuracy of estimating TF activities by expanding the coverage of regulons from high-confidence prior knowledge. Müller-Dott et al. [57] notes that while there are several methods available for inferring TF activities, these methods rely on prior knowledge of the TGs that are regulated by each TF. However, the existing knowledge of regulons is often incomplete or inaccurate, which can limit the accuracy of TF activity estimation. CollecTRI was created to overcome this limitation, by incorporating multiple resources like public databases, text mining and manual curation. Specifically, Müller-Dott et al. combined existing regulon knowledge with gene co-expression networks and gene ontology term annotations to identify additional TGs that are likely to be regulated by each TF. CollecTRI was experimentally tested, in which it significantly improved the accuracy of TF activity estimation compared to other networks like DoRothEA.

Expression Count Measures
Expression count measures are quantitative measures of the gene expression levels in a biological sample. Three different RNA sequencing expression count measures are given for cell lines from CMP: read count data, FPKM data and TPM data.

Read Counts
Read count (RC) measurements is the simplest approach to quantifying gene expression by RNA sequencing. In this measurement, the number of raw reads that align to a particular gene is counted and used as a measure of the level of gene expression. The basic idea behind read count quantification is that the more reads that align to a gene, the higher is the expression of that gene. Transcript lengths are not considered in raw read counts [58].

FPKM Values
Fragments per kilobase of transcript per million mapped reads (FPKM) is a commonly used normalisation method in RNA sequencing experiments to quantify gene expression levels [58]. To calculate FPKM values, the first step is to determine the total number of reads in a sample. This count is then divided by a million to obtain the 'per million' scaling factor. To further account for differences in gene length, the counts are divided by the length of the corresponding gene in kilobases. This division gives the FPKM value, representing the expression level of a gene normalised for both sequencing depth and gene length.

TPM Values
Transcripts per million (TPM) is also used as an RNA sequencing measure to quantify gene expression levels [58]. TPM values are very similar to FPKM values, but TPM normalises the read counts by gene length first,

yielding reads per kilobase (RPK), and then divides the RPK values by the 'per million' scaling factor. One notable advantage of TPM is that the sum of TPM values for all genes within a sample is always the same. This property ensures that TPM values represent the proportion of transcripts contributed by each gene, making it easier to compare the relative expression levels within and between samples.

P-value Thresholds
The decoupleR tools return a list of inferred TFs and their corresponding activity levels. This 'TF - activity level' couple comes with an associated p-value. The p-value provides an estimate of the statistical significance of the obtained activity score. Three different p-value thresholds were selected for analysis in this project: 0.05, 0.2 and no threshold.

Output Classification
As mentioned earlier, PROFILE provides both normalised and binarised biomarker profiles. The outputs were analysed separately in this project. In addition, an analysis of binarising the obtained normalised output were experimented with.

2.3.3 Inferring and Integrating Biomarker Activities into the DrugLogics Pipeline: A Workflow Explanation

For ease of use, scripts that automated the process of inferring and integrating biomarker activities seamlessly into the DrugLogics Pipeline were created for each of the four software tools. The tool-specific scripts can be accessed at the related *GitHub folder*. Due to the multiple combinations of parameters utilised, a total of 60 different training data files were generated per dataset (cell line).

A brief description of this process is given in this section. A detailed step-by-step explanation of the tool-specific scripts, and the workflow from omics data to a formatted training data file can be found in Appendix D. All tools were carried out in the statistical programming language R, which is frequently used for quantitative analysis because of its convenient use with data importing and visualisation [59].

DecoupleR Tools Workflow
The decoupleR tools were downloaded via the decoupleR package, which was in this project was installed with *Bioconductor*. ULM, MLM and CONSENSUS were used with the same procedure to infer biomarker activities from omics data. In short, this process involved first importing and reading the necessary inputs: the molecular matrix and the regulatory network. The specific decoupleR tool was selected for and executed with the selected network, resulting in a list of inferred TF activity levels. The inferred activity levels were subsequently binarised, and the TF annotations were converted from HGNC annotations to the annotations in the respective logical model utilised. Finally, the output was formatted to match the format in the DrugLogics pipeline, using a specialised function created in this project.

PROFILE Workflow
As mentioned previously, only the part of PROFILE that relates to extraction of cell line activity levels was utilised in this project. The files necessary to exceute PROFILE was downloaded from the *PROFILE repository* . PROFILE was generally executed by following the related *tutorial*. This included utilising a script from the PROFILE repository that relates to extracting activity levels, with minor changes. This script takes several inputs to infer activities, and the once relevant for this project are:

- *Omics data* in HGNC annotations.
- *A model file* describing the conversion from HGNC annotations to the specific annotations in the logical model.

When downloading the PROFILE repository, the input files related to omics data for numerous cell lines are

present. All cell lines analysed in this project were part of the data files that were already present in the repository. As a result, in order to run PROFILE with the selected logical models the only task that was necessary to do was to create the model file describing the conversion from HGNC annotations to the specific annotations in the model. Creating the model files comprised of making a tab-separated file with two columns, one with HUGO annotations, and one corresponding column with the annotations in the specific model network. Separate model files were created for each of the CASCADE 1.0, Lu and Park cancer models. The output of PROFILE was also formatted to match the format in the DrugLogics pipeline by using the specialised function.

2.4 PART II: Prediction of Drug Synergies With the DrugLogics Pipeline

This section presents the workflow of using the inference tool-generated training data files with the DrugLogics pipeline to generate synergy predictions.

2.4.1 Running the DrugLogics Pipeline

The DrugLogics pipeline can generally be executed by following the *tutorial for synergy prediction*. The DrugLogics pipeline was executed both with utilising a Docker image and by using Java and Maven. However, the pipeline needs to be executed separately for each training data file. Given that the number of training data files generated in this project is in the thousands, a script was generated to automate the execution of the DrugLogics pipeline. The script facilitates to automatically run numerous training data files subsequently in a loop, without the need for excessive manual labour. The automated scripts are made available on *GitHub*.

In addition, running the DrugLogics pipeline demands a lot of processing power of a computer. Depending of the size and complexity of the logical model utilised, running the pipeline on a regular computer can take up to several hours just for one simulation. Utilising a more powerful processor makes it possible to decrease the compilation time of the pipeline dramatically. Given the extent of this analysis, it was desirable to utilise an external server for this purpose. An external server is a computer system that may grant access to users over the internet. The NTNU SSB1 server was used in this project by connecting with Secure SHell (SSH) keys, and utilising terminal commands. This facilitated to run the pipeline on the external SSB1 server from an internal computer, and to move directories to and from the server.

The process of running the DrugLogics pipeline on the SS1 server can be summarised by the following steps:

1. Moving the necessary files for running the DrugLogics pipeline from an internal computer to the SSB1 server, by using terminal commands.
2. Connecting to the SSB1 server using SSH keys.
3. Executing the DrugLogics pipeline on the SSB1 server using the automated script.
4. Moving the directories with the pipeline outputs from the SSB1 server to an internal computer, by using terminal commands.

This workflow facilitated the possibility of running the DrugLogics pipeline with numerous training data files at minimal cost of manual labour and internal processing power, while significantly reducing the compilation time. As a result, it was possible to investigate a large collection of tools, parameters, cell lines and cancer models in this analysis.

2.4.2 Calculating AUC ROC Values

AUC (Area Under the Curve) ROC (Receiver Operating Characteristic) plots was used as the statistical measure of the synergy results in this project. AUC ROC curves are commonly used to evaluate the performance of binary classifiers [60]. A ROC curve is a plot of the true positive rate (TPR) versus the false positive rate (FPR) at various classification thresholds. The TPR is the proportion of positive cases that are correctly identified by the model, while the FPR is the proportion of negative cases that are incorrectly classified as positive. Essentially, the AUC ROC curve will represent the trade-off between sensitivity (TPR) and specificity (1-FPR). A perfect classifier would have a TPR of 1 and an FPR of 0, resulting in a point at the upper left corner of the AUC ROC plot. Figure 2.11 displays an example of a ROC curve with the calculated AUC.

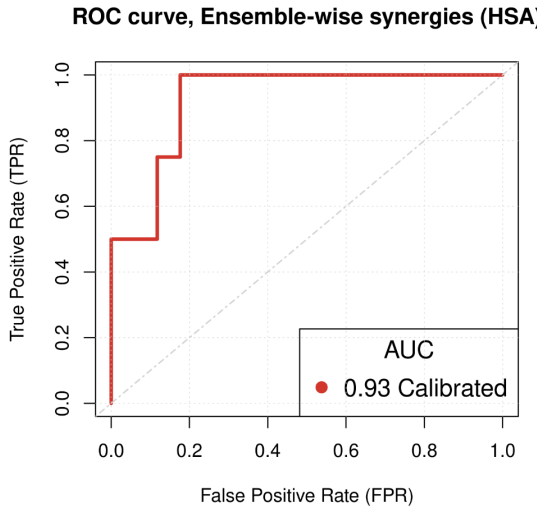


Figure 2.11: Example of a ROC curve with the calculated AUC. Figure retrieved from Flobak et al. [30] with CASCADE 1.0, HSA synergy method, and 150 models calibrated to the AGS cell line steady state.

A random classifier, on the other hand, would (in theory) have a diagonal ROC curve with an AUC of 0.5, indicating that its performance is no better than chance. In general, a higher AUC indicates better classifier performance, with an AUC of 1 indicating perfect classification.

AUC ROC values are commonly used by the DrugLogics group to evaluate synergy predictions, due to its convenience with binary input. When calculating AUC ROC values, the obtained synergy results from Drabme are compared to drug synergies that are known to be true, so-called 'gold standard drug synergies'. AUC ROC values may be calculated from the ensemble-wise synergies file created by Drabme, by comparing the scores of the identified synergies to the gold standard synergies. As mentioned earlier, the ensemble-wise synergies file is constructed by comparing the growth value of the collection of Boolean models from Gitsbe when they are perturbed by *pairwise* combinations of drugs, to the growth value when they are perturbed with *single* drugs. If the growth of a pairwise combination is lower than the additive effect of two single drugs, that is an indication of a drug synergy. The more negative the value is, the stronger the synergistic drug combination. When calculating AUC ROCs, the synergy values in the ensemble-wise synergies file are compared to the gold standard synergies as a TPR-FPR trade-off. This ultimately means that the higher the AUC ROC value, the better the model's ability to distinguish between synergistic and non-synergistic drug combinations (according to the gold standard synergies).

Automated scripts to interpret hundreds of synergy results simultaneously and calculate subsequent AUC ROC values were generated in this project, and can be accessed at *GitHub*.

Relating Training Data Quality to AUC ROC Values

It is the collection of logical models that are used to calculate the growth value in Drabme that links the quality

of the training data to the the resulting AUC ROC values. As mentioned earlier, the fitness of the logical models in Gitsbe are directly calculated based on the information in the training files. Models will obtain a high fitness if they capture the biology specified in the training data in a precise way. High fitness models will proceed to Drabme, and be used to calculate the growth value. As a result, the precision of the inferred biomarker activities in the training data will affect the precision of the generated models, and consequently the drug synergy predictions and the resulting AUC ROC values.

2.5 Part III: Statistical Analysis of Synergy Results

This section presents the workflow of a statistical evaluation of the synergy results. In order to evaluate the significance of the synergy results, and thus the inference performance of the selected software tools, a statistical analysis was conducted. The goal was to investigate if the training data generated by the inference tools statistically improved the synergy predictions over randomly calibrated models. In order to do this, bootstrapping of random calibration data (training data) was selected as a basis of comparison.

Bootstrapping is a statistical technique used to estimate the accuracy of an estimate or a test statistic [61]. The method involves creating multiple re-sampled datasets by randomly sampling with replacement from the original dataset. The resampled datasets will have the same size as the original dataset, and can be used to generate multiple estimates of the same statistic. Figure 2.12 displays the workflow of the bootstrapping process conducted in this project.

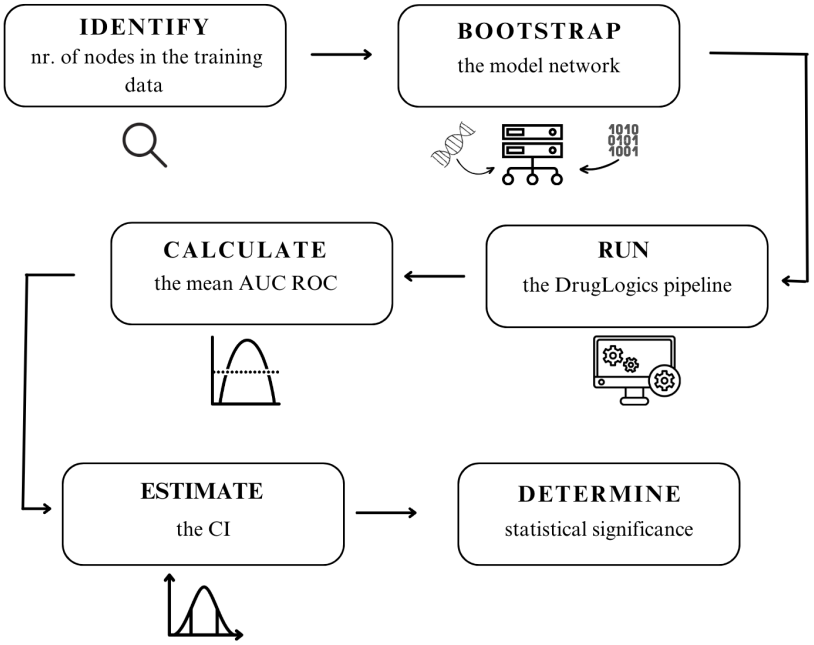


Figure 2.12: Simplified workflow of the bootstrapping process conduced in this project. The first step consisted of identifying the number of nodes in the inference result of each tool, and then generate a hundred random training files for the same number of nodes. Next, the DrugLogics pipeline was executed with the bootstrapped training files and the mean AUC ROC value was calculated for each bootstrapping and used as the population parameter since it may represent an 'average' random synergy result. 95% confidence interval was subsequently computed of the bootstrapped mean, and used to determine the statistical significance of the obtained synergy results.

The bootstrapping was conducted for each model by randomly and independently sampling nodes from the respective model network. The random training data was designed so that the amount of random activities matched the amount of actual biomarker activities of the synergy result that was to be statistically ana-

lysed. This means that the number of nodes in every synergy result had to be identified, and a bootstrapping sampling of that respective model network had to be conducted for *the same number of nodes*. The bootstrapping sampling was then conducted 100 times. Next, the expression value of each bootstrapped node would randomly be selected as 0 or 1. This resulted in a hundred bootstrapped training files that would represent a collection of random training data. An automated script was created for the bootstrapping process, and can be found at *GitHub*. The DrugLogics pipeline was then utilised with the hundred bootstrapped training data files to generate a representation of random synergy predictions.

Further on, the bootstrapping results was used to estimate confidence intervals (CI) with R. CIs are a range of values that are expected to contain the true value of a population parameter with a certain level of confidence. In this project, the mean AUC ROC of the random calibration data was used as the population parameter since it may represent an 'average' random synergy result. The 95% CI was subsequently computed of the bootstrapped mean, and would represent an interval were if the bootstrapping was re-did, the new mean would be within that interval (with a 95% confidence). If one assumes that the bootstrapped mean represents a random synergy result, then whether the synergy result at question is within that CI would tell if this result is likely to be obtained at random. If the synergy result is outside of the 95% CI, it may be defended that it is less than 5% likely that the result would be obtained by using random calibration data. Whether that synergy results is significantly better or worse than the bootstrapped results depend on if the result is above the upper limit of the CI or below the lower limit, respectively. On the other hand, if the synergy result is within the CI, there is a probability above 5% of obtaining that result with random training data, and the null-hypothesis will not be rejected. In this case, the null hypothesis will be that the synergy result is generated at random. As a result, if the synergy results is above the CI, that would mean that the hypothesis that the synergy result could be generated by using random calibration data could be rejected with a 95% confidence.

3 Results and Discussion

The Results and Discussion section of this report presents the findings of a systems biology analysis that evaluated the efficacy of four software tools, namely ULM, MLM, CONSENSUS, and PROFILE, in inferring biomarker activities from omics data for logical model calibration. As outlined in Materials and Methods, this analysis was conducted using a five-step approach:

1. Infer biomarker activities using the software tools
2. Fit activity data to the format of the DrugLogics pipeline
3. Predict drug synergies with the DrugLogics pipeline
4. Calculate AUC ROC values
5. Conduct a statistical analysis

The findings of this project are presented and analysed in three main sections, focused on inference outcomes, synergy predictions, and the statistical analysis. The insights from these three parts are integrated to discuss the feasibility of identifying an optimal approach for biomarker activity inference to calibrate logical models in the DrugLogics pipeline. Due to the complexity of the analysis, the discussion is focused on results related to the gastric adenocarcinoma cell line (AGS) CASCADE 1.0 model. This model was chosen as the primary focus due to its reliability and robustness, having been extensively curated, analysed and utilised in the DrugLogics pipeline previously. These factors make CASCADE 1.0 a suitable candidate for a comprehensive in-depth analysis. The results related to the Lu and Park models with their associated cell lines were utilised to investigate whether the findings of tool performances (and tool parameters) could be generalised across network topologies and dataset characteristics. The synergy results and statistical analysis related to the Lu and Park models are included in Appendix E.

3.1 PART I: Creating Training Data for the DrugLogics Pipeline

This section aims to identify and discuss observed trends and differences in inferred biomarker activities. The discussion is in turn focused on general trends across tools, within tools and with respect to tool parameters. The insights gained from this analysis highlights the strengths and limitations of the tools' inference methodologies, and are used as a foundation for explaining the synergy results in section 3.2.

3.1.1 Gold Standard Biomarker Activities

A set of 'gold standard' biomarker activities for the AGS gastric cancer (GC) cell line were established as a basis of comparison for the obtained inference results (Table 3.1). In this context, 'gold standard' biomarker activities relates to Boolean activity levels that may be expected in the AGS cell line.

The gold standard biomarker activities were constructed based on the findings of Flobak et al. [30]. Their findings included a manually curated analysis of 219 observations of activity levels in AGS steady state from 72 papers. These gold standard activities are only defined for nodes in the CASCADE 1.0 network. In addition, nodes with less than three independent reports on their activity levels were excluded. In instances where conflicting activity levels were encountered, the activity value with the highest level of scientific support was selected. Reports on conflicting findings and less sustained activity observations of Flobak et al. [30] can be found in Appendix F.

Table 3.1: Boolean gold standard activity levels that can be expected to be inferred from AGS cell line data based on an AGS steady state manually curated analysis by Flobak et al. [30].

Biomarker	Boolean activity value
MYC	1
TP53	0
TCF7_f	1
NFκB_f	1
AKT_f	1
ERK_f	1
MMP_f	1
JNK_f	0
BCL2	1
GSK3_f	0
KRAS	1
S6K_f	1
BAX	0
CASP3	0
PIK3CA	1
PTEN_g	0
RAC_f	1
CCND1	1
CTNNB1	1
CASP8	0
MAPK14	0

The underlying reason for these expected activity levels may be explained by the biological function of the components in the AGS signalling network. For instance, the expected active state of the transcription factor (TF) MYC may be attributed to its critical role in cell cycle progression, apoptosis, and cellular transformation. Overexpression of MYC is found to contribute to GC development by activating a cascade of downstream signalling pathways, which are involved in promoting cell growth and division, inhibiting apoptosis and enhancing angiogenesis. In fact, overexpression of MYC has been observed in over 40% of GC incidents, and its activation has been associated with poor prognosis and aggressive tumour behaviour [62] [63]. TP53, however, was found to be inactive in AGS by Flobak et al. [30]. This may be explained in that TP53 is a tumour suppressor gene which plays an important role in maintaining genome stability by regulating the cell cycle, promoting apoptosis and triggering the repair of damaged DNA [64]. Because of this, TP53 is known as the 'guardian of the genome' [65].

Another example is CTNNB1 and TCF7_f (TCF7 family) which are TFs that were mostly found to be active in AGS by Flobak et al. [30]. This may for instance be explained by their relation to the Wnt/β signalling pathway. As mentioned in the Introduction, aberrant activation of the Wnt/β signalling pathway is a known characteristic of many types of cancers, including GC [6]. CTNNB1 and TCF7_f are both key components needed in active states to maintain the functionality of this pathway [66] [67].

3.1.2 General Inference Trends

To highlight the differences in inferred activity levels by various tools, it is necessary to first analyse general trends across them. Figure 3.1 displays inferred biomarker activities from the AGS cell line by each software tool. Each column represents a different set of inferred biomarker activities obtained from using a unique set of tool parameters. As the intent is to first provide a global assessment across tools, the parameters are not visualised in this figure. Tool parameters are discussed in section 3.1.4.

As observed in Figure 3.1, PROFILE infers a significantly larger number of biomarker activities compared to

the decoupleR tools, which at most inferred a total of seven activities. The reason for this is that the decoupleR tools are only capable of inferring regulator activities, which in the case of transcriptomics are TFs, such as MYC, TP53, FOXO_f, TCF7_f, EGR1 CTNNB1, and NFκB_f. PROFILE, in contrast, returns activity levels of both TFs, enzymes, receptors, genes, peptides and so on. As a result, the decoupleR tools are only capable of inferring a subset of the activities that PROFILE returns.

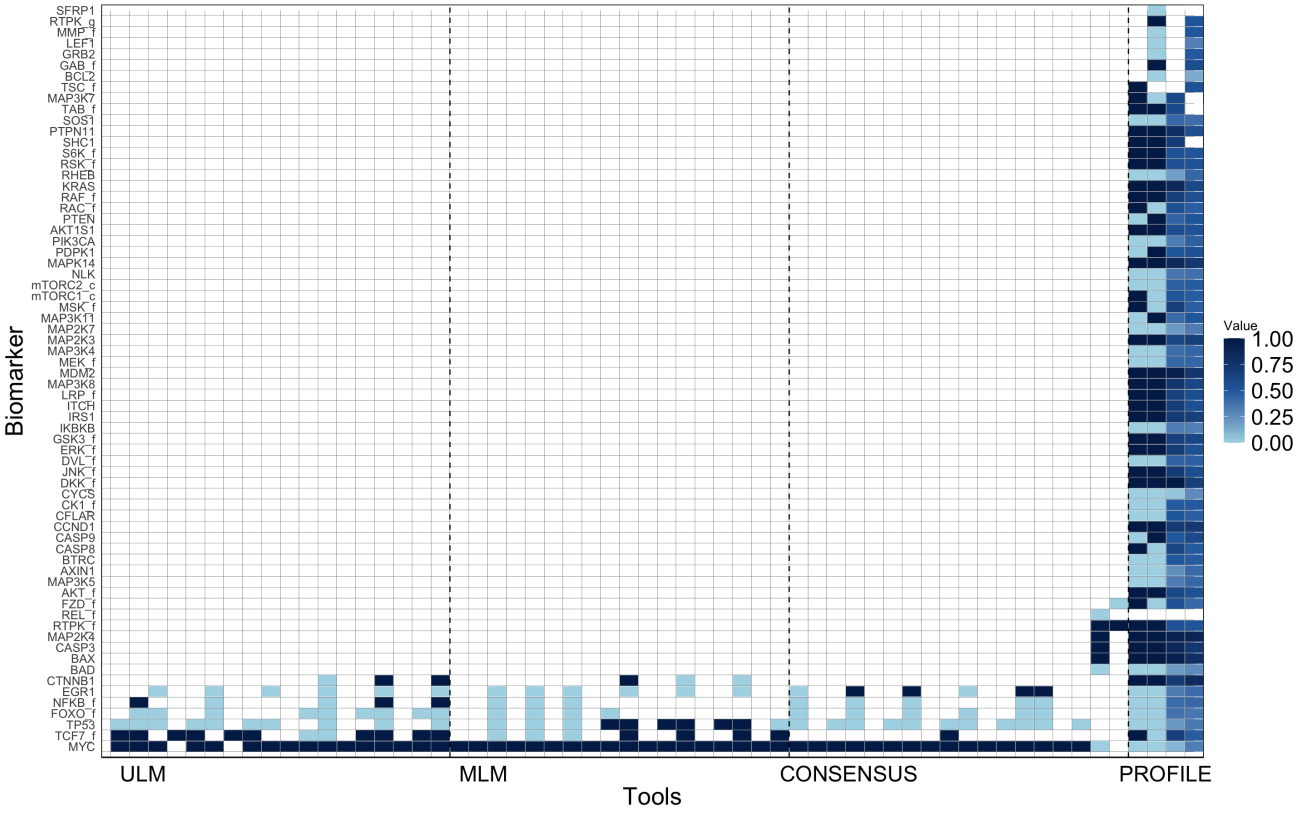


Figure 3.1: Inferred biomarker activities from AGS cell line transcriptomics. Biomarker activities were inferred using the bioinformatic tools ULM, MLM, CONSENSUS and PROFILE with different combinations of parameters. Each column represents a different set of tool parameters. The tool parameters include different regulatory networks, expression count measures, p-value thresholds and output classifications. Colours indicate inferred Boolean activity values (1=dark, 0=light).

Another noticeable trend is that the tools consistently infer the same activity values for particular nodes, while the activities of other nodes differ more across tools. For instance, the inference of active MYC is observed by almost all tools, which is in compliance with its gold standard activity (Table 3.1). TP53 was mainly expected to be inferred in an inactive state from the AGS cell line, and although this is generally true, it is inferred as active by some tools.

FOXO_f, which is short for 'FOXO family', was also one of the biomarkers inferred with a consistent activity level my multiple tools. FOXO_f was not included by Flobak et al. [30] as a gold standard due to limited papers reporting its activity level in AGS. However, one paper suggested its possible active state in AGS, as seen from the less sustained observations in Appendix F. However, other studies report that FOXO factors can function as tumour suppressors by promoting cell cycle arrest, apoptosis and DNA repair. Aberrant modulation of FOXO factors leading to their inactive state has also specifically been linked to the development and progression of gastric adenocarcinoma [68] [69] [70]. Further supporting this, a study from 2018 specifically linked the presence of FOXO3, a member of the FOXO family, to the inhibition of AGS cell growth [71]. Consequently, it seems like the consistent inference of inactive FOXO_f may be reasonable, although its link to AGS is less reported on.

Similarly, EGR1 is inferred as inactive by several tools. EGR1 is also part of the less sustained observations

of Flobak et al. [30], with one paper from 2004 suggesting its inactivity in AGS. Reviewing more recent literature further support this. EGR1 is a TF involved in cellular regeneration, angiogenesis, cell growth and programmed cell death [72]. Ko et al. [72] found that proliferation of AGS cells could be inhibited through a signalling pathway involving EGR1. They found that EGR1 promoted the upregulation of molecules involved in inducing apoptosis, such as the p21 gene, whose function lies in the induction of cell cycle arrest. Through these downstream genes, EGR1 inhibition was linked to inhibiting apoptosis in AGS. The study of Kim et al. [73] further support this. As a result, it seems like the inference of inactive EGR1 from multiple tools may be expected from AGS data. However, there were also some instances where active EGR1 was inferred.

Finally, NFκB, CTNNB1 and TCF_7 were expected to be inferred in active states, and for the most part, this is observed (Figure 3.1). However, some tools inferred these TFs in inactive states from the AGS cell line. Except for one paper suggesting findings of active NFκB in AGS [74], there was not found any supporting literature of these observations in AGS.

These deviating results may be caused by the different algorithms and parameters of the software tools, which is discussed in section 3.1.3 and 3.1.4. The contradictory findings of biomarker activities also brings to light the challenges of investigating them. It is important to approach biomarker inference results from software workflows with caution, as there may be multiple intertwined factors causing them. For instance, as our understanding of regulatory interactions enhance through research, the functional and regulatory roles of network components becomes better understood. This improved understanding results in better characterisation and description of certain components, for example in network resources like DoRothEA and CollecTRI. As mentioned in Materials and Methods, only regulons with a confidence level of C or higher were considered in this analysis, meaning that TF-target gene (TG) interactions with low confidences were discarded. As the decoupleR tools utilise these networks to infer TF activities, they may be better equipped at inferring accurate activity levels of some TFs if all of their interactions in the regulatory networks are accurate and have high confidence levels. Consequently, as more knowledge is gained about a particular regulator, its true activity level may be more likely to be consistently identified by software tools due to its well-characterised and understood functions in biological networks. As a result, it is important to note that a skewed information bias towards some TFs may lead to a more consistently correct inference rate of these TFs compared to less understood once.

3.1.3 Comparing Tool-Specific Inference Results

In this section, the focus is to analyse biomarker activities specific to each tool. The aim of this analysis is to highlight the underlying methodologies and statistical algorithms of each tool, and their impact on biomarker inference. Additionally, this section aims to increase confidence in using certain tools to infer biomarker activities to calibrate logical models in the DrugLogics pipeline.

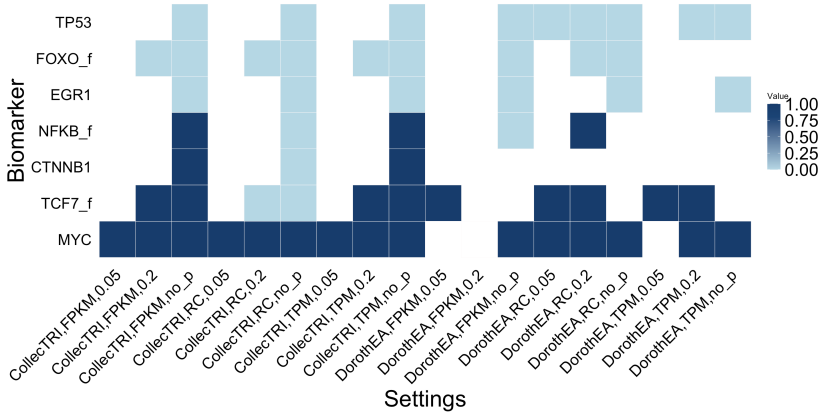
To strictly analyse the tools, the inference results of the decoupleR tools are analysed within the same combinations of parameters, and the PROFILE results are analysed independently. Figure 3.2 and Figure 3.3 provides a detailed view of the tool-specific inference results from AGS transcriptomics.

DecoupleR tools

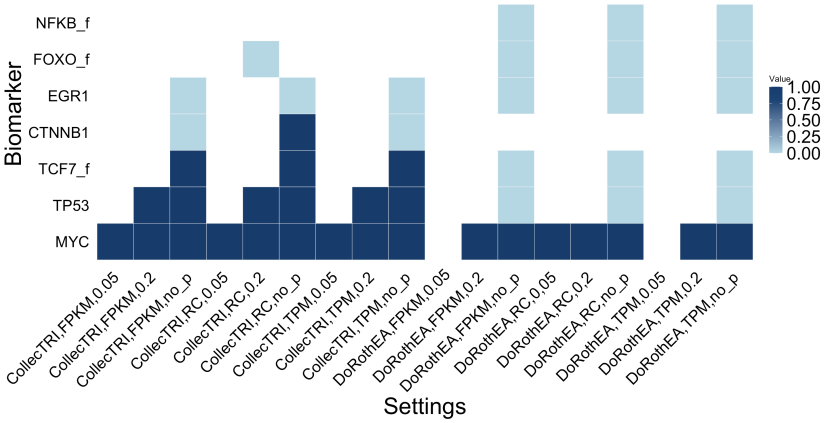
ULM and MLM

As previously outlined, the methodologies employed by ULM and MLM are distinct, despite being based on linear models. ULM examines each regulator's activity individually to estimate the correlation between a *single* TF and the TGs in a sample. Conversely, MLM takes into account the possible interactions between TFs and their *joint* influence on TGs. These differences in methodologies may cause differing sensitivities to the threshold used to identify a TF as a regulator of the genes in the data. As ULM considers the activity of each regulator independently, it might be more sensitive to detect small variations in the data of individual genes. In contrast, by considering the interactions between multiple regulators and their joint impact, MLM may be

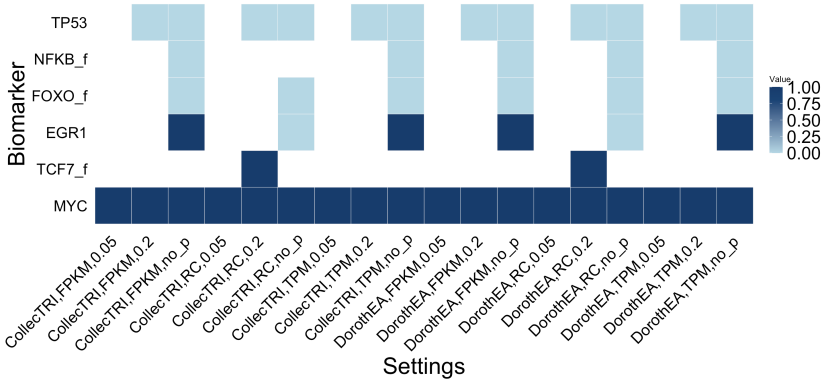
more robust against small variations in the data. To infer a TF, MLM may require substantial variations in the genes expression values. This difference in sensitivity could let ULM infer more TFs than MLM under the same circumstances.



(a) ULM



(b) MLM



(c) CONSENSUS

Figure 3.2: Tool-specific inference results of decoupleR tools from AGS transcriptomics. The inference results are divided into tool-specific panels, and each column represents a different set of tool parameters. The tool parameters include different regulatory networks, expression count measures and p-value thresholds. The colours indicate the inferred Boolean activity values (1=dark, 0=light).

The suggestion that ULM may infer more TF activities than MLM is an evident trend when analysing Figure 3.2 (a) and (b). For instance, with the parameters DoRothEA, read count (RC) and a p-value threshold of 0.05 and 0.2, ULM identified three (MYC:1, TCF7_f:1, TP53:0) and five TF activities (MYC:1, TCF7_f:1 and TP53:0, FOXO_f:0, NFκB_f:1), respectively. MLM identified only one TF (MYC:1) in both cases. However,

it is important to note that this trend does not always hold true, as MLM produces comparable or even higher numbers of inferred activities than ULM in some cases, for instance with the parameters 'DoRothEA, FPKM, no p-value'. Several factors can contribute to this variability. For example, in situations where the expression of a particular TG is regulated by the combined effect of multiple regulators, MLM might be better equipped to identify those complex interactions than ULM. As a result, the number of inferred activities by ULM and MLM might differ depending on the expression data and the complexity of the regulatory network used.

When comparing the inference results of ULM and MLM to the gold standard biomarkers (Table 3.1) is seems as ULM mainly infers biomarkers in accordance to the gold standard activities. In fact, ULMs deviations from the gold standards are only observed a few times (two times for NFκB_f and TCF7_f and one time for CTNNB1). MLM in contrast, infers both NFκB_f, TCF7_f, CTNNB1 as inactive and TP53 as active several times. Neither of these results are in accordance with the gold standards. NFκB_f:0, TCF7_f:0 and TP53:0 are all inferred three times each and CTNNB1:0 is inferred two times. In addition, both ULM and MLM identifies inactive FOXO_f and EGR1. As mentioned earlier, these findings may be expected in AGS, although their activities in the cell line is not thoroughly documented. Nevertheless, one potential limitation of both MLM and ULM is that they assume a linear relationship between the omics features and the biological activity. In reality, this relationship may be more complex, which may cause the biologically incorrect inference results.

To conclude, it seems like the overall deviation of MLM from the gold standard activities of AGS is more pronounced compared to ULM. These results might be caused if the dataset contains small variations, or that the TGs in the data are mostly regulated by single TFs, as ULM might more equipped to asses these precisely.

CONSENSUS
As mentioned in Materials and Methods, CONSENSUS combines the results of the top-performing decoupleR tools (ULM, MLM and WSUM), by using a double-tailed z-score transformation. By combining the different methodologies of ULM and MLM, both single and joint relations may be taken into consideration. Thus, this approach can lead to even more robust and reliable results. In fact, most of the biomarker activities identified by CONSENSUS (Figure 3.2 (c)) are in compliance with the expected gold standards. However, CONSENSUS identifies inactive NFκB_f and active EGR1 as regulators (when eliminating the p-value thresholds), which is is not expected in the AGS cell line. This is further discussed in section 3.1.4. Nevertheless, CONSENSUS is mostly able to infer expected biomarker activities according to the gold standards. In addition, when combining the results from multiple tools, the inferred biomarker activities from CONSENSUS remain relatively limited in comparison to other methods.

However, the inference results of CONSENSUS may introduce a level of complexity that can be difficult to interpret in a biological way, especially if the results from different workflows are contradictory. For example, if one tool identifies a TF as highly active in a given sample, while another tool identifies the same TF as having low or no activity in the same sample, it can be challenging to reconcile these seemingly contradictory results to a combined score in a biological meaningful way. Therefore, it is important to carefully consider the limitations and potential sources of variability when using the CONSENSUS approach, and to be cautious when interpreting the biological meaning of its results.

PROFILE
All inferred biomarker activities of PROFILE are not analysed here, but some links are drawn to compare the result to the decoupleR tools. Similar to the decoupleR tools, FOXO_f, TP53, EGR1 and NFκB_f is inferred as rather inactive by PROFILE (Figure 3.3). MYC, however, is also inferred as inactive by PROFILE, which deviates from decoupleR results and its gold standard activity (Table F).

In contrast, PROFILE identifies some none-TF gold standard activities that the decoupleR tools are not able of inferring, namely AKT_f, ERK_f, MMP_f, JNK_f, GSK3_f, S6K_f, PTEN_g, RAC_f and CASP8. BCL2 and KRAS were inferred by PROFILE at relatively low and high expression values, respectively. This is in compliance with the gold standard activity for KRAS, but not for BCL2. The rest of the these biomarkers were inferred with normalised activity levels close to 0.5, making them difficult to interpret due to the Boolean nature

of the gold standards. This highlights one of the limitation of utilising Boolean values to describe biological processes, which is discussed further in section 3.5.5.

To conclude, it seems like PROFILE overall is able to capture a broader set of components with known relations to the AGS cell line than the decoupleR tools. However, the continuous activity levels of some of these components are difficult to interpret in a biological meaningful way. In addition, PROFILE identifies a large number of entities in which there was not found experimental observations on in AGS (Appendix F). While these nodes may be of importance, they may also add noise in the model calibration.

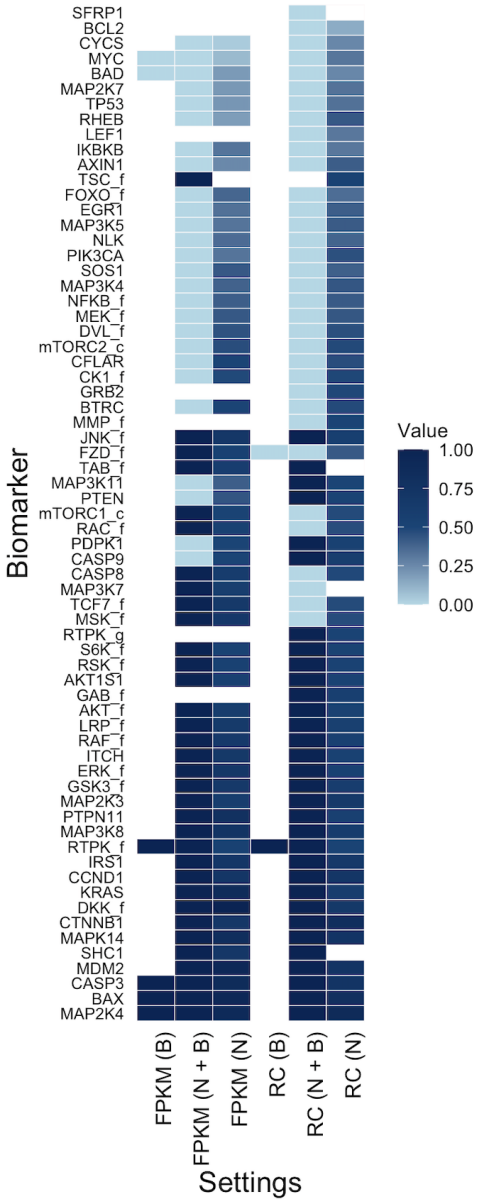


Figure 3.3: Inferred biomarker activities of PROFILE from AGS cell line transcriptomics. Each column represents a different set of tool parameters. The tool parameters include different expression count measures and output classifications. Colours indicate inferred Boolean activity values (1=dark, 0=light). The letter B denotes binarised output classification and the letter N denotes normalised output classification.

3.1.4 Comparing Parameter-Specific Inference Results

The differences in inferred biomarker activities with respect to tool parameters are examined in this section. This analysis offers additional insights into underlying factors contributing to the observed inference activities, and highlights possible optimal parameters for each tool.

DecoupleR Parameters
The three different decoupleR parameters, namely GRNs, expression count measures and p-value thresholds are identified and discussed in this section. The decoupleR tools differ in the degree they are affected by parameters. Figure 3.2 highlights that ULM exhibits a greater variance in inferred activity levels across different parameters than both MLM and CONSENSUS. This increased variance may be due to ULM single regulator estimating approach, making it more sensitive to detect variations resulting from different expression data and regulatory networks. By considering complex relationships, ULM and CONSENSUS may be more robust towards these variations, with CONSENSUS being the overall least affected by differing parameters.

Gene Regulatory Networks
Reviewing the inference results reveals that utilising different GRNs significantly impacts the inferred activities of some tools. Figure 3.2 (a) displays an example of this, where for 'TPM, 0.2', the inferred biomarker activities with DoRothEA are MYC:1 TCF7_f:1 TP53:0, while for CollecTRI they are MYC:1, TCF7_f:1 and FOXO_f:0. While there are instances where the use of both networks infer identical biomarker activities, the inferred TFs seem to significantly vary with different networks for ULM and MLM (with greater p-values).

As mentioned earlier, ULM inferred opposite expression values of TCF7_f than what is expected by its gold standard activity (Table 3.1). This incident is exclusively the case when ULM used CollecTRI as a GRN. However, for the same incident in MLM, it is exclusively observed with DoRothEA. In contrast, TP53's activity level was incorrectly inferred by MLM with CollecTRI. NFκB's activity level was incorrectly inferred compared to the gold standards by ULM with both GRNs. In CONSENSUS, the results are almost identical with CollecTRI and DoRothEA. Based on the variability of these results, concluding with an optimal GRN for ULM and MLM cannot clearly be accomplished based on this analysis. The effect of GRNs is further discussed in light of the synergy results in section 3.2.3.

The obtained differences resulting from GRNs may be explained by their different information bases. CollecTRI and DecoupleR differ in their sources and quality of prior knowledge that they use. DoRothEA's human regulons with confidence levels of C or higher consists of about 13 000 TF-TG interactions. In contrast, CollecTRI is triple the size and has increased coverage of the TF-TG interactions obtained from combining different resources. Therefore, depending on the specific characteristics of the data and the tools used for inference, the different GRNs may contribute to the identification of different sets of biomarker activities. However, the use of different regulatory networks does not seem to affect the inference results of CONSENSUS. In fact, its inference result with DoRothEA is almost identical to the inference results with CollecTRI with the same parameters. As mentioned earlier, this might be caused by the fact that CONSENSUS is a more robust tool in terms of sensitivity to small changes in the data, by combining the results obtained from several tools. However, in the benchmark of Müller-Dott et al. [57] the inference results of CONSENSUS was consistently better than when using other resources like DoRothEA. This is further discussed in section 3.2.3.

Expression Count Measures
As described in Materials and Methods, RCs represents the raw number of reads mapped to a specific gene, while FPKM (Fragments Per Kilobase of transcript per Million mapped reads) and TPM (Transcripts Per Million) are both normalisation methods. Some patterns can be observed from Figure 3.2, with respect to these three parameters. First, it seems that using FPKM, TPM or RC does not have a dramatic impact on the inferred biomarker activities for most tools. For example, CONSENSUS and ULM consistently infer MYC:1, TP53:0 and FOXO_f:0 regardless of whether FPKM, TPM or RC was used. There are a few exceptions to this trend. For example, CONSENSUS only infers TCF7_f when RC is used. As mentioned earlier, active TCF7_f is

expected in the AGS cell line, so it seems like it is missed when FPKM and TPM expression values are used. Overall, using RC data seems to have a more significant impact on the inferred activities compared to FPKM or TPM. For example, with CollecTRI ULM infers TCF7_f, NFκB_f and CTNNB1 as active with FPKM and TPM, but they are all inferred as inactive with RC. From a biological point of view, these biomarker activities are expected to be active in AGS (Table 3.1). This is also true in MLM and CONSENSUS with CTNNB1 and EGR1, respectively, although the literature findings on EGR1 in AGS were not very strong.

In addition, when comparing the number of inferred biomarker activities of the decoupleR tools, using FPKM and TPM data yield slightly lower numbers *in a few cases* compared to using RC data. This is not a evident trend, but in the cases where the numbers differ, the highest number is inferred with RC five out of eight times. This can be attributed to the fact that both FPKM and TPM normalisation methods take into account the differences in gene length, which RC does not. From a technical point of view, normalising the data can decrease the range of expression values, making the variations in the dataset smaller, thus making the biomarkers harder to detect for some tools, like MLM. This may explain the slightly higher inference rate and the generally larger deviation from the expected biomarker activities when raw RC are utilised compared to FPKM and TPM.

Overall, it appears that the choice of input data, particularly between FPKM and TPM, does not highly affect the inferred activities for most tools. However, the choice of RC versus FPKM and TPM seems to have a more significant impact on the inferred activities, particularly for ULM and MLM. However, it is important to note that while there seems to be some differences in the inference based on the different expression values, these differences are relatively minor, and most tools infer activities relatively consistent across these parameters.

P-value Thresholds
Different p-values cause different biomarker activity inference, as is evident when looking at Figure 3.2. When the p-value threshold is eliminated, more biomarkers may be identified. However, this may also potentially increase the number of incorrect inferences. When analysing the inference results with respect to the gold standard biomarkers (Table 3.1) this seems to be the case. Cutting off the p-value threshold increased the rate of deviations from the expected inference activities for both ULM, MLM and CONSENSUS compared to having a strict p-value threshold. Conversely, using a more stringent threshold (0.05 or 0.2) results in fewer biomarker activities detected, but with higher confidences. As the p-values are estimates of the confidence of an inferred biomarker activity, it is expected that allowing higher p-values increases the rate of incorrect inferences.

PROFILE parameters
In this subsection, the differences in inference results when using the different parameters of PROFILE, namely expression count measures and output classification is identified and discussed.

Expression Count Measures
Same as for the decoupleR tools, there can be observed some differences in the inference of biomarker activities when using FPKM and RC with PROFILE (Figure 3.3). This is especially evident with the binarised output where, except for one node (RTPK_f), different nodes are inferred with FPKM and RC values. Six nodes were inferred when using FPKM data and only two were inferred when using RC data. Together with the decoupleR inference results, these results suggest that the choice of expression count measure can impact the inferred biomarker activities, likely due to the normalisation method affecting gene expression levels.

When looking for differences in the normalised output, PROFILE infers lower levels of MYC, TP53 and CTNNB1 expressions when using FPKM data than with RC data. TCF_7 and CASP3 on the other hand, is inferred at a higher expression rate when FPKM is used compared to RC. MMP_f and BCL2 are not inferred with FPKM but are inferred with very low expression values with RC. Comparing these results to the expected inference results, there is no evident trend in where the faulty inference trends are observed (Table 3.1). As a result, it is not possible to conclude with an optimal expression count measure for PROFILE at this point.

Output Classification
Looking at PROFILE's inference results it seems like most of the genes fall within the categories of unimodal or zero-inflated distributions (Figure 3.3), as the number of biomarker activities in the normalised output are highest. As a much larger number of biomarker activities are normalised than binarised, one can observe a much higher rate of expected biomarker activities in the normalised training data. Consequently, the rate of biomarker activities which deviates from the gold standard activities, and entities with no experimental observations in AGS is also higher in the normalised output.

Looking at the binarised output of PROFILE (Figure 3.3), except for MYC, there are no similarities in the inferred nodes of PROFILE and the the nodes inferred by the decoupleR tools. MYC is also inferred as inactive by PROFILE, which according to the gold standard biomarkers is not expected in AGS (Table 3.1). This is also the case for both BAX, CASP3 and MAP2K4 which are inferred as active but expected to be inactive according to the observations of Flobak et al. (Appendix F). In contrast, BAD is inferred in accordance to the less sustained observations. As for the remaining nodes in the binarised output, no evidence was found by Flobak et al. [30] supporting their relation to the AGS cell line (Appendix F). Overall, it seems like both the normalised an binarised output of PROFILE introduces some noise and unknown activities. The effect of training data sizes and deviations of PROFILE is further discussed in section 3.2.3.

3.2 PART II: Prediction of Drug Synergies With the DrugLogics Pipeline

The second part of the results in this project relates to the generation of synergy predictions, using the inferred biomarker activities as training data in the DrugLogics pipeline. Each training data file generated a unique synergy prediction, resulting in a corresponding unique AUC ROC value. The obtained AUC ROC values range from 0 to 1, with a score of 0 indicating that no correct synergy predictions were found, and 1 indicating a perfect synergy prediction, in accordance to a set of predefined gold standard synergies [30].

Trends and differences in AUC ROC results are identified and discussed in this section. The discussion is in turn focused on general trends across tools, within tools and with respect to tool parameters. The previous findings of inference results in section 3.2 are used to explain observed patterns and deviations. The goal is to investigate if biomarker activities inferred by specific tools perform better than others as calibration data to tailor logical models in the DrugLogics pipeline.

3.2.1 General Trends in Synergy Predictions

General trends in synergy predictions, using AUC ROC values, are addressed in this section. Table 3.2 displays a small extract of the inferred biomarker activities from the AGS cell line, and corresponding AUC ROC values using the CASCADE 1.0 model. All synergy results related to the AGS cell line can be found Table G.1 in Appendix G.

Table 3.2: An extract of inferred biomarker activities from the AGS cell line, and corresponding AUC ROC values from the CASCADE 1.0 model with the DrugLogics pipeline.

Tool	Inferred biomarker activities	AUC ROC
CONSENSUS	MYC:1, TP53:0	0.82
CONSENSUS	EGR1:0, FOXO f:0, MYC:1, TP53:0	0.77
ULM	MYC:1, TCF7_f:1, FOXO_f:0	0.72
CONSENSUS	MYC:1	0.68
ULM	EGR1:0, FOXO_f:0, MYC:1, NFKB_f:0, TP53:0	0.66
MLM	MYC:1, TP53:1	0.47
CONSENSUS	EGR1:1, FOXO_f:0, MYC:1, TP53:0, NFKB_f:0	0.29

Looking at Table 3.2 it is evident that the inference of specific biomarker activities results in high AUC ROC values. For instance, the combination of MYC:1 and TP53:0 appears with the high AUC ROC value of 0.82. This suggests that these two biomarker activities may be useful in calibrating logical models for predicting accurate synergies for AGS. When MYC:1 is inferred alone the resulting AUC ROC value drops to 0.68, so it seems as the presence of TP53:0 contributes to calibrating more accurate AGS models. This is also highlighted by the fact that when MYC:1 and the biologically incorrect TP53:1 are inferred, the AUC ROC value is significantly lower (0.47). As mentioned previously, these biomarker activities are expected in AGS, so given that the logical model captures the biology of AGS, it makes sense that inference of these biomarker activities results in high AUC ROC values. This is also evident when looking at the CASCADE 1.0 model network (Figure 2.4), as MYC is directly connected to Prosurvival with a positive link. Active TP53 in contrast, contributes to Antisurvival through a series of links, and hence the inactivation of TP53 would be expected to contribute to AGS in the network model.

Another general observation from the synergy results is that the addition of biologically incorrect biomarker activities may decrease the AUC ROC values. This can be seen from the combination of EGR1:1, FOXO_f:0, MYC:1, NFκB_f:0 and TP53:0 which result in an AUC ROC value of 0.29. Knowing that MYC:1 and TP53:0 alone results in a much higher AUC ROC value, it seems like the presence of EGR1:1, FOXO_f:0 and NFκB_f:0 contributes to decreasing the precision of the synergy predictions. As discussed previously, a strong connection of *active* NFκB_f in AGS activity was found, so from a biological point of view the presence of NFκB:0 in the training data results in inaccurate AGS models, and hence, a lower AUC ROC value. This result also make sense in regards to the topology of CASCADE 1.0, as NFκB_f contributes, through series of different nodes, to the inactivation of Antisurvival. In addition, the active inference of EGR1 also deviates from some scientific findings.

Overall, the general trends highlights that it can be even more important to prioritise exclusion of incorrect biomarkers compared to prioritise inclusion of correct biomarkers, when it comes to constructing models that can accurately predict synergies in the Gitsbe module of the DrugLogics pipeline. This is for example evident from the fact that the AUC ROC result connected to MYC:1 increased by 0.15 when adding the biologically correct inference value of TP53 (TP53:0), but decreased by 0.21 when adding the incorrect TP53:1. In addition, there seems to be some biomarker activities that are more detrimental to have in the training data file in order to obtain high AUC ROC values. For instance, when EGR1:0, FOXO f:0, MYC:1, TP53:0 are inferred, the AUC ROC value is 0.77, and it decreases with as little as 0.10 when the biologically incorrect biomarker activity of NFKB_f is added. This may suggest that NFKB_f does not plays a vital role in the CASCADE 1.0 network, as its incorrect inference did not have detrimental outcomes on the AUC ROC value.

3.2.2 Trends in Tool Performances

In this section, the performances of ULM, MLM, CONSENSUS and PROFILE is highlighted and discussed based on the obtained AUC ROC values. Figure 3.4 displays synergy predictions in form of AUC ROC values resulting from each training file with biomarker inferences from the AGS cell line. As seen from Figure 3.4, the tool performances vary depending on parameter combinations. However, in this section the focus is on discussing general trends in each tool, by comparing AUC ROC values within the same parameter combinations. Synergy results is discussed with respect to the tool-specific parameters in section 3.2.3.

Comparing the performances of the decoupleR tools with the same parameters reveals that CONSENSUS generally outperforms both ULM and MLM. This trend is evident across multiple parameters, including different normalisation methods, regulatory networks and p-value thresholds. This suggests that CONSENSUS may be the most effective inference tool overall, as it is able to combine multiple sources of information to achieve a robust collection of biomarker activities. In fact, within the 18 different combinations of decoupleR parameters tested, CONSENSUS achieves the highest synergy performance in six of them, and results are tied with other tools in five more. However, it is worth noting that in several cases the improvement in performance achieved

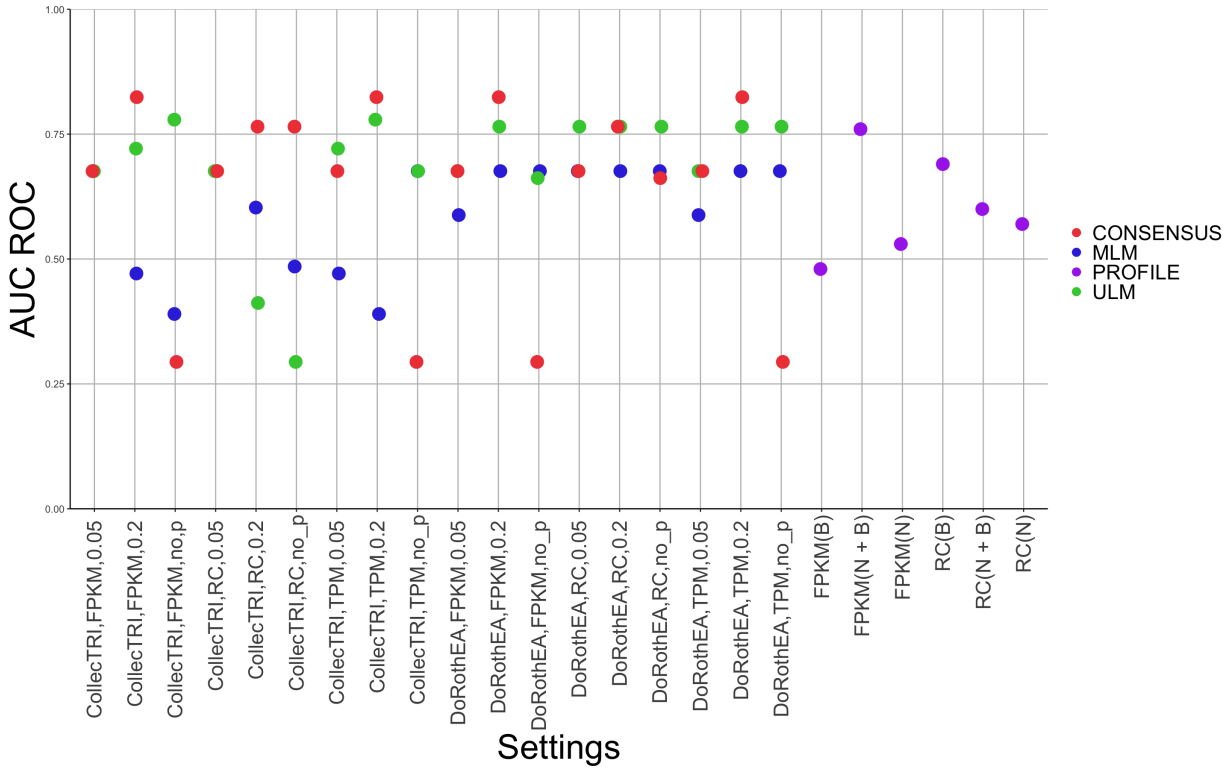


Figure 3.4: Combined synergy results in form of AUC ROC values obtained from using AGS CASCADE 1.0 with the DrugLogics pipeline. Each column represents a different set of parameters. The tool parameters include different regulatory networks, expression count measures, p-value thresholds and output classifications. The colours in the figure refer to the respective software tool utilised for biomarker activity inference.

by CONSENSUS is relatively modest compared to ULMs.

Another trend in tool performances observed from Figure 3.4 is that the AUC ROC values of ULM are generally higher than those of MLM. In fact, ULM receives a higher value than MLM in 13 out of 18 combinations of parameters. This is especially evident with the 'CollecTRI,FPKM,no_p' parameters, where ULM has an AUC ROC value close to 0.8 while MLMs is close to 0.4. Similarly, in 'CollecTRI,RC,0.2', ULM has an AUC ROC value of 0.41 while MLM has an AUC ROC value of 0.60. In DoRothEA with FPKM and $p < 0.2$, ULM has an AUC ROC value of 0.77 while the one of MLM is 0.68.

To explain the observed patterns in tool performances, it is necessary to revisit the inference results. As discussed earlier, CONSENSUS was found as the most robust tool in terms of inferring biomarker activities that are biologically expected from the AGS cell line, and to be least affected by the differing parameters. As a result, it may not be surprising that the synergy results of CONSENSUS are consistently high as well. Except for four cases where the p-value is eliminated (which is discussed later in section 3.2.3), all CONSENSUS results are within the range of 0.66 to 0.82. Interestingly, so are almost all the results of ULM (except for two cases: 'CollecTRI,RC,0.2' and 'CollecTRI,RC,no_p'). This highlight that while CONSENSUS outperforms ULM in almost all cases, the inference results from both tools culminates in relatively consistent and high synergy predictions. Now, the reason why the inference results ULM outperforms MLM, and why CONSENSUS outperforms both ULM and MLM in the DrugLogics pipeline may lie in the tool's ability to infer a conservative but precise collection of biomarker activities. As discussed previously, CONSENSUS and MLM are more conservative inference approaches than ULM in terms of number of inferred activities. However, this might be attributed to higher validity in the more conservative inference methods, where there are fewer incorrectly inferred biomarker activities. This relates to the way the training data is used to calibrate logical models in the Gitsbe algorithm. In this context, a precise biomarker refers to two things: 1) it has a strong association to the

function of the biological network that is modelled, and 2) the correct activity value is inferred.

As mentioned earlier, it is the collection of Boolean models that are used to calculate the growth value in Drabme that links the quality of the training data to the the AUC ROC values. In the genetic algorithm of Gitsbe, the fitness score of each individual model is calculated by comparing the output of the model with the expected values in the training data. High fitness models are passed to Drabme and are used to calculate the growth value. As a result, the precision of inferred biomarker activities in the training data affects the precision of the generated models in Gitsbe, and consequently the drug synergy predictions from Drabme and the resulting AUC ROC values.

Both missing and imprecise biomarker activities may affect the precision of the generated models in Gitsbe. This may be explained by taking a closer look at the effect missing and imprecise activities in the training data have in the progression of the Gitsbe algorithm. If the training data contains biologically *imprecise* information, such as MYC:0, the fitness score of the model may be misleading. This is because the model is being trained to fit to biologically incorrect information, which can lead to the generation of models with high fitness scores that are not biologically representative of the disease state. This may result in the generation of a model collection that proceeds to Drabme with high fitness scores, but low representation of the biology in the cancer network. In contrast, if a biologically important biomarker activity is *missing* from the training data, for example the absence of MYC:1, the resulting models will not properly account for its effects, which also can lead to biologically defective models. Specifically, the fitness scores will be calculated based on incomplete information about the cancer system, and the resulting models may not accurately reflect the true dynamics of the cancer in the presence of the missing biomarker. In both cases, a defective model collection is used to calculate growth values in Drabme. This may in turn cause lack of, or incorrect drug synergies which ultimately results in lower AUC ROC values when comparing these to the gold standard drug synergies. In the case of a restricted biomarker collection, there will most likely be less imprecise biomarker activities compared to a larger inference collection. However, there will also most likely be missing important biomarker activities from the training data, compared to a larger collection. From the obtained AUC ROC values it seems like ULM overall scores higher with respect to this trade-off than MLM, and that CONSENSUS scores overall higher than both. This might not be surprising as MLM was found to infer smaller collections of biomarkers, and deviate the most from the expected inference results, while CONSENSUS was found to infer a small collection of biomarkers and deviate the least. ULM on the other hand, generally inferred the most biomarkers of all, but had lower deviation rates than MLM.

However, there are cases where it may be more difficult to explain the obtained AUC ROC values based on biological interpretation of the inference results. For example in the case of 'CollecTRI, FPKM, 0.2' CONSENSUS infer MYC:1 and TP53:0 (which as discussed earlier are expected in AGS) with the AUC ROC value of 0.82. With the same parameters MLM infers MYC:1 and TP53:1 with an AUC ROC of 0.47. In this case, it seems like the presence of active TP53 dramatically lowered the AUC ROC value, which makes sense as it is expected to be inferred as inactive in AGS, as can also be observed from the CASCADE network structure (Figure 2.4). However, the inference results from ULM with these parameters is harder to explain in a biological meaningful way. ULM infers MYC:1, TCF7_f:1 and FOXO_f:0, with an AUC ROC of 0.72 (Table 3.2). Based on background knowledge provided in section 3.2 it is known that FOXO_f inactivation, MYC activation and TCF7_f activation can be expected in AGS function. So how come the AUC ROC value of ULM is lower then of CONSENSUS, if both tools inferred biologically precise biomarkers? This question needs to be addressed with respect to the workflow of the DrugLogics pipeline.

It could be speculated that the obtained AUC ROC value is within a reasonable uncertainty, as the synergy predictions from the DrugLogics pipeline are based on a number of evolutions of model mutations. The higher the evolution number (up to a plateau, beyond which no improvement is seen), the more cycles of mutations the models undergo to fit more accurately to the observations in the training data. As a result, it is possible that the number of evolutions may have limited the models ability in fitting to the training data in a more precise way, and that increasing the number of evolutions might result in a higher AUC ROC value. To investigate

this, the number of evolutions was increased from 50 (the default) to 500. For the specific case of MYC:1, TCF7_f:1 and FOXO_f:0 this yielded an AUC ROC value of 0.65, which is even lower than before. A possible explanation for this may be that the evolution process of mutations in Gitsbe is less constrained when lower numbers of biomarkers are fixed in the training data. The evolutionary algorithms may explore some additional trajectories in the mutation phase when it is not penalised for having multiple activity levels fixed. Further supporting this is the findings of the master project by Thea Hettasch (*'Effect of Calibration Data Subsetting on Boolean Model Calibration and Drug Synergy Predictions'*, May 2023), which indicates that there may be a prevalence within the DrugLogics pipeline for smaller collections of biomarker activities. Other factors that may affect the obtained synergy results is discussed in section 3.5.

PROFILE, in contrast, has the lowest overall AUC ROC values out of all four tools. This indicates that it may not be as effective as the other tools at identifying precise and restricted biomarker sets. However, there are a few instances where PROFILE performs relatively well, such as with FPKM where the normalised output is consecutively binarised. The reason why the binarised version of the normalised output yields a higher AUC ROC vale might simply be because the binarisation process shifted more inference values in the 'right' direction according to the gold standards. However, this might just be a 'lucky case', as the opposite results are observed with RC. As previously stated, the inference results of PROFILE are quite different to the inference results of decoupleR. To give a short recap, the binarised results contained very few biomarkers as expected and the normalised results contain a lot of expected biomarkers (according to the gold standards), but also a lot of unexpected once. As a result, it may not be surprising that the synergy results of PROFILE are consistently lower than of the decoupleR tools, as the trade-of of inferring a restricted but precise activity collection seem to be shifted with either low precision or too much noise. The possible prevalence of the DrugLogics pipeline towards smaller training data collections may also provide an explanation for the low AUC ROC results of PROFILE. However, it is worth noting that PROFILE was only evaluated in a limited number of parameters, so its performance may differ in other cases.

Overall, CONSENSUS appears to have the overall highest performance in inferring biomarker activities from the AGS transcriptomics to calibrate logical models in the DrugLogics pipeline. MLM and ULM have relatively consistent performances, but ULM generally outperforms MLM. PROFILE has lower AUC ROC values overall, but performs relatively well with some specific parameters. However, the performance of all tools are highly dependent on the specific parameters used, which is discussed in the following section.

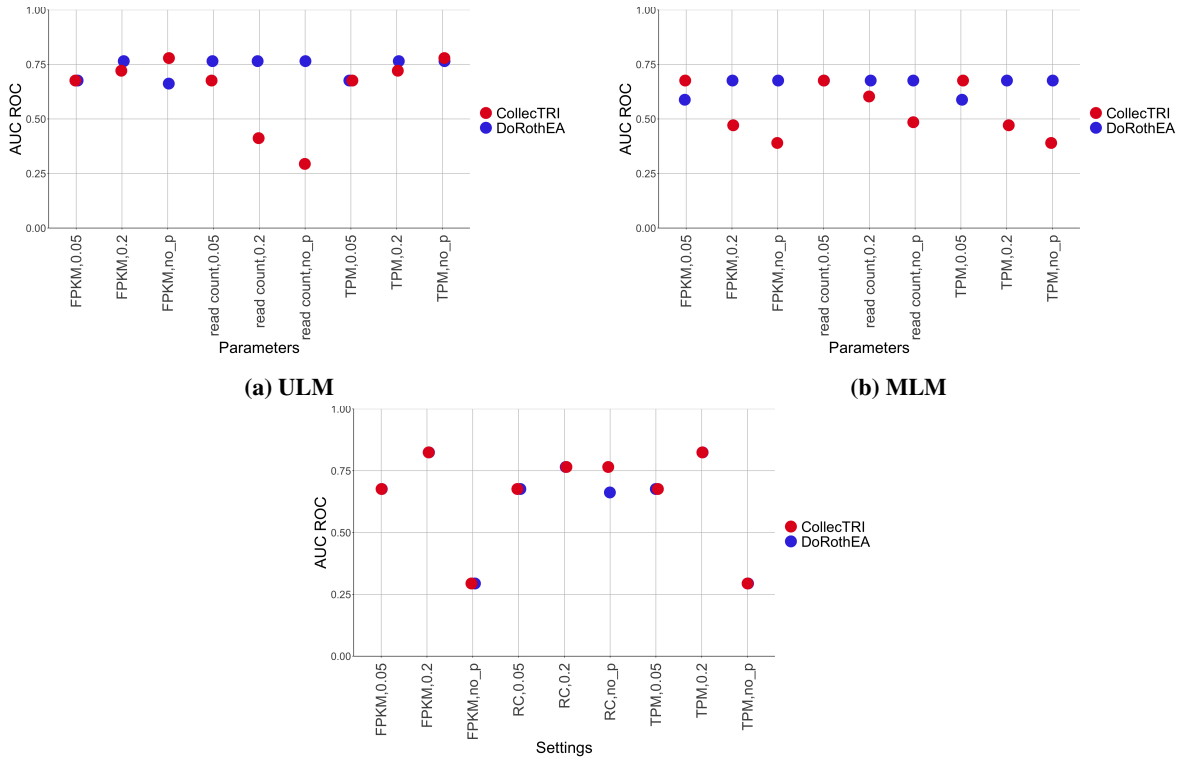
3.2.3 Trends in the Performance of Tool Parameters

AUC ROC results with respect to each tool parameter is analysed and discussed in this section, aiming to provide additional insights into the observed variance in the overall prediction results. The goal is also to highlight optimal parameters for each tool.

DecoupleR Parameters

Gene Regulatory Networks
Figure 3.5 displays synergy predictions in form of AUC ROC values for all tools, where the colours refer to the GRN utilised. As previously discussed, the GRNs had almost no effect on the inference of CONSENSUS, so as expected the synergy results display the same (Figure 3.5 (c)).

Interestingly, the performance of ULM and MLM seems to be higher and more consistent when using DoRothEA, compared to using CollecTRI. From the inference analysis there were not any obvious trends of one GRN leading to the inference of more precise biomarkers than the other, as the use of both GRNs seem to cause deviations in a relatively equal rates. However, as mentioned earlier, this might relate to the fact that some biomarkers are more significant than others in the CASCADE 1.0 network topology. It seems that for ULM and MLM utilising DoRothEA might lead to the identification of biomarkers that are more significant, or in the case where it identifies incorrect biomarker activities, they might be less detrimental than the ones identified when using



(a) ULM **(b) MLM**

(c) CONSENSUS

Figure 3.5: AGS CASCADE 1.0 synergy results from the DrugLogics pipeline highlighted for gene regulatory networks. Synergy predictions in form of AUC ROC values are divided into tool-specific panels. Colours refer to the respective gene regulatory network utilised. Each column in the figure represents a different set of parameters, including different expression count measures and p-value thresholds.

CollecTRI.

When looking at MLMs synergy results in Figure 3.5 (b), there are six cases where CollecTRI noticeably has a lower AUC ROC value. In these six cases, TP53 is inferred in an active state. While MLM infers TCF7_f as inactive with DoRothEA, that does not seem to impact the synergy results, indicating that TP53 has a more significant contribution to the characteristics of AGS (at least in the CASCADE 1.0 network). This might also be attributed to the fact that TCF7_f's role in CASCADE is to activate MYC, which is already active in these incidents. Similarly, when analysing Figure 3.5 (a) there are two incidents of CollecTRI with a particularly low AUC ROC value, which relates to the incidents where TCF7_f, NFκB_f and CTNNB1 are all inferred with opposing expression values according to the gold standards. NFκB_f is also identified with an incorrect inference value with DoRothEA, but this does not affect the synergy outcomes. In this case it seems that the additive effect of these three imprecise biomarkers have a very damaging effect on the synergy results. As a result, for ULM and MLM it seems like utilising CollecTRI overall increases the risk of identifying imprecise biomarker activities with more damaging effects on the model calibration process, than when using DoRothEA. Lastly, it is worth noting that MLM seems to be most affected by the regulatory network used, as is seen from Figure 3.4.

As mentioned earlier, the underlying reason why CollecTRI and DoRothEA may display such large variations is due to their differing sizes of TF-TG interactions, and the corresponding coverage of those regulons. CollecTRI contains more TF-TG interactions and has a greater regulon coverage than DoRothEA. CollecTRI was also found in the benchmark of Müller-Dott et al. to generally perform better at identifying perturbed TFs [57]. Müller-Dott et al. utilised CONSENSUS to benchmark the GRNs, so the almost identical results of DoRothEA and CollecTRI with CONSENSUS seem to be a bit contradictory. This may be due to the fact that the TFs

identified by both GRNs are probably some of the most studied TFs in cancer, meaning that they are likely adequately covered by both GRNs. In addition, only a few TFs are compared in this analysis, while more 907 datasets, corresponding to 450+ TFs was utilised in the benchmark. Furthermore, an early version of the CollecTRI GRN was utilised in this project. More accurate signs of TG-TG interactions are included in the latest version, which may have caused the deviations of these results from the benchmarking of Müller-Dott et al. [57].

Expression Count Measures

Figure 3.6 displays synergy predictions in the form of AUC ROC values for all tools, where the colours refer to the respective expression count measure utilised. From the figure, it is evident that AUC ROC values obtained with the different expression measures are mostly similar for all tools, with a few exceptions. It is especially evident that the use of FPKM and TPM data seems to produce identical results in almost all cases. This suggests that the normalisation methods utilised by FPKM and TPM does not significantly affect the synergy predictions. As utilising FPKM and TPM data mostly resulted in the same inferred biomarkers activities, this is in line with the expected outcomes.

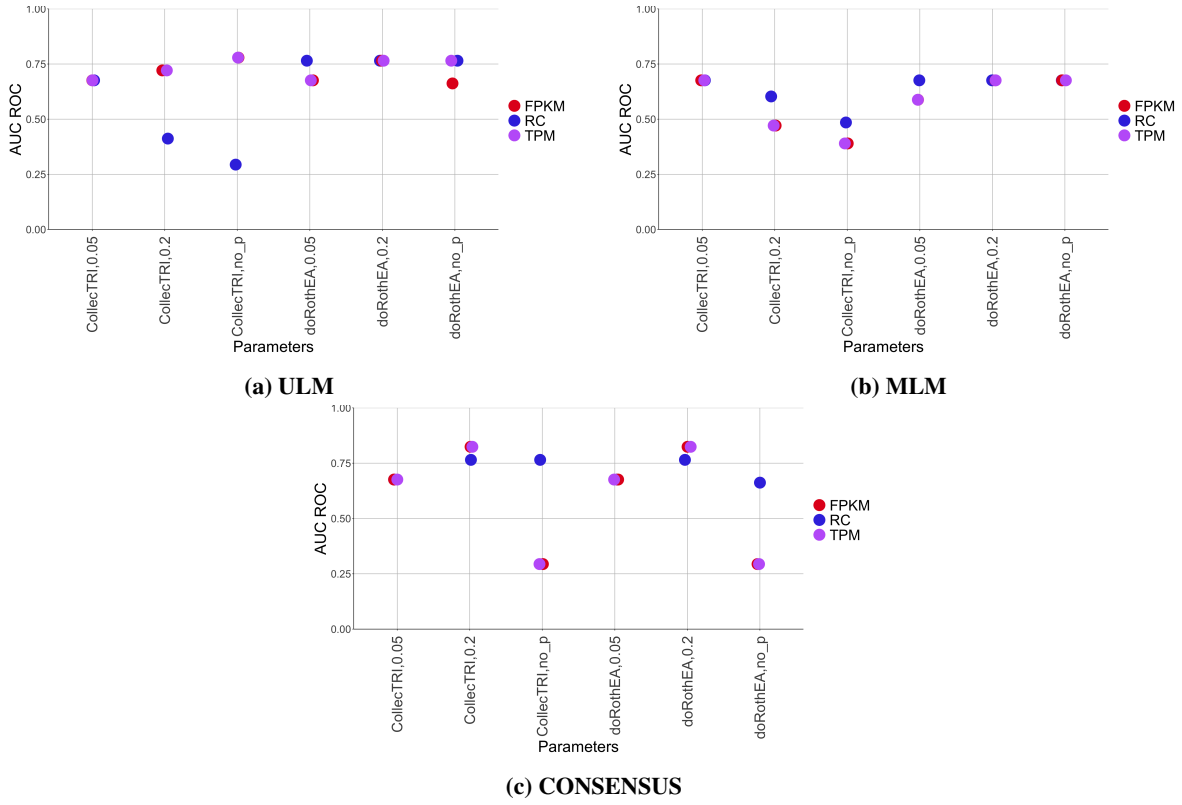


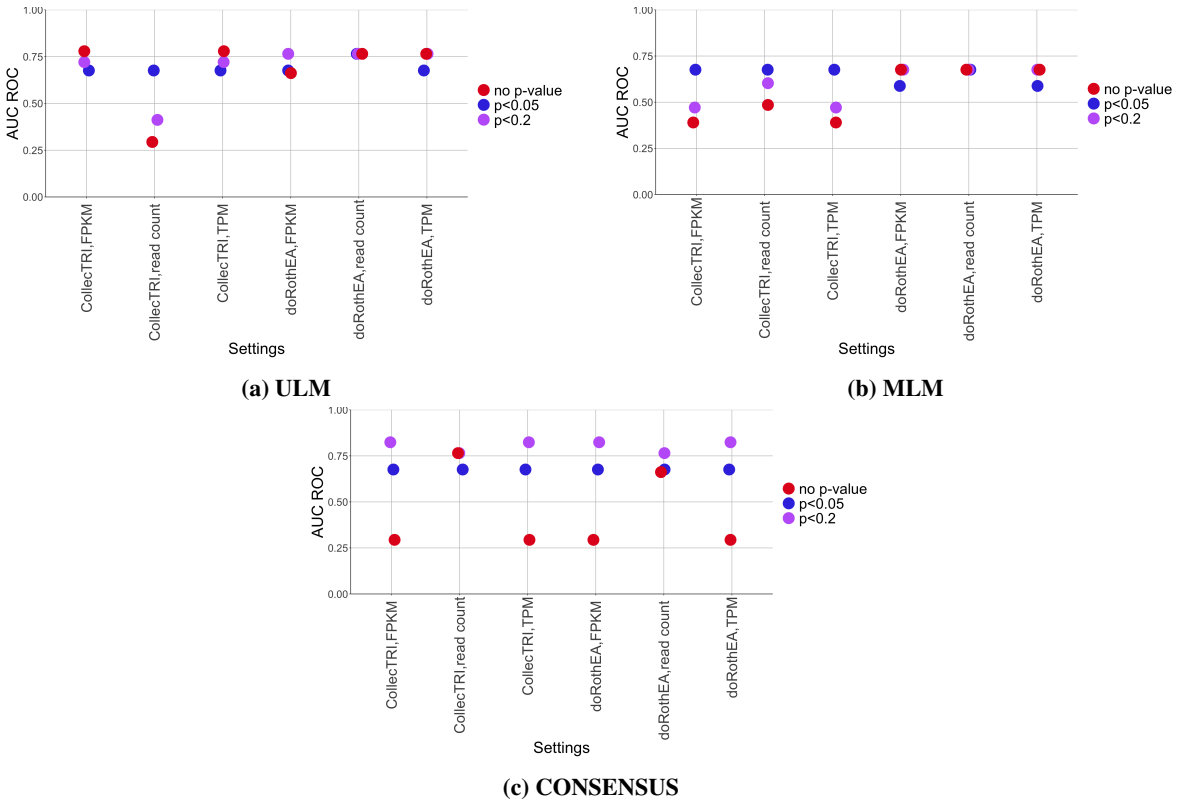
Figure 3.6: AGS CASCADE 1.0 synergy results from the DrugLogics pipeline highlighted for expression count measures. Synergy predictions in form of AUC ROC values are divided into tool-specific panels. Colours refer to the respective expression count measure utilised. Each column in the figure represents a different set of parameters, including different gene regulatory networks and p-value thresholds.

However, there are a few instances where the AUC ROC values differ when using different expression data. For example, the combination of CollecTRI with higher p-values in both ULM and MLM result in some differences when using RC compared to TPM and FPKM data. As mentioned previously, these differences can be attributed to the fact that FPKM and TPM are normalisation methods. However, it seems like utilising RC provides the most consistently high AUC ROC values for both MLM and CONSENSUS, while utilising TPM and FPKM provided the overall highest results for ULM. MLM seems to be overall least affected by the expression count measures.

An explanation for why RC may result in higher synergy results for MLM and CONSENSUS is that, as mentioned earlier, these two tools may require larger variations in the data to infer biomarker activities than ULM. As there may be larger variations in the RC data, this might explain their ability to draw more precise biomarker activities from RC data than with normalised data, as there will generally by smaller variations there. However, this may not always be the case, as utilising raw RC's may cause imprecise inferences as it does not take into account important biological aspects of the data. In contrast, ULM might be able to detect these small variations in the normalised data. For ULM, the larger differences in RC data might provide as noise, which ULM is more sensitive to detect as significant variation, leading to imprecise detection of biomarkers. In conclusion, utilising TPM and FPKM data seems to result in more reliable biomarkers when CONSENSUS and MLM is utilised, while RC seems to be the better choice for ULM.

P-value Thresholds

Figure 3.7 displays synergy predictions in form of AUC ROC values for all tools, where colours refer to the respective p-values utilised. As mentioned in Materials and Methods, p-values are used to filter out biomarker activities with low confidences.



(a) ULM **(b) MLM**

(c) CONSENSUS

Figure 3.7: AGS CASCADE 1.0 synergy results from the DrugLogics pipeline highlighted for p-value thresholds. Synergy predictions in form of AUC ROC values are divided into tool-specific panels. Colours refer to the respective p-values utilised. Each column in the figure represents a different set of parameters, including different expression count measures and gene regulatory networks.

Looking at Figure 3.7 (a) and (b), it is evident that the AUC ROC values obtained with the three p-value thresholds are similar in several cases. However, there are some instances where the AUC ROC values differ more. In the cases where the p-values differ the most, it seems that eliminating the p-value causes the most damage to the synergy predictions. In fact, the lowest AUC ROC values of both ULM and MLM result from eliminating the p-value threshold. In addition, the results indicate that utilising the strict p-value of 0.05 results in the most overall highest synergy predictions with both ULM and MLM.

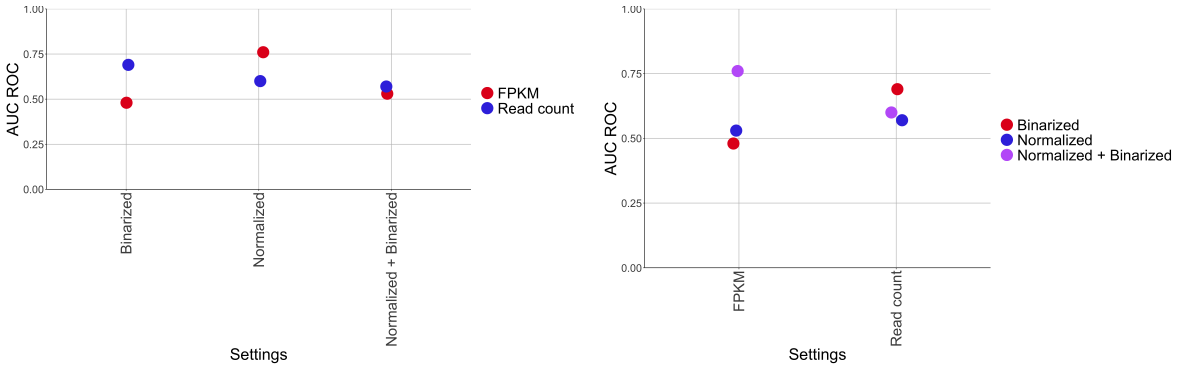
For CONSENSUS it seems evident that eliminating the p-value threshold dramatically decreases the synergy

predictions (Figure 3.7 (c)). Contrasting, the highest AUC ROC values of CONSENSUS consistently result from utilising 0.2 as the p-value threshold. This might suggest that for CONSENSUS, having a very low p-value threshold does not allow for inference of enough biomarker activities, since it already is a restrictive inference tool. As mentioned, lack of important biomarker activities in the training data might lead to faulty models in Gitsbe by not providing enough relevant information. This may in turn result in defective synergy predictions and a lower AUC ROCs. However, eliminating the threshold allows for too much noise in the training data, also resulting in inaccurate model generations in Gitsbe and ultimately incorrect synergy predictions and low AUC ROC results. The observations of the inference results supports this claim, as a higher rate of imprecise biomarkers were found when eliminating the p-value threshold. For MLM and ULM this trend is less obvious, which might be due to the fact that their methodologies originally are less strict at inferring biomarkers. As a result, increasing or decreasing the p-value threshold might not have as dramatic effect on their inferred biomarkers, as for CONSENSUS.

In conclusion, while the AUC ROC values obtained with different p-values are mostly similar, there may be instances where one p-value threshold outperforms the others. Overall, having a relative strict p-value threshold seems to generally result in more precise biomarker inferences for all three tools.

PROFILE Parameters

Figure 3.8 displays the synergy predictions of PROFILE in form of AUC ROC values where the colours refer to the different parameters utilised. Figure 3.8 (a) display the effect of expression count measures and Figure 3.8 (b) displays the effect of utilising binarised, normalised or normalised output that is subsequently binarised.



(a) Expression count measures **(b) Output classification**

Figure 3.8: AGS CASCADE 1.0 synergy results from the DrugLoigcs pipeline highlighted for PROFILE parameters. Synergy predictions are displayed in form of AUC ROC values. Colours refer to the respective PROFILE parameters investigated. Each column in the figures represents a different set of remaining parameters.

The AUC ROC values displayed in Figure 3.8 (a) imply that RC might provide *slightly* more consistently high synergy results than FPKM, although utilising FPKM resulted in the overall highest AUC ROC value. Panel (b) suggests that using normalised output that is subsequently binarised results in better synergy predictions than using normalised data. Utilising the binarised data leads to the highest and the lowest synergy result in the two incidents. Since analysing the parameters in PROFILE is based on very few observations, and the trends are very weak, it not appropriate to draw any conclusions in regards to optimal parameters from this analysis.

3.3 PART III: Statistical Analysis of Synergy Results

The third part of the results in this project relates to the statistical analysis of the synergy predictions. This analysis was executed to investigate to what degree the training data generated by the software tools statistically improved the model calibration over randomly calibrated models. In order to do so, the AUC ROC values

resulting from the tools were compared to bootstrapping of random training data. Consecutive confidence intervals (CI) associated with the bootstrapping were calculated and used together in the statistical analysis. The trends and differences in inference significance is identified and discussed in this section.

Figure 3.9 displays the five highest AUC ROC values related to each inference tool applied with the AGS cell line. The AUC ROC values are displayed together with the mean value (the black lines) of the random bootstrapping and its corresponding confidence interval (CI) (the red intervals). Figure 3.9 displays that the AUC ROC values of CONSENSUS and ULM are consistently higher than the bootstrapped mean values. This suggests that these tools are indeed performing better than randomly selecting biomarker activities from the disease network. Furthermore, the CIs do not overlap with the obtained AUC ROC values, further indicating that the added value of the bioinformatic tools is statistically significant.

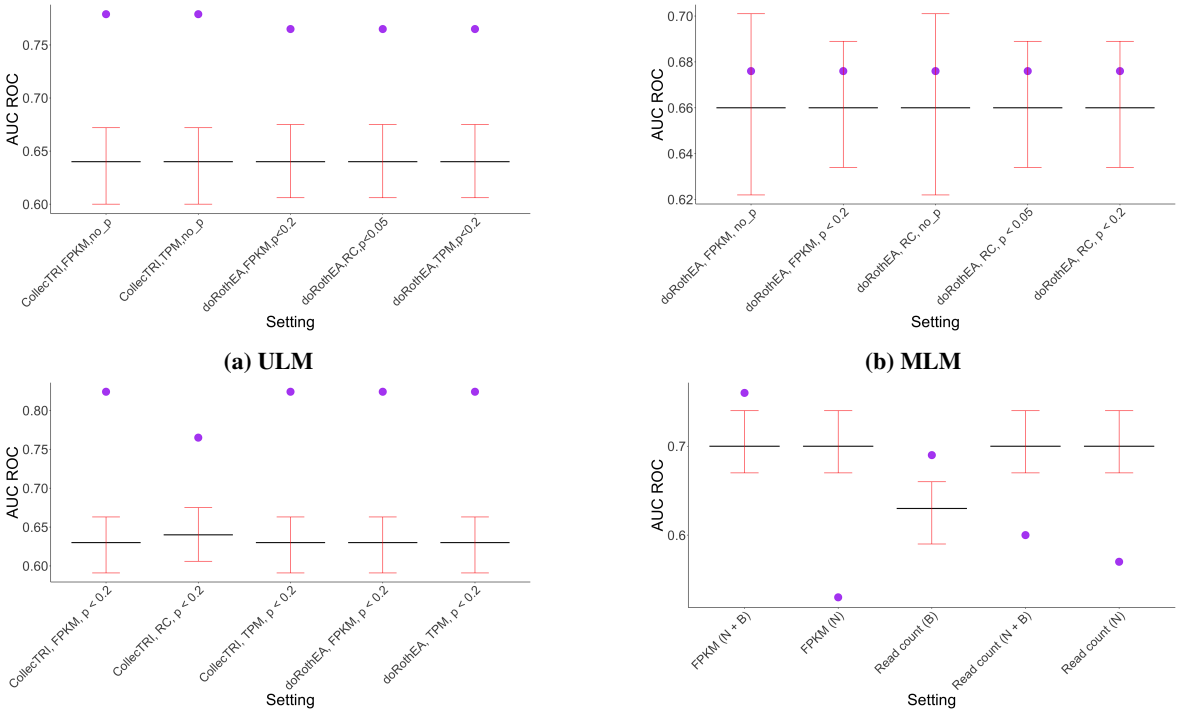


Figure 3.9: Synergy results resulting from the top five inferences of each software tool compared to bootstrapping of random training data of the CASCADE 1.0 model. Purple dots are AUC ROC values obtained by each software tool, and black lines indicate the bootstrapped mean of random training data. Red lines indicate associated confidence intervals.

However, when looking at the results obtained from MLM one can observe that the AUC ROC values are consistently close to the bootstrapped mean, which suggests that the tools are not performing significantly better than random at inferring precise biomarkers from AGS transcriptomics. The CIs for each AUC ROC value also overlap with the bootstrapped mean, which further supports the conclusion that the tool is not significantly outperforming random chance.

When it comes to PROFILE, the obtained AUC ROC values are not consistently close to the bootstrapped mean. The CIs do not overlap with the obtained results, but the performance of the tool varies considerably, with some results performing better and others performing worse than random training data. However, it is worth noting that two of the AUC ROC values are higher than the bootstrapped mean and the corresponding CIs do not overlap, which suggests that this tool *may* hold promise at performing better than random chance. As seen from the analysis of the master project by Thea Hettasch, taking a specific subsets of expression data may perform better than utilising the whole set. Further investigation is needed to determine the statistical significance of PROFILE as an inference tool in order to draw any final conclusions.

To conclude, it seems like MLM and PROFILE are not consistently performing significantly better than random chance at inferring precise biomarkers from AGS transcriptomics to calibrate logical models in the DrugLogics pipeline. In contrast, ULM and CONSENSUS provide statistically significant results according to the findings in this analysis, further supporting the previous findings. However, it is important to note that these conclusions are drawn from inference results from *one* cell line and the use of *one* logical model in the DrugLogics pipeline. Additional analysis with more data is necessary to fully evaluate the performance of these tools.

3.4 Consistency of Trends Across Logical Models and Cell Lines

Whether the trends found from inferences from the AGS cell line and model calibration with the CASCADE 1.0 model is consistent with other model topologies and cell line datasets is discussed in this section. The section do not provide an in depth analysis of the results related to the Lu and the Park models, but their global trends with respect to tool performances and parameter settings are analysed.

3.4.1 Consistency of Tool Performances

Consistency Across Cell Lines
When looking at the synergy results of the Lu and Park models in Appendix E, is it evident that there are not any strong trends indicating that one or more tools consistently outperform the others. In fact, the tool performances differ quite dramatically from cell line to cell line. For instance, with the Park model and the COLO_205 cell line, ULM and MLM seem to outperform CONSENSUS, while with the HCT-116 cell line, the opposite trend seems to be the case (Figure E.1 in Appendix E). However, in both cases, inferences of MLM result in one of the highest synergy scores. This is also the case with the cell line SW48, where MLM seem to slightly outperform ULM and CONSENSUS a few times, in addition to providing the highest overall AUC ROC score. As for the SW620 cell line, there is not a consistent performance pattern across parameter settings, but ULM and CONSENSUS inferences resulted in the highest synergy scores. When it comes to PROFILE, it generally displays a large range of AUC ROC values, with some results comparable to the decoupleR results, but also often some among the lowest values.

The tool performances also differ quite significantly from cell line to cell line with the Lu model (Figure E.2 in Appendix E). However, with the COLO_205 and HCT-116 cell lines, ULM and MLM seem to outperform CONSENSUS in most cases. With HCT-116, MLM also consistently outperform ULM. With the SW48 and the SW620 cell lines the results from all three decoupleR tools are strikingly similar. In addition, PROFILE performs consistently equal or better than the decoupleR tools, and is actually the highest ranking tool in both the HCT-116 and the SW620 cell lines. As a result, it is hard to conclude about an optimal tool across different cell lines, as the characteristics of the datasets seem to significantly impact the inference performances of the tools. However, there seem to be a slight advantage of using ULM and MLM over CONSENSUS, which is conflicting from the findings with the AGS cell line.

Consistency Across Logical Models
Figure E.1 and Figure E.2 in Appendix E also highlights fact that the tools performances differ within the *same cell lines* across the two models. This suggest that the topology of the logical models significantly impacts the synergy results. In fact, for the HCT-116 cell line, CONSENSUS was consistently one of the highest performing tools with the Park model, while with the Lu model it was consistently the lowest performing tool. This is also evident with the SW48 and SW620 cell lines, where the results were rather similar across tools with the Lu model, while there was a relatively large variations with the Park model. This highlights one of the main challenges and greatest weaknesses with the analysis of this project. As the characteristics of the model topology highly influences the synergy results, the conclusions drawn of the tool performances cannot be generalised across models. The limitations of model topologies in this analysis is further discussed in section 3.5.4. Nevertheless, when looking at the results obtained from the statistical analysis related to the Park model,

the inference results from all decoupleR tools seem to generally perform better than random calibration data. For the Lu model, however, the results are often found within or below the CI of the random calibration data.

3.4.2 Consistency of Optimal Tool Parameters

As for optimal tool parameters, there are some trends from the AGS CASCADE 1.0 results that persist across models and cell lines, while other trends are changed.

CONSENSUS
One trends that seems to persist across models and cell lines is that different parameters seem to have little to no effect on the synergy results of CONSENSUS, as can bee seem from Figures E.3 - E.8 in Appendix E. This is especially evident with different GRNs, where the AUC ROC values resulting from DoRothEA and CollecTRI are close to identical across cell lines and models. This was also the case with AGS and CASCADE 1.0. As for using different expression count measures, varying between FPKM and TPM seem to have relatively low impacts on the results of CONSENSUS across these cell lines and models as well. However, using RC with CONSENSUS did not result in consistently higher AUC ROC values than FPKM/TPM across all models and cell lines, as was indicated with AGS and CASCADE 1.0. Nevertheless, using RC resulted in some cases of slightly higher AUC ROC values compared to FPKM/TPM across cell lines. When it comes to p-values for CONSENSUS, there also seem to be a less consistent trend across cell lines and models compared to with AGS and CASCADE 1.0. In fact, the p-value thresholds seem to not highly affect CONSENSUS overall, although eliminating the p-value results in better or worse results compared to the more strict thresholds in some cell lines. This is especially evident with the Lu model and the COLO_205 cell line, where the AUC ROC values are higher when eliminating the p-value in all cases. However, given that eliminating the p-value when using CONSENSUS results in the largest range of AUC ROC values, it might be safer to keep a threshold to secure consistent results.

ULM
For ULM, using DoRothEA was previously found to consistently improve the synergy results over CollecTRI. However, using CollecTRI seems to be the better choice across models and cell lines, as can be seem from Figure E.9 and Figure E.10 in Appendix E. When it comes to utilising different expression count measures, the results across models and cell lines displayed less variance than with the AGS CASCADE 1.0 results. In fact, the results are very similar and almost identical in certain cases. However, it seems that using RC provides a slight advantage in some cell lines, which was *not* the case with with AGS and CASCADE 1.0. For the p-values, there does not seem to be an evident trend across models and cell lines here either, as many of the synergy results are very similar across p-values. However, it seems that eliminating the p-value resulted in slightly more differing AUC ROC values, with some higher and some lower than with p-value thresholds. This means that having a threshold may be the best option for ULM, to obtain consistent results.

MLM
When it comes to MLM, using DoRothEA also consistently resulted in the highest overall AUC ROC values with AGS and CASCADE 1.0. Same as for ULM, using CollecTRI might seem to be the better choice across models and cell lines, although the improvements are rather small and not consistent for all cell lines (Figure E.15 and Figure E.16 in Appendix E). Also similar to ULM, the results of expression count measures are very similar across models and cell lines, with RC displaying a slightly larger variation than FPKM/TPM. However, using RC seemed to slightly improve the synergy results, might make it the safest options with this tool. As for the p-values, eliminating the threshold when using MLM seems to result in slightly lower AUC ROC values in both models for most cell lines, although this is not an evident trend.

PROFILE
Due to limited parameter combinations in PROFILE it was not possible to draw any firm conclusions from the AGS cell line with the CASCADE 1.0 model. When reviewing the results from the Lu and the Park models

with respect to different output classifications (Figure E.21 and Figure E.22 in Appendix E) it still seems like the results differ quite significantly without any obvious trends across cell lines and models, making it hard to conclude with an optimal output. However, there might be a slight tendency of normalised output performing a bit better for some cell lines. When it comes to different expression count measures with PROFILE, it is still not possible to conclude with an optimal one, as there is no evident trends in the results.

3.5 Limitations and Critical Reflections

As described throughout this section, the obtained inference and synergy results can be explained based on biological knowledge of the modelled cancer system, and knowledge of the methodologies and statistics of the software tools. While this study provides insights into the performance of inference tools, it is important to acknowledge the limitations that may impact the findings. These limitations are discussed in the following paragraphs, together with possible suggestions to address the issues.

3.5.1 Training Data Binarisation

As mentioned in Materials and Methods, all training data from the inference tools were binarised prior to running the DrugLogics pipeline (except for the normalised training data in PROFILE). There are several reasons why this was done, one of them being that the CASCADE 1.0 model can be considered a Boolean model, although its output nodes are multivalued. This is due to the fact that the output nodes are not directly utilised in the pipeline, as it is the binary values of the nodes *connected to them* that are used to calculate the growth value. As these nodes are Boolean, the model can be considered as Boolean. Consequently, the nodes in the networks of the generated models in Gitsbe, which the training data is compared to, are also Boolean. As a result, also having Boolean values in the training files seemed reasonable. In addition, as mentioned earlier, interpreting the inference results from a biological point of view is more complicated for normalised expression values.

However, binarising the training data means that the obtained inference results of the tools are shifted to an extreme value. This results in inference results that do not reflect the accurate findings of the tools, and may cause biases when comparing the tool performances. It is important to note that the data was scaled between 0 and 1 prior to the binarisation, in order to maintain the integrity of the data. Nevertheless, if the extreme scenario that MLM's inferred activity of TP53 was found and scaled to be 0.51 it would be 1 after the binarisation, while if ULM inferred activity of TP53 was 0.50 after the scaling it would be 0 after the binarisation process. Consequently, it would seem that ULM is much more effective at inferring the true activity level of TP53 than MLM, while in reality they are almost identical. While the biological interpretation might have been more complicated, the synergies found by the pipeline might have been more precise in terms of displaying a true inference performance of the different tools. As a result, it is important to keep this bias in mind when looking at the findings of this study. Utilising continuous training data might be desirable if a more comprehensive analysis is to be conducted on this topic.

3.5.2 Data Quality and Complexity

As mentioned earlier, the characteristics of the data can have a significant impact on the performance of inference tools at identifying biomarkers. One such characteristic is the amount of variation present in the data, which refers to the range of expression values. As previously stated, MLM and CONSENSUS may require more substantial variations in the data to infer a biomarker than ULM. In addition, the number of samples in the dataset can affect the ability of both ULM and MLM methodologies to detect significant relationships between regulators and molecular features. With smaller sample sizes, there might not be sufficient statistical power to detect subtle effects for MLM. In contrast, if the dataset contains many genes that are regulated by the

combined effect of multiple regulators, MLM might be more capable than ULM at identifying these regulators precisely. As a result, the specific characteristics of the AGS cell line might cause inference advantages for some tools, leading to a better performance than when datasets with other characteristics were utilised. This may explain some of the observed variance of tool performances resulting from the AGS cell line compared to the other cell lines.

3.5.3 AUC ROC as a Statistical Measure

While ROC curves are useful for evaluating the performance of drug response prediction models, they also have limitations, particularly in their use of true positive rate (TPR) and false positive rate (FPR).

One limitation of using AUC ROC values as a statistical measure of tool performances is that it can be affected by the characteristics of data used to generate them. If the data is imbalanced, AUC ROC values may be misleading [75]. An imbalanced AUC ROC value may for instance be a very high true negative rate, and a very low true positive rate, which is the case in most of the inference results in this analysis. In such imbalanced scenarios, the AUC ROC value can be misleading because it focuses on the overall performance of the classifier across all possible classification thresholds. It measures how well the model ranks the positive samples against the negative samples. However, if the one class dominates the samples, the classifier might perform very well in correctly classifying the majority class (high true negative rate) but poorly on the minority class (low true positive rate). Despite the poor performance on the minority class, the AUC ROC can still be high because it considers the overall ranking of the samples. This may lead to incorrect conclusions regarding the performance of the tools.

In addition, if the TPR and FPR are imbalanced, the resulting AUC ROC value may be difficult to interpret. The AUC value does not provide information about the rate of TPs and FPs, which can make it difficult to interpret the biological meaning of the obtained values. This can ultimately limit the ability to draw meaningful conclusions of the underlying biological mechanisms that contributed to the low or high AUC ROC value. In this project it was desirable to take both the TPR and the FPR into account, and due to the extensive number of synergy results it was not feasible to analyse each AUC ROC curve. Nonetheless, if it is desirable to assign different weights to the TPR or FPR, it might be appropriate to conduct an analysis of the obtained AUC ROC curves, or calculate each rate separately.

To obtain a more robust analysis, different statistical measures could be added in addition to AUC ROC, for instance AUC PR (Precision-Recall). PR curves are another way of evaluating the performance of binary classifiers, only that they focus on the precision and recall metrics rather than the TPR and FPR like ROC [75]. PR are particularly useful when the positive class is rare (which was the case in this analysis), as they are more informative than ROC curves in this scenario. PR curves may also be is good for assessing if a positive/negative prediction has a real-world usage merit, by assessing if the prediction result matches reality. For instance in situations where there is a significant class imbalance (very low or high prevalence of positive samples), precision-recall can provide valuable insights by considering the trade-off between precision (the proportion of correctly predicted positive samples) and recall (the proportion of actual positive samples correctly identified by the model). Utilising such additional statistical measures together with the AUC ROC may provide more accurate assessments of the performance of the tools, and should be applied if a more comprehensive analysis is to be conducted on this topic.

3.5.4 Using Synergy Predictions From the DrugLogics Pipeline as Measures of Tool Performances

The goal of this project was to identify alternative approaches of generating biomarker activity data to improve model calibration in the DrugLogics pipeline. However, the use of synergy predictions generated by the DrugLogics pipeline as a measure of the relative performance of inference tools is subject to bias, which may lead

to erroneous conclusions.

Boolean Models as Cancer Systems

One limitation of the DrugLogics pipeline is that it does not account for the dynamic changes in the biological system. As mentioned in Materials and Methods, Boolean models are an essential part of the DrugLogics pipeline, as they allow for the simulation of complex biological networks.

One of the limitations of using Boolean models in the DrugLogics pipeline is that they assume that the interactions between genes and proteins are static and binary, meaning that they either exist or do not exist. However, in reality, the interactions are more complex, such as quantitative or temporal interactions, which cannot be captured by Boolean models. Boolean models does not account for the time-dependent changes in gene expression, which may be critical in understanding the response of cancer cells to different drugs. Therefore, the predictions of the pipeline may not always reflect the actual biological behaviour, as the model utilised to generate them does not accurately reflect the biology of the disease. As a result, while the inference results of a tool may accurately reflect the cellular system, the synergy predictions may not be in line with the expected results as certain assumptions are made by the Boolean model that deflect from a 'real' biological system. This has the possibility of introducing biases in the obtained results, that may lead to incorrect conclusions regarding the tools inference precision. The Boolean model's limitations should be taken into account when interpreting the results, and further research is needed to develop more sophisticated models of biological cancer systems.

Deficiencies in the Prior Knowledge

Another limitation of the DrugLogics pipeline is that it heavily relies on prior knowledge. Because of the drawbacks of this, the predictions of the pipeline may not be what is biologically observed.

Deficiencies in the Logical Models

Gaps in the prior knowledge of the logical models used by the DrugLogics pipeline may also affect the synergy results. Since the models used in the pipeline are derived from prior knowledge, modifications to the prior knowledge must be expected to affect the predictive performance of the model [42].

The logical rules of the models are based on experimental data and prior knowledge of the cancer systems, and are used to determine how the components of the system interact with each other. However, the models are sensitive to the logical equations used to describe these interactions. Small changes in the equations can lead to significant changes in the behaviour and predictions of the model. Thus, if the prior knowledge network is incomplete or incorrect, it may lead to false predictions of synergistic drug combinations. Flobak et al. [42] states that *"We find that our approach is somewhat sensitive to errors in the training data, and even more sensitive to errors in the prior knowledge, indicating that curation quality is paramount to our modeling approach"*. If a biologically significant biomarker is missing from the logical model, the presence of this biomarker in the training data is not considered. This may mean that important information inferred from the tools may be discarded, possibly leading to false conclusions regarding their precision. In addition, if the interactions between the entities in the logical model do not accurately describe the true biological processes, this might also cause improper conclusions. For example in the case of TCF7_f, some tools inferred it as active while others found it to be inactive. The way TCF7_f is included in the Boolean network, its incoming and outgoing interactions, are decisive for whether inferring it as 0 or 1 results in the highest synergy predictions. Remembering the fact that different studies found contradicting findings on several of the biomarker activities included in the CASCADE 1.0 model (Appendix F), some of the interactions in the model could be incomplete or based on outdated information. Therefore, it is crucial to have reliable and up-to-date prior knowledge to obtain accurate predictions with the DrugLogics pipeline.

As described in Materials and Methods, the networks of the Lu and the Park models was simplified by reducing the number of nodes and edges, prior to using them in the DrugLogics pipeline. Such reductions can result in losing some important interactions that contribute to the synergy predictions. This may be the case with the the Lu model, as its synergy results are overall lowest, suggesting that it may be of lower curation quality than the CASCADE 1.0 and the Park model, or that the modifications done significantly damaged the quality of the

model. The Lu model also had the most comprehensive reduction out of the two. Inconsistent behaviour of the Lu model was also fount in the analysis of Thea Hettach, further supporting this claim. The Park model also displays generally lower AUC ROC values than the CASCADE 1.0 model, at least with the COLO_205 and HCT-116 cell lines. This may also suggest that the predictive performance of the Park model in the DrugLogics pipeline is not as consistent and reliable as the CASCADE 1.0 model. As a result, the CASCADE 1.0 seems to have the overall highest predictive performance, suggesting that the results and conclusions drawn form this model may be the most reliable. Nevertheless, such deficiencies in model networks, and their significant impact on the synergy results in the DrugLogics pipeline, makes it difficult to generalise findings of tool performances across models.

Deficiencies in the Gold Standard Drug Synergies
Similarly as with the logical models, deficiencies in the gold standard drug synergies have the possibility of directly influencing of the synergy results.

In the case of CASCADE 1.0, the gold standard drug synergies refer to a set of four synergies: PI-PD, PI-5Z, PD-AK and AK-5Z. In this case, the four gold standard synergies were experimentally tested by Flobak et al. [30]. However, there may be other relevant drug synergies for AGS that are not included in this set. This particular experiment was limited to investigation of seven nodes in the CASCADE 1.0 network, and their 21 pairwise combinations [30]. Analysing a larger collection of nodes, possibly with a larger logical model for AGS, may result in the discovery of novel drug synergies. If novel drug synergies were included in the gold standards, that would highly affect the obtained synergy results, as it content is directly used to calculate the AUC ROC values. In turn, this might have affected the conclusions drawn of the tool performances.

Nevertheless, these seven nodes were selected based on the availability of chemical inhibitors for targeting them in biological experiments. As a result, more information is needed in regards to chemical inhibitors in order to scale up this experiment. Such limitations would also relate to the gold standard drug synergies found by Jaaks et al. [46].

The Methodology of the DrugLogics Pipeline
As mentioned previously, there are some assumptions made by the pipeline that may affect the synergy results, such as the number of evolutions specified in the configuration file. However, 50 evolutions is above the plateau level of fitness found by Flobak et al. [42]. In addition to the number of evolutions, a number of models per generation, and a number of models selected for the next generation can also be specified in the configuration file. Furthermore, one can specify the number of mutations, the fitness threshold used to stop the evolutions and so on. While the same set of parameters in the configuration file was utilised throughout this project, the results and the conclusions drawn from them might have looked differently if another set of parameters was chosen.

As mentioned earlier, there may also be some biases in the pipeline towards preferring low numbers, or at least not having excessive numbers, of biomarker activities in the training data, as found by another master project by Thea Hettasch. As a result, tools which infers a high number of biomarkers (like PROFILE) may have been given a lower AUC ROC value than what can be biologically explained.

3.6 Guidelines for Selecting Inference Tools for the DrugLogics Pipeline

In this section, the main objective of this project, to assess whether an optimal inference tool (and tool-specific parameters) could be identified, is discussed. Recommendations and guidelines is given for selecting inference tools for generating training data for the DrugLogics Pipeline.

One thing that might be obvious from this analysis is the fact that there might not be one optimal tool that under all circumstances will infer precise biomarkers from omics data to use as training data in the DrugLogics pipeline. However, the analysis highlighted some trends that may be utilised as guidelines for selecting an

appropriate tool for different purposes.

CONSENSUS has proven to be a robust tool at inferring precise biomarkers from the AGS cell line, decreasing the likelihood of identifying incorrect biomarker activities. The overall highest synergy predictions from the AGS cell line and the CASCADE 1.0 model were obtained by utilising training data generated by CONSENSUS, which makes it a promising tool for identifying biomarker activities to calibrate logical models in the DrugLogics pipeline. However, the consistently high performance of CONSENSUS was not an evident trend across other model topologies and dataset characteristics. This decreases the confidence of CONSENSUS as the most robust tool, as it seemed to be slightly outperformed by ULM and MLM in other models and cell lines. However, the fact that CONSENSUS is least affected by the different parameter conditions seem to persist across models and cell lines. There also seemed to be a slight preference towards using a p-value threshold of either 0.05 or 0.2 when utilising CONSENSUS, as it seems that eliminating this threshold may damage the robustness and preciseness of the tool. There might also be a slight preference towards utilising RC data over FPKM/TPM with CONSENSUS. The analysis did not reveal any preferential GRN, as the inference results of CONSENSUS was consistent across these parameters for almost all models and cell lines. However, the inference results of CONSENSUS, as mentioned earlier, may introduce a level of complexity that can be difficult to interpret in a biological way, especially if the results from different workflows are contradictory. If considering implementation of CONSENSUS in the DrugLogics pipeline, this complexity needs to be taken into account.

Although generally outperformed by CONSENSUS, ULM also proved as an inference tool capable of identifying expected biomarker activities as checked against gold standards from the AGS cell line. The inference results of ULM resulted in constantly high synergy results when applied in the DrugLogics pipeline, although the results of CONSENSUS were overall highest. Nevertheless, there is also ground for marking ULM as a promising tool to identify biomarkers for use as training data in the DrugLogics pipeline, as ULM displayed great promise with the other models and cell lines as well. As for the most optimal parameters in ULM, there are less consistent trends than for CONSENSUS, so there is need for a more thorough analysis in order to draw any strict conclusions. Based on the obtained results from the AGS cell line, there is an indication that utilising DoRothEA as a GRN, with a strict p-value threshold and FPKM/TPM data, provides the most robust inference of biomarker activities. However, the results from the other models and cell lines suggested that CollecTRI and RC data might be a better choice. Given that using CollecTRI improved more results, and that the observed advantage of DoRothEA with AGS and CASCADE 1.0 were only observed in specific cases (high p-values), it may be advisable to use CollecTRI as the preferred GRN with ULM. The same can be said for recommending the use of RC data. However, these analyses should be repeated with the latest version of the CollecTRI regulons. Nevertheless, the suggestion of using a strict p-value from the AGS and CASCADE 1.0 results seemed to persist across the other models and cell lines.

From the results of AGS and CASCADE 1.0, MLM proved the least robust at inferring precise activities to calibrate logical models in the DrugLogics pipeline. Its synergy results had a larger degree of variance, its inference result displayed the most deviations from the expected inference results, and it also obtained the overall lowest AUC ROC values compared to the other decoupleR tools. However, its synergy results when utilising the DoRothEA network are rather consistent, and also relatively high. In addition, MLM also displayed great promise with the other models and cell lines, which indicates that it may be of value after all. Same as for ULM the optimal parameters seemed to be different with the other cell lines and models, suggesting CollecTRI as the optimal GRN. Same as for ULM, the general recommendation of a GRN for MLM would also be CollecTRI, as the observed improvements of DoRothEA was only observed a few times (also with high p-values) and more results of other cell lines are improved with CollecTRI. Contrasting, the recommendation to utilise a stringent p-value and RC data derived from the AGS and CASCADE 1.0 outcomes remained relatively consistent across models and cell lines.

Based on the obtained results from PROFILE and the limited variations of parameters utilised, it would be the least recommended tool to utilise to generate training data for the DrugLogics pipeline. This is because there of the great deviations from the expected inference results, and also its large variance and range in synergy results.

However, if PROFILE is utilised, it seems like normalised output produced slightly highest synergy results for some cell lines. However, this is not a recommendation, as the trends were very weak. For the same reason, it was not possible to conclude with an optimal expression count measure based on this analysis. However, using different subsets of the PROFILE output may hold promise at increasing its value as a inference tool for the DrugLogics pipeline.

Due to the fact that the results from CASCADE 1.0 may hold the most significance with respect to comparing tool performances due to its high caution quality and predictive power, the overall recommended tool for further research is CONSENSUS. This is also due to the fact that CONSENSUS displayed the most consistent inference results of TFs across all models and cell lines, and its relative strong statistical significance (at lest with the CASCADE 1.0 and Park models). However, based on the analysis of Thea Hettach, and the displayed potential of PROFILE with the Lu and Park models, an additional alternative may be to supplement the inferences of CONSENSUS TFs with subsets of the activity levels of other entities by PROFILE. The analysis of Thea Hettach may be useful for determining an appropriate sub-setting approach for the PROFILE outputs.

The recommended tool-specific parameters based on the analysis conducted in this project are summarised in Table 3.3

Table 3.3: Recommendations of tool-specific parameters.

	CONSENSUS	**ULM**	**MLM**	**PROFILE**
P-value cutoff	0.05 / 0.2	0.05 / 0.2	0.05 / 0.2	-
GRN	DorothEA/CollecTRI	CollecTRI	CollecTRI	-
Expression count measure	Read counts	Read counts	Read counts	?
Output classification	-	-	-	Normalised?

Finally there are some additional factors that should to be taken into consideration before selecting a software tool for biomarker inference, two of which are briefly presented here.

Computational Efficacy
One advantage of ULM is that it may be computationally faster than MLM, as it only analyses one variable at a time. The computational cost of using CONSENSUS is even higher than that of using ULM or MLM alone. However, for the purpose of this project there was no noticeable difference of the three decoupleR tools, as they all compiled in a matter of seconds. However, running the PROFILE script to infer biomarker activities was more time consuming (up to 15 minutes per run). If a large-scale analyses is to be conducted, it might be wise to consider the computational efficiency of the tools. If a smaller analysis is to be conducted, this might not be of much relevance.

The Characteristics of the Data
As mentioned earlier, if the data contains much noise or is derived from a complex biological system with strong connections between the regulators in the network, the relationships between regulators and molecular features might be more difficult to discern by ULM, so MLM or CONSENSUS might be more appropriate tools to infer biomarkers from such data. In contrast, if the dataset is small, or has low variations and low noise, ULM might be the preferred tool as might be able to detect these variations better than MLM. In reality, one might not know the quality, complexity or characteristics of the data prior to analysis. In these cases, it seems like CONSENSUS is the safest choice, as it includes the characteristics of both ULM and MLM.

4 Conclusion

4.1 A Brief Summary

Computational modelling of cellular systems is a powerful tool to increase our understanding of biological processes and disease mechanisms. Logical modelling, in particular, is useful in building simplified models of complex biological systems that have predictive power to suggest new therapeutic drug combinations. By using personalised cancer data, these models can be calibrated to individual patients. By integrating experimental data and prior knowledge, computational models such as the DrugLogics pipeline can simulate the behaviour of patient cancer cells and predict the effect of drug combinations on those cells.

During this project, software tools capable of inferring high-quality biomarker activities from omics data were identified. It was also found that *in silico* inference of precise biomarker activity states can improve calibration of logical models in the DrugLogics pipeline, optimise cell line specific drug synergy predictions, and reduce time spent on manual curation of biomarkers. While there was not found an optimal tool that can infer precise biomarkers under all circumstances, some trends identified in this project can support the recommendation on the use of new tools and approaches in the DrugLogics pipeline. CONSENSUS proved the most consistent and robust inference tool and is the overall recommended tool for implantation in the DrugLogics pipeline to calibrate logical models. ULM and MLM may may also provide viable options, but their performance were more affected by parameter settings and dataset characteristics. Using the whole output of PROFILE is not recommended for generating training data for the DrugLogics pipeline, but sub-settings its output in combination with the CONSENSUS TFs may be of value for further research. The project also highlighted some optimal parameters for precise inference, which may be summarised in that a p-value threshold of 0.2 or 0.05 and the regulatory network CollecTRI may be applied for optimal inferences with the decoupleR tools. However, the analysis should be redone with the latest version of CollecTRI to confirm this. The recommended expression count measures based on this analysis are read counts for all decoupleR tools. If PROFILE is used, normalised expression data resulted in the slightly higher predictive performances in this analysis. However, there are multiple external factors that may have influenced these results, like the characteristics and complexity of the data, limitations of utilising Boolean models, using AUC ROC values as a statistical measure and deficiencies in the prior knowledge of the logical models. The inconsistency of tool performances across model topologies and dataset characteristics limits the overall value of the findings, as the findings of tool performances cannot be generalised across the studied logical models and cell lines.

4.2 The Value of the Research

The findings of this project may be utilised as guidelines for selecting software tools capable of inferring biomarkers that may be used as training data to calibrate logical models in the DrugLogics pipeline. The results also provide a basis of further research aiming to improve accurate inference of biomarker activities. Further research in this area can empower oncologists and healthcare professionals to make more informed decisions about the most suitable treatment options for individual patients. Ultimately, the utilisation of logical modelling and biomarker inference in cancer care can lead to more personalised and effective treatments, potentially improving patient outcomes and quality of life.

In addition, the findings of this master's thesis provide some guidance on the parameters that may provide most optimal for accurate inference for the tools analysed, at least for ULM, MLM and CONSENSUS. The scripts produced in this project may also automate the process of inferring biomarker activities, converting node names from HGNC annotations to model specific annotations, and formatting and creating a training data file ready to use with the DrugLogics pipeline. Automated scripts created for calculating AUC ROC values from numerous synergy files may also be of value. In addition, the inconsistent or absent inference of some TFs may be utilised at indicators of lack of proper annotated regulatory interactions of these TFs in prior knowledge

networks utilised.

4.3 Recommendations for Further Work

To achieve the goal of identifying an optimal software tool that can accurately and precisely infer biomarkers from omics data which can be leveraged to calibrate logical models for predicting drug synergies for personalised cancer treatment in the DrugLogics pipeline, there are several recommendations for future work that could be explored.

One possible avenue for further analysis would be to explore the use of additional data types, such as proteomics. Proteomics data is the measure of expression levels of proteins in a cell, which can provide complementary information to transcriptomics data. Utilising more types of data to evaluate the inference power of tools may provide strength to the observed findings of the inference tools, or additional insights into the differences and limitations of the them. In the same way, it may be interesting to explore additional logical models and cell lines. Additional logical models that may be interesting to add to this analysis could be CASCADE 2.0 and CASCADE 3.0. CASCADE 2.0 [76] and CASCADE 3.0 [37] which are extensions of CASCADE 1.0 were additional nodes and edges have been added to the network. These models could provide valuable insights into the inference power of the software tools by seeing if the added nodes may change or strengthen the performance outcomes of the tools.

Another area that would be interesting to explore could be the use of an R package created by the DrugLogics group called EMBA (an acronym for 'Ensemble (boolean) Model Biomarker Analysis'). EMBA aims to analyse the predictive performance and synergy prediction output of the models generated by Gitsbe module in the DrugLogics pipeline, and identify significant nodes (biomarkers) within the Boolean networks that significantly contribute to the observed synergies or produce better predictions [77]. These nodes may contribute to the accurate prediction of experimentally observed synergies or improve overall prediction performance. By identifying significant nodes within the Boolean networks that contribute to the observed synergies or improve overall prediction performance, that might indicate node activities that might aught to be included in the training data to improve the calibration of the logical models.

Another important area is to explore additional inference tools, such as ROMA. ROMA is specifically designed to efficiently and accurately compute the activity of gene sets or modules with coordinated expressions [78]. The activity quantification in ROMA is based on a simple uni-factor linear model of gene regulation that approximates the expression data of a gene set by its first principal component. ROMA can be used in various contexts, ranging from estimating differential activities of transcriptional factors to identifying overdispersed pathways in single-cell transcriptomics data. Additional workflows that may be worth looking into can be Reverse Causal Reasoning (RCR) [79] or High Throughput Pathway Interpretation and Analysis (hiPathia) [80].

Finally, it could be interesting to explore the use of a pool of software tools to identify novel nodes that might be of relevance to a specific logical cancer model, but are not yet incorporated in the model. For instance, nodes that are frequently identified by multiple tools from omics data related to a specific cancer type may represent nodes that should be included in the logical model of the cancer system. By incorporating these nodes into the model, it may be possible to improve the predictive power of the logical models, and thus obtain more accurate drug synergy predictions for personalised treatment strategies.

Bibliography

1. World Health Organization. WHO report on cancer: setting priorities, investing wisely and providing care for all. Tech. rep. 2020. Available from: https://www.who.int/publications/i/item/9789240001299
2. World Health Organization. WORLD CANCER REPORT : cancer research for cancer development. Tech. rep. Lyon, 2020
3. Quaresma M, Coleman MP and Rachet B. 40-year trends in an index of survival for all cancers combined and survival adjusted for age and sex for each cancer in England and Wales, 1971-2011: A population-based study. The Lancet 2015 Mar; 385:1209–10. DOI: 10.1016/S0140-6736(14)61396-9
4. Hanahan D and Weinberg RA. Hallmarks of cancer: The next generation. Cell 2011 Mar; 144:646–54. DOI: 10.1016/j.cell.2011.02.013
5. Giancotti FG. Deregulation of cell signaling in cancer. FEBS Letters 2014 Aug; 588:2558–9. DOI: 10.1016/j.febslet.2014.02.005
6. Zhang Y and Wang X. Targeting the Wnt/β-catenin signaling pathway in cancer. Journal of Hematology and Oncology 2020 Dec; 13:1–4. DOI: 10.1186/s13045-020-00990-3
7. Yap TA and Workman P. Exploiting the cancer genome: Strategies for the discovery and clinical development of targeted molecular therapeutics. Annual Review of Pharmacology and Toxicology 2012; 52:549–50. DOI: 10.1146/annurev-pharmtox-010611-134532
8. May JE, Donaldson C, Gynn L and Ruth Morse H. Chemotherapy-induced genotoxic damage to bone marrow cells: Long-term implications. Mutagenesis 2018 Sep; 33:241–2. DOI: 10.1093/mutage/gey014
9. Botchkarev VA. Molecular Mechanisms of Chemotherapy-Induced Hair Loss. Tech. rep. Boston: Department of Dermatology, Boston University School of Medicine, 2003 :72–3
10. Goodwin S, McPherson JD and McCombie WR. Coming of age: Ten years of next-generation sequencing technologies. 2016 Jun. DOI: 10.1038/nrg.2016.49
11. Kircher M and Kelso J. High-throughput DNA sequencing - Concepts and limitations. BioEssays 2010 Jun; 32:524–5. DOI: 10.1002/bies.200900181
12. Warr A, Robert C, Hume D, Archibald A, Deeb N and Watson M. Exome sequencing: Current and future perspectives. G3: Genes, Genomes, Genetics 2015; 5:1543–4. DOI: 10.1534/g3.115.018564
13. Horgan RP and Kenny LC. 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics. The Obstetrician & Gynaecologist 2011 Jul; 13:189–95. DOI: 10.1576/toag.13.3.189.27672
14. Hasin Y, Seldin M and Lusis A. Multi-omics approaches to disease. Genome Biology 2017 May; 18:1–2. DOI: 10.1186/s13059-017-1215-1
15. Kitano H. Systems Biology: A Brief Overview. SCIENCE 2002 Mar; 295:1662. Available from: http://www.stke.org/
16. Mariottini C and Iyengar R. System Biology of Cell Signaling. *Handbook of Systems Biology: Concepts and Insights*. Elsevier, 2012 Nov :311–9. DOI: 10.1016/B978-0-12-385944-0.00016-2
17. Alcocer-Cuarón C, Rivera AL and Castaño VM. Hierarchical structure of biological systems: A bioengineering approach. Bioengineered 2014 Oct; 5. DOI: 10.4161/bioe.26570
18. Barabási AL and Oltvai ZN. Network biology: Understanding the cell's functional organization. Nature Reviews Genetics 2004 Feb; 5:101–3. DOI: 10.1038/nrg1272
19. Barabási AL, Gulbahce N and Loscalzo J. Network medicine: A network-based approach to human disease. Nature Reviews Genetics 2011 Jan; 12:56–7. DOI: 10.1038/nrg2918
20. Moraes F and Góes A. A decade of human genome project conclusion: Scientific diffusion about our genome knowledge. Biochemistry and Molecular Biology Education 2016 May; 44:215. DOI: 10.1002/bmb.20952

21. Carvunis AR, Roth FP, Calderwood MA, Cusick ME, Superti-Furga G and Vidal M. Interactome Networks. *Handbook of Systems Biology: Concepts and Insights*. Elsevier, 2012 Nov :45–8. DOI: 10.1016/B978-0-12-385944-0.00003-4

22. Chu LH and Chen BS. Construction of a cancer-perturbed protein-protein interaction network for discovery of apoptosis drug targets. BMC Systems Biology 2008 Jun; 2:1–17. DOI: 10.1186/1752-0509-2-56

23. Li L, Zhang K, Lee J, Cordes S, Davis DP and Tang Z. Discovering cancer genes by integrating network and functional properties. BMC Medical Genomics 2009; 2:4–14. DOI: 10.1186/1755-8794-2-61

24. Oti M, Snel B, Huynen MA and Brunner HG. Predicting disease genes using protein-protein interactions. Journal of Medical Genetics 2006 Aug; 43:691–6. DOI: 10.1136/jmg.2006.041376

25. Bulyk ML and Walhout AJ. Gene Regulatory Networks. *Handbook of Systems Biology: Concepts and Insights*. Elsevier, 2012 Nov :65–88. DOI: 10.1016/B978-0-12-385944-0.00004-6

26. Palsson B. The challenges of in silico biology. Nature Biotechnology 2000; 18:1147–50. DOI: 10.1038/81125

27. Karlebach G and Shamir R. Modelling and analysis of gene regulatory networks. Nature Reviews Molecular Cell Biology 2008 Oct; 9:770–4. DOI: 10.1038/nrm2503

28. Albert R and Thakar J. Boolean modeling: A logic-based dynamic approach for understanding signaling and regulatory networks and for making useful predictions. Wiley Interdisciplinary Reviews: Systems Biology and Medicine 2014; 6:353–61. DOI: 10.1002/wsbm.1273

29. Davidich MI and Bornholdt S. Boolean Network Model Predicts Knockout Mutant Phenotypes of Fission Yeast. PLoS ONE 2013 Sep; 8:1–7. DOI: 10.1371/journal.pone.0071786

30. Flobak Å, Baudot A, Remy E, Thommesen L, Thieffry D, Kuiper M and Lægreid A. Discovery of Drug Synergies in Gastric CancerCells Predicted by Logical Modeling. PLoS Computational Biology 2015 Aug; 11:1–20. DOI: 10.1371/journal.pcbi.1004426

31. Hood L and Friend SH. Predictive, personalized, preventive, participatory (P4) cancer medicine. Nature Reviews Clinical Oncology 2011 Mar; 8:184–7. DOI: 10.1038/nrclinonc.2010.227

32. Al-Lazikani B, Banerji U and Workman P. Combinatorial drug therapy for cancer in the post-genomic era. 2012 Jul. DOI: 10.1038/nbt.2284

33. Loewe S. The Problem of Synergism and Antagonism of Combined Drugs. Tech. rep. Salt Lake City: University of Utah Collecge of Medicine, Department of Pharmacology, 1953 :285–9

34. Greco WR, Bravo G and Parsons JC. The Search for Synergy: A Critical Review from a Response Surface Perspective. Tech. rep. 1995 :332–3

35. National Cancer Institute. A to Z List of Cancer Drugs. Available from: https://www.cancer.gov/about-cancer/treatment/drugs

36. Jeon M, Kim S, Park S, Lee H and Kang J. In silico drug combination discovery for personalized cancer therapy. BMC Systems Biology 2018 Mar; 12:59–66. DOI: 10.1186/s12918-018-0546-1

37. Tsirvouli E, Touré V, Niederdorfer B, Vázquez M, Flobak Å and Kuiper M. A Middle-Out Modeling Strategy to Extend a Colon Cancer Logical Model Improves Drug Synergy Predictions in Epithelial-Derived Cancer Cell Lines. Frontiers in Molecular Biosciences 2020 Oct; 7. DOI: 10.3389/fmolb.2020.502573

38. Park JC, Jang SY, Lee D, Lee J, Kang U, Chang H, Kim HJ, Han SH, Seo J, Choi M, Lee DY, Byun MS, Yi D, Cho KH and Mook-Jung I. A logical network-based drug-screening platform for Alzheimer's disease representing pathological features of human brain organoids. Nature Communications 2021 Dec; 12:1–11. DOI: 10.1038/s41467-020-20440-5

39. Pirkl M, Hand E, Kube D and Spang R. Analyzing synergistic and non-synergistic interactions in signalling pathways using Boolean Nested Effect Models. Bioinformatics 2016 Mar; 32:893–9. DOI: 10.1093/bioinformatics/btv680

40. Strimbu K and Tavel JA. What are biomarkers? Tech. rep. 6. National Institute of Health, 2010 Nov :1–5. DOI: 10.1097/COH.0b013e32833ed177

41. The DrugLogics Initiative. Available from: https://druglogics.eu/

42. Flobak Å, Zobolas J, Vazquez M, Steigedal TS, Thommesen L, Grislingås A, Niederdorfer B, Folkesson E and Kuiper M. Logical modeling: Combining manual curation and automated parameterization to predict drug synergies. bioRxiv preprint server for Biology 2021 Jul :4–37. DOI: 10.1101/2021.06.28.450165. Available from: https://doi.org/10.1101/2021.06.28.450165

43. Wooten DJ, Meyer CT, Lubbock AL, Quaranta V and Lopez CF. MuSyC is a consensus framework that unifies multi-drug synergy metrics for combinatorial drug discovery. Nature Communications 2021 Dec; 12:1–4. DOI: 10.1038/s41467-021-24789-z

44. Fraser TR. On the antagonism between the actions of active substances. British Medical Journal 1872; 2:457–9. DOI: 10.1136/bmj.2.617.457

45. Lu J, Zeng H, Liang Z, Chen L, Zhang L, Zhang H, Liu H, Jiang H, Shen B, Huang M, Geng M, Spiegel S and Luo C. Network modelling reveals the mechanism underlying colitis-associated colon cancer and identifies novel combinatorial anti-cancer targets. Scientific Reports 2015 Oct; 5:1–13. DOI: 10.1038/srep14739

46. Jaaks P, Coker EA, Vis DJ, Edwards O, Carpenter EF, Leto SM, Dwane L, Sassi F, Lightfoot H, Barthorpe S, Meer D van der, Yang W, Beck A, Mironenko T, Hall C, Hall J, Mali I, Richardson L, Tolley C, Morris J, Thomas F, Lleshi E, Aben N, Benes CH, Bertotti A, Trusolino L, Wessels L and Garnett MJ. Effective drug combinations in breast, colon and pancreatic cancer cells. Nature 2022 Mar; 603:166–73. DOI: 10.1038/s41586-022-04437-2

47. Park SM, Hwang CY, Choi J, Joung CY and Cho KH. Feedback analysis identifies a combination target for overcoming adaptive resistance to targeted cancer therapy. Oncogene 2020 May; 39:3803–20. DOI: 10.1038/s41388-020-1255-y

48. Richter M, Piwocka O, Musielak M, Piotrowski I, Suchorska WM and Trzeciak T. From Donor to the Lab: A Fascinating Journey of Primary Cell Lines. Frontiers in Cell and Developmental Biology 2021 Jul; 9:1–3. DOI: 10.3389/fcell.2021.711381

49. Cell Model Passports. Available from: https://cellmodelpassports.sanger.ac.uk/

50. Casali PG and Vyas M. Data protection and research in the European Union: a major step forward, with a step back. Analysis of Oncology 2020; 32:15–8

51. Alvarez MJ, Shen Y, Giorgi FM, Lachmann A, Ding BB, Hilda Ye B and Califano A. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. Nature Genetics 2016 Aug; 48:838–45. DOI: 10.1038/ng.3593

52. Beal J, Montagud A, Traynard P, Barillot E and Calzone L. Personalization of logical models with multi-omics data allows clinical stratification of patients. Frontiers in Physiology 2019; 10:1–16. DOI: 10.3389/fphys.2018.01965

53. Badia-I-Mompel P, Vélez J, Braunger J, Geiss C, Dimitrov D, Müller-Dott S, Taus P, Dugourd A, Holland CH, Ramirez Flores RO and Saez-Rodriguez J. decoupleR: Ensemble of computational methods to infer biological activities from omics data. bioRxiv preprint server for Biology 2021 :1–21. DOI: 10.1101/2021.11.04.467271

54. Teschendorff AE and Wang N. Improved detection of tumor suppressor events in 1 single-cell RNA-Seq data. bioRxiv preprint server for Biology 2020 :2–15. DOI: 10.1101/2020.07.04.187781

55. Andrade C. Z Scores, Standard Scores, and Composite Test Scores Explained. Indian Journal of Psychological Medicine 2021 Nov; 43:555–7. DOI: 10.1177/02537176211046525

56. Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D and Saez-Rodriguez J. Benchmark and integration of resources for the estimation of human transcription factor activities. Genome Research 2019; 29:1363–75. DOI: 10.1101/gr.240663.118

57. Müller-Dott S, Tsirvouli E, Vázquez M, Ramirez Flores RO, Badia-I-Mompel P, Fallegger R, Laegreid A and Saez-Rodriguez J. Expanding the coverage of regulons from high-confidence prior knowledge for accurate estimation of transcription factor activities. bioRxiv preprint server for Biology 2023 Mar :1. DOI: 10.1101/2023.03.30.534849. Available from: https://doi.org/10.1101/2023.03.30.534849

58. Zhao Y, Li MC, Konaté MM, Chen L, Das B, Karlovich C, Williams PM, Evrard YA, Doroshow JH and McShane LM. TPM, FPKM, or Normalized Counts? A Comparative Study of Quantification Measures for the Analysis of RNA-seq Data from the NCI Patient-Derived Models Repository. Journal of Translational Medicine 2021 Dec; 19:1–4. DOI: 10.1186/s12967-021-02936-w

59. Giorgi FM, Ceraolo C and Mercatelli D. The R Language: An Engine for Bioinformatics and Data Science. 2022 May. DOI: 10.3390/life12050648

60. Kumar R and Antony GA. A Review of Methods and Applications of the ROC Curve in Clinical Trials. Biostatistics 2010; 44:659–70. Available from: https://doi.org/10.1177/009286151004400602

61. Fairus Mokhtar S, Yusof ZM and Sapiri H. Confidence Intervals by Bootstrapping Approach: A Significance Review. Malaysian Journal of Fundamental and Applied Sciences 2023; 19:30–42

62. Calcagno DQ, Leal MF, Assumpçaão PP, Smith MdAC and Burbano RR. MYC and gastric adenocarcinoma carcinogenesis. 2008 Oct. DOI: 10.3748/wjg.14.5962

63. Calcagno DQ, Freitas VM, Leal MF, Souza CR de, Demachki S, Montenegro R, Assumpção PP, Khayat AS, Smith MD, Santos AK dos and Burbano RR. MYC, FBXW7 and TP53 copy number variation and expression in Gastric Cancer. BMC Gastroenterology 2013 Sep; 13:1–9. DOI: 10.1186/1471-230X-13-141

64. Li X, Cheung KF, Ma X, Tian L, Zhao J, Go MY, Shen B, Cheng AS, Ying J, Tao Q, Sung JJ, Kung HF and Yu J. Epigenetic inactivation of paired box gene 5, a novel tumor suppressor gene, through direct upregulation of p53 is associated with prognosis in gastric cancer patients. Oncogene 2012 Jul; 31:3419–28. DOI: 10.1038/onc.2011.511

65. Wang Z and Sun Y. Targeting p53 for novel anticancer therapy. Translational Oncology 2010; 3:1. DOI: 10.1593/tlo.09250

66. Xu X, Liu Z, Tian F, Xu J and Chen Y. Clinical significance of transcription factor 7 (TCF7) as a prognostic factor in gastric cancer. Medical Science Monitor 2019; 25:3957–63. DOI: 10.12659/MSM.913913

67. Devereux TR, Stern MC, Flake GP, Yu MC, Zhang ZQ, London SJ and Taylor JA. CTNNB1 mutations and β-catenin protein accumulation in human hepatocellular carcinomas associated with high exposure to aflatoxin B1. Molecular Carcinogenesis 2001; 31:68–73. DOI: 10.1002/mc.1041

68. Fan C, Liu S, Zhao Y, Han Y, Yang L, Tao G, Li Q and Zhang L. Upregulation of miR-370 contributes to the progression of gastric carcinoma via suppression of FOXO1. Biomedicine and Pharmacotherapy 2013 Jul; 67:521–6. DOI: 10.1016/j.biopha.2013.04.014

69. Chen YH, Li CL, Chen WJ, Liu J and Wu HT. Diverse roles of FOXO family members in gastric cancer. World Journal of Gastrointestinal Oncology 2021; 13:1367–75. DOI: 10.4251/wjgo.v13.i10.1367

70. Liu Y, Ao X, Jia Y, Li X, Wang Y and Wang J. The FOXO family of transcription factors: key molecular players in gastric cancer. 2022 Jul. DOI: 10.1007/s00109-022-02219-x

71. Gao Y, Qi W, Sun L, Lv J, Qiu W and Liu S. FOXO3 Inhibits Human Gastric Adenocarcinoma (AGS) Cell Growth by Promoting Autophagy in an Acidic Microenvironment. Cellular Physiology and Biochemistry 2018 Sep; 49:335–46. DOI: 10.1159/000492884

72. Ko H, Kim JM, Kim SJ, Shim SH, Ha CH and Chang HI. Induction of apoptosis by genipin inhibits cell proliferation in AGS human gastric cancer cells via Egr1/p21 signaling pathway. Bioorganic and Medicinal Chemistry Letters 2015 Oct; 25:4191–5. DOI: 10.1016/j.bmcl.2015.08.005

73. Kim SJ, Kim JM, Shim SH and Chang HI. Shikonin induces cell cycle arrest in human gastric cancer (AGS) by early growth response 1 (Egr1)-mediated p21 gene expression. Journal of Ethnopharmacology 2014 Feb; 151:1064–71. DOI: 10.1016/j.jep.2013.11.055

74. Kundu JK, Shin YK and Surh YJ. Resveratrol modulates phorbol ester-induced pro-inflammatory signal transduction pathways in mouse skin in vivo: NF-κB and AP-1 as prime targets. Biochemical Pharmacology 2006 Nov; 72:1506–13. DOI: 10.1016/j.bcp.2006.08.005

75. Saito T and Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS ONE 2015 Mar; 10:1–21. DOI: 10.1371/journal.pone.0118432

76. Niederdorfer B, Touré V, Vazquez M, Thommesen L, Kuiper M, Lægreid A and Flobak Å. Strategies to Enhance Logic Modeling-Based Cell Line-Specific Drug Synergy Prediction. Frontiers in Physiology 2020 Jul; 11:1–15. DOI: 10.3389/fphys.2020.00862

77. Zobolas J, Kuiper M and Flobak Å. emba: R package for analysis and visualization of biomarkers in boolean model ensembles. Journal of Open Source Software 2020 Sep; 5:1–2. DOI: 10.21105/joss.02583

78. Martignetti L, Calzone L, Bonnet E, Barillot E and Zinovyev A. ROMA: Representation and Quantification of Module Activity from Target Expression Data. Frontiers in Genetics 2016 Feb; 7:1–11. DOI: 10.3389/fgene.2016.00018

79. Catlett NL, Bargnesi AJ, Anthony S, Seagaran T, Ladd W, Elliston KO and Pratt D. Reverse causal reasoning: applying qualitativecausal knowledge to the interpretation ofhigh-throughput data. BMC Bioinformatics 2013; 14:1–13

80. Peña-Chilet M, Esteban-Medina M, Falco MM, Rian K, Hidalgo MR, Loucera C and Dopazo J. Using mechanistic models for the clinical interpretation of complex genomic variation. Scientific Reports 2019 Dec; 9:1–8. DOI: 10.1038/s41598-019-55454-7

Appendix

A Versions of R packages, Databases and Tools

The versions of of R packages, tools and databases utilised in this project are displayed in Table A.1.

Table A.1: Versions of R packages, databases and tools utilised in this project.

Resource	Version
Bioconductor	3.15
CollecTRI	11.11.22
Cytoscape	3.8.2
DecoupleR R version	2.2.2
Docker	4.13.0
Drabme	1.2.1
Druglogics-synergy module	1.2.1
Gitsbe	1.3.1
Human DoRothEA	1.12.0
Java	8
Maven	3.6.0
R	4.2.2
RStudio	2022.12.0+353
Visual Studio Code	1.78.0

B Logical equations

Symbols used to represent logical operators:

- & = AND
- | = OR
- ! = NOT

B.1 CASCADE 1.0 Model

The logical formalism of the CASCADE 1.0 model by Flobak et al. [30] are displayed in Table B.1 and Table B.2.

Table B.1: Boolean rules of the CASCADE 1.0 model of Flobaek et al. [30].

Boolean rule
TAB1 = !p38alpha
CFLAR = AKT & !ITCH
IKKA = AKT
DKK1 = DKK1gene
pras40 = !AKT
Axin = !LRP
TCF = betacatenin | !NLK
cMYC = TCF
RTPKgene = FOXO
TSC2 = GSK3 & !(IKKB | AKT | RSK | ERK)
SFRP1gene = !cMYC
Caspase9 = CytochromeC
p53 = p38alpha & !MDM2
DKK1gene = TCF & !cMYC
GAB = GRB2 & !ERK
AKT = PDK1 | mTORC2
ASK1 = !AKT
RSK = ERK & PDK1
SHP2 = GAB
Ras = SOS | SHP2
MEKK4 = Rac
S6K = PDK1 | mTORC1
MKK3 = ASK1 | TAK1
PDK1 = PI3K & !PTEN
MEK = Raf | MAP3K8 | !ERK
DUSP1 = p38alpha | MSK
BAD = !AKT & !RSK
BAX = p53
TAK1 = TAB1
RTPK = (RTPKgene | MMP) & !(p38alpha | MEK)
CK1 = !LRP
Egr1 = !TCF
SOS = GRB2 | !ERK
BCL2 = !BAD
MKK7 = TAK1 | GRAP2
LRP = (Fz | ERK | JNK1 | p38alpha) & !DKK1
GRB2 = SHC1
MAP3K8 = IKKB
Caspase8 = !CFLAR
FOXO = !(AKT | NLK)

Table B.1 – *Continued from previous page*

Boolean rule
GSK3 = !(LRP | RSK | S6K | ERK | p38alpha | Dvl | AKT)
Raf = Ras | !(Rheb | AKT | ERK)
ITCH = JNK1
MLK3 = Rac
PTENgene = Egr1
p38alpha = (MKK3 | MKK4) & !DUSP1
IKKB = TAK1 & !p53
MSK = ERK | p38alpha
MDM2 = (AKT | MDM2gene) & !S6K
MDM2gene = NFkB | p53
DUSP6 = ERK | mTORC1
NFkB = IKKA | IKKB | MSK
JNK1 = (MKK7 | MKK4) & !DUSP1
ERK = MEK | !DUSP6
Rheb = !TSC2
Rac = Dvl | mTORC2
CytochromeC = BAX & !BCL2
betacatenin = IKKA | !betaTrCP
MKK4 = MEKK4 | MLK3 | TAK1 | GRAP2
mTORC2 = TSC2 & !S6K
SHC1 = RTPK | !PTEN
IRS1 = !(S6K | ERK | IKKB)
mTORC1 = (Rheb | RSK) & !pras40
NLK = TAK1
Dvl = Fz
betaTrCP = Axin & GSK3 & CK1
SFRP1 = SFRP1gene
Fz = !SFRP1
MMP = LEF
PI3K = GAB | IRS1 | Ras
LEF = betacatenin
PTEN = PTENgene & !GSK3
GRAP2 = !p38alpha

Table B.2: Multivalues Boolean rules of the CASCADE 1.0 model of Flobak et al. [30].

Multivalued rules
CCND1 = RSK + TCF
Caspase37 = Caspase8 + Caspase9
Antisurvival = Caspase37 + FOXO
Prosurvival = CCND1 + cMYC

B.2 Lu Model

The logical formalism of the original Lu et al. [45] model are displayed in Table B.3.

Table B.3: Boolean rules of the original Lu et al. model [45].

Boolean rule
ATM = ROS
ASK1 = ROS & !P21
AKT = PI3K & ! (PP2A | CASP3)
BAX = ((TBID | P53) & PP2A) & !AKT
BCATENIN = !(GSK3B | APC)
BCL2 = (STAT3 | NFKB) & !(P53 | PP2A)

Table B.3 – *Continued from previous page*

Boolean rule
CASP3 = (CASP8 | CASP9) & !IAP
CASP8 = FADD &! (CFLIP | P21)
CASP9 = CYTC & !(IAP | P21)
CERAMIDE = SMASE & !SPHK1
CFLIP = NFKB
COX2 = S1P & TNFR
CYCLIND1 = (BCATENIN | STAT3 | JUN) & !GSK3B
CYTC = MOMP
EP2 = PGE2
ERK = MEK
FAS = CTL
FADD = TNFR | FAS
FOS = ERK
GP130 = IL6
GSK3B = !(EP2 | AKT)
IAP = (NFKB | STAT3) & !SMAC
IKB = !IKK
IKK = (AKT | (S1P&TNFR))
JAK = GP130 & !SOCS
JNK = ASK1 | MEKK1
JUN = ((BCATENIN | ERK) & JNK) & !GSK3B
MDM2 = (P53 & AKT) & !(GSK3B | ATM)
MEK = RAF | ROS
MEKK1 = CERAMIDE | TGFR | TNFR
MOMP = (BAX | TBID | CERAMIDE) & !BCL2
NFKB = !IKB
P21 = (P53 | SMAD) & !(GSK3B | CASP3)
P53 = (PTEN | JNK | ATM) & !MDM2
PGE2 = COX2
PI3K = EP2 | RAS &!PTEN
PP2A = CERAMIDE & !AKT
PTEN = P53 & !(NFKB | JUN)
RAF = CERAMIDE | RAS
RAS = EP2 | GP130
ROS = TNFR &! SOD
SOD = NFKB | STAT3
S1P = SPHK1
SMAC = MOMP
SMAD = TGFR & !JUN
SMAD7 = SMAD | NFKB
SMASE = P53 | FADD
SPHK1 = ERK | TNFR
STAT3 = JAK
SOCS = STAT3
TBID = CASP8 & !BCL2
TGFR = TGFB &!SMAD7
TNFR = TNFA
TREG = (IL10 | DC) &! IL6
TNFA = MAC
TH2 = IL4 &! (IFNG | TGFB)
TH1 = (IL12 | IFNG) &! (IL10 | TGFB | IL4)
TGFB = TREG
MAC = (IFNG | CCL2) &! IL10
IL6 = MAC | DC | NFKB
IL4 = DC | TH2
IL12 = DC | MAC
IL10 = TREG | TH2
IFNG = TH1 | CTL
CTL = IFNG &! TGFB

Table B.3 – *Continued from previous page*

Boolean rule
DC = (CCL2 | TNFA) &! IL10
CCL2 = NFKB
Prosurvival = (FOS & CYCLIND1) & !(P21 | CASP3)
Antisurvival = CASP3

The logical formalism of the modified Lu et al. [45] model are displayed in Table B.4.

Table B.4: Boolean rules of the modified Lu et al. model [45].

Boolean rule
ATM = ROS
ASK1 = ROS & !P21
AKT = PI3K & ! (PP2A | CASP3)
BAX = ((TBID | P53) & PP2A) & !AKT
BCATENIN = !(GSK3B | APC)
BCL2 = (STAT3 | NFKB) & !(P53 | PP2A)
CASP3 = (CASP8 | CASP9) & !IAP
CASP8 = FADD &! (CFLIP | P21)
CASP9 = CYTC & !(IAP | P21)
CERAMIDE = SMASE & !SPHK1
CFLIP = NFKB
COX2 = S1P & TNFR
CYCLIND1 = (BCATENIN | STAT3 | JUN) & !GSK3B
CYTC = MOMP
ERK = MEK
FADD = TNFR | FAS
FOS = ERK
GP130 = IL6
GSK3B = !(EP2 | AKT)
IAP = (NFKB | STAT3) & !SMAC
IKB = !IKK
IKK = (AKT | (S1P & TNFR))
JAK = GP130 & !SOCS
JNK = ASK1 | MEKK1
JUN = ((BCATENIN | ERK) & JNK) & !GSK3B
MDM2 = (P53 & AKT) & !(GSK3B | ATM)
MEK = RAF | ROS
MEKK1 = CERAMIDE | TGFR | TNFR
MOMP = (BAX | TBID | CERAMIDE) & !BCL2
NFKB = !IKB
P21 = (P53 | SMAD) & !(GSK3B | CASP3)
P53 = (PTEN | JNK | ATM) & !MDM2
PI3K = EP2 | RAS &!PTEN
PP2A = CERAMIDE & !AKT
PTEN = P53 & !(NFKB | JUN)
RAF = CERAMIDE | RAS
RAS = EP2 | GP130
ROS = TNFR &! SOD
SOD = NFKB | STAT3
S1P = SPHK1
SMAC = MOMP
SMAD = TGFR & !JUN
SMAD7 = SMAD | NFKB
SMASE = P53 | FADD
SPHK1 = ERK | TNFR
STAT3 = JAK
SOCS = STAT3
TBID = CASP8 & !BCL2

Table B.4 – *Continued from previous page*

Boolean rule
TGFR = !SMAD7
IL6 = NFKB

B.3 Park Model

The logical formalism of the original Park et al. [47] model are displayed in Table B.5.

Table B.5: Boolean rules of the original Park et al. model [47].

Boolean rule
Akt = mTOR2 & PDK1 & !PP2A
AMPK = p53
AP_1 = ATF2 | beta_catenin | CREB | ELK1 | ERK | JNK | PP2A | p90RSK & !GSK3beta
APAF1 = CytC | FOXO3 | E2F1
APC = !Frizzled
ATF2 = JNK | p38 | ATM
ATM = DNA_damage | E2F1 & !(CyclinG | Wip1)
AXIN = beta_catenin
Bax = c_Myc | JNK | p53 | Smad2_3_4 | GSK3beta | PP2A | FOXO3 & ! (Akt | ERK | S6K | p90RSK | Bcl_2)
Bcl_2 = STAT | Akt | ERK | CREB | p90RSK & !(JNK | p53 | Smad2_3_4 | PP2A)
beta_catenin = !APC | !(APC | (AXIN& GSK3beta) | p53)
BRaf = Ras | Src & !(ERK | Akt)
BRCA1 = CHK2 | ATM | E2F1 & !Akt
c_Myc = mTOR1 | beta_catenin | ERK | p38 & !(FOXO3 | GSK3beta | Smad2_3_4 | PP2A)
CASP3 = CASP9& !XIAP
CASP9 = CytC | APAF1 & !(Akt | XIAP)
Cdc25 = !(CHK1 | CHK2) | PP2A | MK2
CHK1 = BRCA1& !Akt
CHK2 = ATM | DNA_PK
CRaf = Ras | Src & !ERK | Akt
CREB = Akt | p90RSK | MSK& !GSK3beta
CyclinA = E2F1& CyclinE & !(Rb | p21 | p27)
CyclinB = Cdc25 & CyclinA)& !CyclinB
CyclinD = AP_1 | beta_catenin | c_Myc | S6K & !(FOXO3 | GSK3beta | p15)
CyclinE = !(CyclinA | CyclinB) & Cdc25 | E2F1 | CyclinD & !(p21 | p27)
CyclinG = !p53
CytC = Bax
DNA_damage = DNA_damage
DNA_PK = DNA_damage
E2F1 = c_Myc | ATM | CHK2 | MK2 & !p14ARF & !(CyclinA | CyclinB | Rb)
EGF = EGF
EGFR = EGF | Src & ! (ERK | c_Myc)
ELK1 = ERK | JNK | p38
ERK = MEK & !(MKPs | PP2A)
FOXO3 = JNK | p38 & !(Akt | ERK)
Frizzled = WNT
GAB1 = EGFR | Grb2 & !ERK
GADD45 = Smad2_3 | (p53 & DNA_damage)
Grb2 = Shc | IRS & !SPRY
GSK3beta = PP2A & !(Akt | S6K | Frizzled | p38 | p90RS)
IRS = IGFR &!(S6K | mTOR1 | JNK)
JAK = IGFR
JNK = MKK4 & MKK7 & !(MKPs | PP2A)
MDM2 = Akt | MDMX | p53 | Wip1 & !(ATM | CyclinE | GSK3beta | p14ARF | Rb | CyclinG | PP2A)
MDMX = Akt | Wip1 & !(ATM | MDM2 | p14ARF)
MEK = BRaf | CRaf & !(PP2A | ERK)
MEKK1 = Grb2 | Shc | Ras

Table B.5 – *Continued from previous page*

Boolean rule
MEKK2 = EGFR & Src
MEKK3 = TAK1& !(PP2A | GAB1)
MEKK4 = GADD45 &!(GSK3beta)
MK2 = p38
MKK3_6 = MEKK3 | MEKK4 | MLKs | TAK1 | TAO & !(Akt | PP2A)
MKK4 = MEKK1 | MEKK2 | MEKK4 | TAK1 | TAO | MLKs & !(Akt | PP2A)
MKK7 = MEKK1 | MEKK2 | TAK1 | TAO | MLKs & !GADD45
MKPs = AP_1 | CREB | ERK | JNK | p38
MLKs = JNK & !Akt
MSK = ERK | p38
mTOR1 = !TSC1_2
mTOR2 = PI3K | Akt & !S6K
p120RasGAP = EGFR & !(SHP2 | Src)
p14ARF = c_Myc | E2F1 | DNA_damage & !(p53 | Wip1)
p15 = Smad2_3_4 & !c_Myc
p21 = p53 | Smad2_3_4 & !(GSK3beta | Akt | c_Myc | MDM2)
p27 = FOXO3 | GSK3beta | Smad2_3_4 & !(Akt | c_Myc | Ras | CyclinD)
p38 = MKK3_6 | MKK4 & !(MKPs | Wip1)
p53 = ATM | JNK | p38 | CHK2 | CHK1 | p90RSK | DNA_PK & !(Bcl_2 | MDM2 | MDMX)
p90RSK = ERK
PDK1 = PIP3
PI3K = EGFR & GAB1 | Ras | Src | IRS & !S6K
PIP3 = PI3K & !PTEN
PP2A = p38 | TGFR & !(mTOR1 | Src)
PTEN = FOXO3 | p53 & !(AP_1 | GSK3beta)
Rac1 = Ras
Ras = Sos & !p120RasGAP
Rb = ATM | !((CyclinD& CyclinE) | CyclinA | CyclinB)
IGFR = Src | FOXO3 & !(c_Myc | mTOR1)
S6K = mTOR1 & PDK1 & !PP2A
Shc = EGFR | IGFR | Src & !(PTEN | PP2A)
SHP2 = GAB1
SMAC = Bax & !Bcl_2
Smad2_3_4 = Smad2_3
Smad2_3 = p38 & JNK | TGFR & !(Smad7 | ERK)
Smad7 = Smad2_3
Sos = Grb2& SHP2 & !(ERK& p90RSK)
SPRY = ERK & !SHP2
Src = EGFR | IGFR & !Src
STAT = JAK | Src
TAK1 = TGFR& !PP2A
TAO = ATM
TGF_beta = TGF_beta
TGFR = TGF_beta & !(Smad7 | ERK)
TSC1_2 = GSK3beta | AMPK & !(ERK | p90RSK | Akt | MK2)
Wip1 = p53
WNT = WNT
XIAP = Akt | STAT & !SMAC

The logical formalism of the modified Park et al. [47] model are displayed in Table B.6.

Table B.6: Boolean rules of the modified Park et al. model [47].

Boolean rule
Akt = mTOR2 & PDK1 & !PP2A
AMPK = p53
AP_1 = ATF2 | beta_catenin | CREB | ELK1 | ERK | JNK | PP2A | p90RSK & !GSK3beta
APAF1 = CytC | FOXO3 | E2F1

Table B.6 – *Continued from previous page*

Boolean rule
APC = !Frizzled
ATF2 = JNK | p38 | ATM
ATM = E2F1 & !(CyclinG | Wip1)
AXIN = beta_catenin
Bax = c_Myc | JNK | p53 | Smad2_3_4 | GSK3beta | PP2A | FOXO3 & ! (Akt | ERK | S6K | p90RSK | Bcl_2)
Bcl_2 = STAT | Akt | ERK | CREB | p90RSK & !(JNK | p53 | Smad2_3_4 | PP2A)
beta_catenin = !APC | !(APC | (AXIN& GSK3beta) | p53)
BRaf = Ras | Src & !(ERK | Akt)
BRCA1 = CHK2 | ATM | E2F1 & !Akt
c_Myc = mTOR1 | beta_catenin | ERK | p38 & !(FOXO3 | GSK3beta | Smad2_3_4 | PP2A)
CASP3 = CASP9& !XIAP
CASP9 = CytC | APAF1 & !(Akt | XIAP)
Cdc25 = !(CHK1 | CHK2) | PP2A | MK2
CHK1 = BRCA1& !Akt
CHK2 = ATM | DNA_PK
CRaf = Ras | Src & !ERK | Akt
CREB = Akt | p90RSK | MSK& !GSK3beta
CyclinA = E2F1& CyclinE & !(Rb | p21 | p27)
CyclinB = Cdc25 & CyclinA)& !CyclinB
CyclinD = AP_1 | beta_catenin | c_Myc | S6K & !(FOXO3 | GSK3beta | p15)
CyclinE = !(CyclinA | CyclinB) & Cdc25 | E2F1 | CyclinD & !(p21 | p27)
CyclinG = !p53
CytC = Bax
E2F1 = c_Myc | ATM | CHK2 | MK2 & !p14ARF & !(CyclinA | CyclinB | Rb)
EGFR = Src & ! (ERK | c_Myc)
ELK1 = ERK | JNK | p38
ERK = MEK & !(MKPs | PP2A)
FOXO3 = JNK | p38 & !(Akt | ERK)
GAB1 = EGFR | Grb2 & !ERK
GADD45 = Smad2_3 | p53 &
Grb2 = Shc | IRS & !SPRY
GSK3beta = PP2A & !(Akt | S6K | Frizzled | p38 | p90RS)
IRS = IGFR &!(S6K | mTOR1 | JNK)
JAK = IGFR
JNK = MKK4 & MKK7 & !(MKPs | PP2A)
MDM2 = Akt | MDMX | p53 | Wip1 & !(ATM | CyclinE | GSK3beta | p14ARF | Rb | CyclinG | PP2A)
MDMX = Akt | Wip1 & !(ATM | MDM2 | p14ARF)
MEK = BRaf | CRaf & !(PP2A | ERK)
MEKK1 = Grb2 | Shc | Ras
MEKK2 = EGFR & Src
MEKK3 = TAK1& !(PP2A | GAB1)
MEKK4 = GADD45 &!(GSK3beta)
MK2 = p38
MKK3_6 = MEKK3 | MEKK4 | MLKs | TAK1 | TAO & !(Akt | PP2A)
MKK4 = MEKK1 | MEKK2 | MEKK4 | TAK1 | TAO | MLKs & !(Akt | PP2A)
MKK7 = MEKK1 | MEKK2 | TAK1 | TAO | MLKs & !GADD45
MKPs = AP_1 | CREB | ERK | JNK | p38
MLKs = JNK & !Akt
MSK = ERK | p38
mTOR1 = !TSC1_2
mTOR2 = PI3K | Akt & !S6K
p120RasGAP = EGFR & !(SHP2 | Src)
p14ARF = c_Myc | E2F1 & !(p53 | Wip1)
p15 = Smad2_3_4 & !c_Myc
p21 = p53 | Smad2_3_4 & !(GSK3beta | Akt | c_Myc | MDM2)
p27 = FOXO3 | GSK3beta | Smad2_3_4 & !(Akt | c_Myc | Ras | CyclinD)
p38 = MKK3_6 | MKK4 & !(MKPs | Wip1)
p53 = ATM | JNK | p38 | CHK2 | CHK1 | p90RSK | DNA_PK & !(Bcl_2 | MDM2 | MDMX)
p90RSK = ERK

Table B.6 – *Continued from previous page*

Boolean rule
PDK1 = PIP3
PI3K = EGFR & GAB1 | Ras | Src | IRS & !S6K
PIP3 = PI3K & !PTEN
PP2A = p38 | TGFR & !(mTOR1 | Src)
PTEN = FOXO3 | p53 & !(AP_1 | GSK3beta)
Ras = Sos & !p120RasGAP
Rb = ATM | !((CyclinD& CyclinE) | CyclinA | CyclinB)
IGFR = Src | FOXO3 & !(c_Myc | mTOR1)
S6K = mTOR1 & PDK1 & !PP2A
Shc = EGFR | IGFR | Src & !(PTEN | PP2A)
SHP2 = GAB1
SMAC = Bax & !Bcl_2
Smad2_3_4 = Smad2_3
Smad2_3 = p38 & JNK | TGFR & !(Smad7 | ERK)
Smad7 = Smad2_3
Sos = Grb2& SHP2 & !(ERK& p90RSK)
SPRY = ERK & !SHP2
Src = EGFR | IGFR & !Src
STAT = JAK | Src
TAK1 = TGFR& !PP2A
TAO = ATM
TGFR = !(Smad7 | ERK)
TSC1_2 = GSK3beta | AMPK & !(ERK | p90RSK | Akt | MK2)
Wip1 = p53
XIAP = Akt | STAT & !SMAC

C DecoupleR Tools and Benchmark Results

A complete list of the decoupleR tools, and their ranking in the benchmarking analysis of Badia-I-Mompel et al. [53] is displayed in Table C.1 and Table C.2, respectively.

Table C.1: The software tools in the decoupleR ensemble. The table is collected from Badia-I-Mompel et al. [53].

Name	Sign	Weight	Permutation	P-value	Range
AUCell	No	No	No	No	0,1
UDT	Yes	Yes	No	No	0,Inf
MDT	Yes	Yes	Yes	No	0,Inf
FGSEA	No	No	Yes	Yes	0,1
GSVA	No	No	No	No	-1,+1
WSUM	Yes	Yes	Yes	Yes	-Inf,+Inf
WMEAN	Yes	Yes	Yes	Yes	-Inf,+Inf
ORA	No	No	No	Yes	0,Inf
ULM	Yes	Yes	No	Yes	-Inf,+Inf
MLM	Yes	Yes	No	Yes	-Inf,+Inf
VIPER	Yes	Yes	Yes	Yes	-Inf,+Inf
CONSENSUS	No	No	No	Yes	0,Inf

Table C.2: DecoupleR tools ranked by their performance in the benchmarking of Badia-I-Mompel et al. [53]. The results are ranked based on the median AUC of the joint distribution of AUC ROCs and AUC PRs.

Method	P-value	Median AUC
CONSENSUS	$<$ 2.2e-16	0.68
MLM	$<$ 2.2e-16	0.67
ULM	$<$ 2.2e-16	0.66
Norm WMEAN/Norm WSUM	$<$ 2.2e-16	0.65
ORA	$<$ 2.2e-16	0.64
Corr WSUM	$<$ 2.2e-16	0.64
UDT	$<$ 2.2e-16	0.64
MDT	5.76e-5	0.63
WSUM	0.144	0.64
VIPER	1	0.62
AUCell	1	0.62
Corr WMEAN	1	0.62
WMEAN	1	0.60
FGSEA	1	0.59
Norm FGSEA	1	0.58
GSVA	1	0.56

D Step-by-Step Workflow of Creating Calibration Data for the DrugLogics Pipeline

D.1 DecoupleR Tools Workflow

The scripts for executing the decoupleR tools MLM, ULM and CONSENSUS can be found *here*.

A detailed step by step explanation of the extraction of biomarker activity by the decoupleR tools, translating the entities to the node naming in the respective models, binarising the activity levels, and then fitting the output to the format of the DrugLogics pipeline is as follows:

- The molecular matrix with omics data from Cell Model Passports(CMP) were imported and read as a *.csv file. A molecular measure was selected, respectively ”TMP value”, ”Read count” or ”FPKM value”, and the data was stored as a data frame.
- A regulatory network was imported, either the DoRothEA network or the CollecTRI network, and confidence levels A, B and C were selected for. The network was also stored as a dataframe.
- The decoupleR tools cannot run with missing values so a check to see if there was any in the data was performed. If present, NA’s (numbers that are not available) and Infs (infinity numbers) were removed.
- A decoupleR tool was selected (MLM, ULM or CONSENSUS) and the tool was utilised with the selected network, a source (tf), a target (target), a mor (mor) and a minsize (5).
- The resulting output were filtered by a selected p-value threshold (0.05, 0.2 or no threshould).
- The filtered output were scaled between 0 and 1 and then subsequently binarised.
- Nodes in the output were converted from HGNC symbols to match the node names in the given model, by using a separate script.
- Finally, the output were fitted to match the output format of the DrugLogics pipeline by using a specialised function.

D.2 PROFILE Workflow

The scripts for executing PROFILE can be found *here*.

A step by step explanation of the changes made to the ’OncoLogics pilot.Rmd’ script from the PROFILE pipeline for extraction of biomarker activity and then fitting the output to the format of the DrugLogics pipeline is as follows:

- A model was selected at the top of the .Rmd file (line 11), either ’Flobak’, ’Lu’ or ’Park’.
- The cancer type, namely ’Gastric Carcinoma’ or ’Colorectal Carcinoma’ was selected for (in line 57).
- The cell lines were selected, namely ”HCT-116”, ”AGS”, ”COLO_205”, ”SW48 or ”SW620”.
- The specific cell line profiles were selected from the output of PROFILE and each activity was matched to a node name by restructuring the output data as a table.
- Finally, the output were fitted to match the output format of the DrugLogics pipeline by using a specialised function.

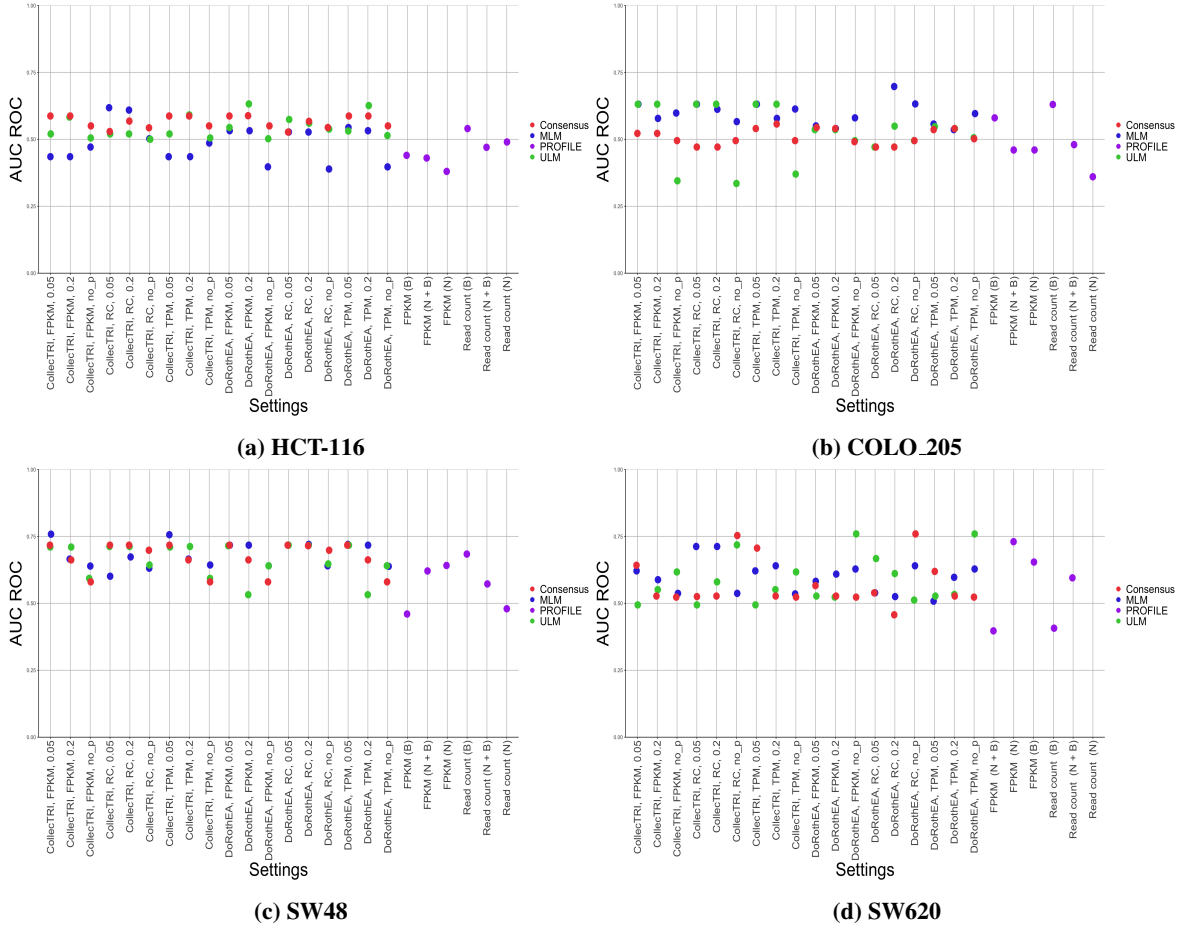
E Synergy and Statistical Results of Lu and Park Models

All figures related to the synergy results and the statistical analysis conducted for the Lu model [45] and the Park model [47] with the HCT-116, COLO_205, SW48 and SW620 cell lines can be found here.

E.1 Synergy Results - Tool Performances

Figure E.1 and Figure E.2 displays the obtained AUC ROC values from utilising the generated training data files with the DrugLogics pipeline with the Park and the Lu model, respectively. Four cell lines were utilised with both models: HCT-116, COLO_205, SW48 and SW620. Each column represents a different set of settings. The dots are coloured based on inference tools utilised.

Park Model

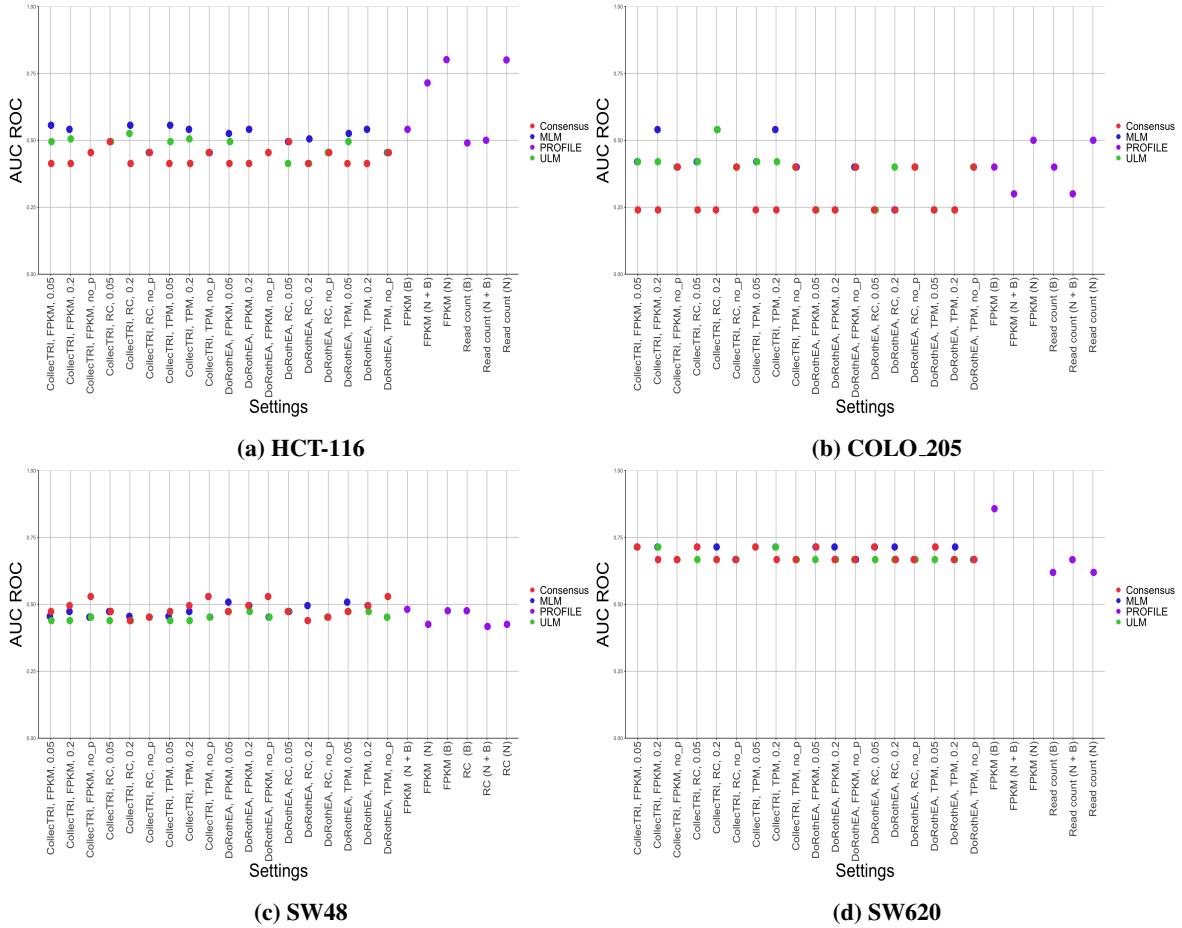


(a) HCT-116 **(b) COLO_205**

(c) SW48 **(d) SW620**

Figure E.1: Synergy results of HCT-116, COLO_205, SW48 and SW620 with the Park model.

Lu Model



(a) HCT-116 **(b) COLO_205**

(c) SW48 **(d) SW620**

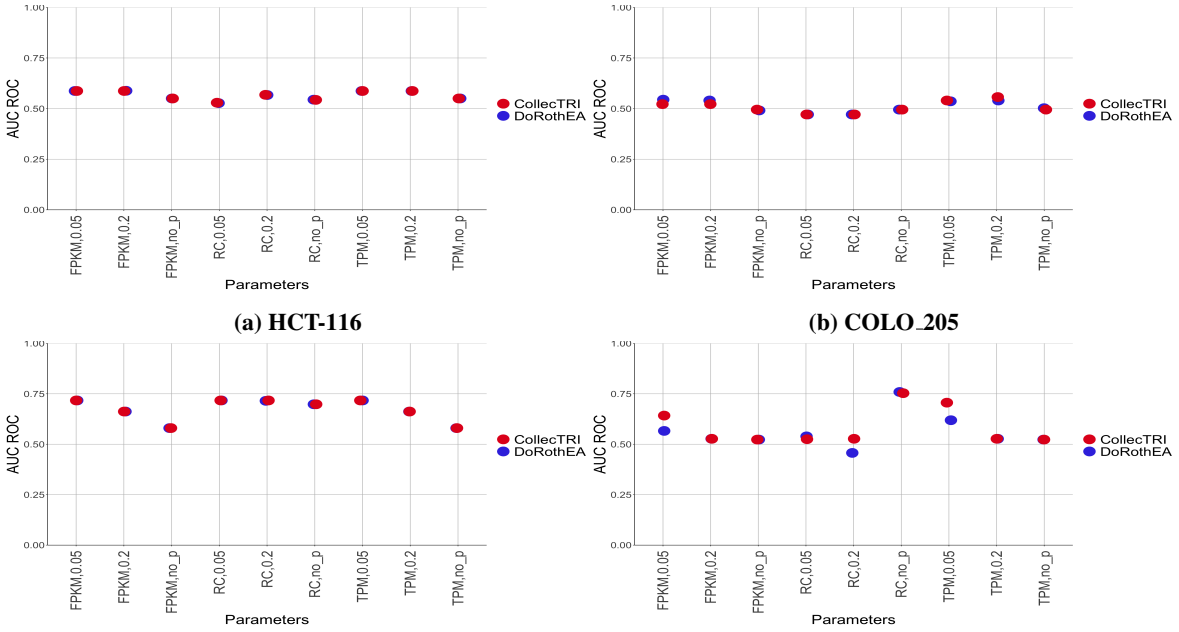
Figure E.2: Synergy results of HCT-116, COLO_205, SW48 and SW620 with the Lu model.

E.2 Synergy Results - Tool Parameters

Figure E.3 - E.24 displays the synergy results of the Lu and the Park models with each respective cell line, highlighted for each software tool and parameter settings.

E.2.1 CONSENSUS

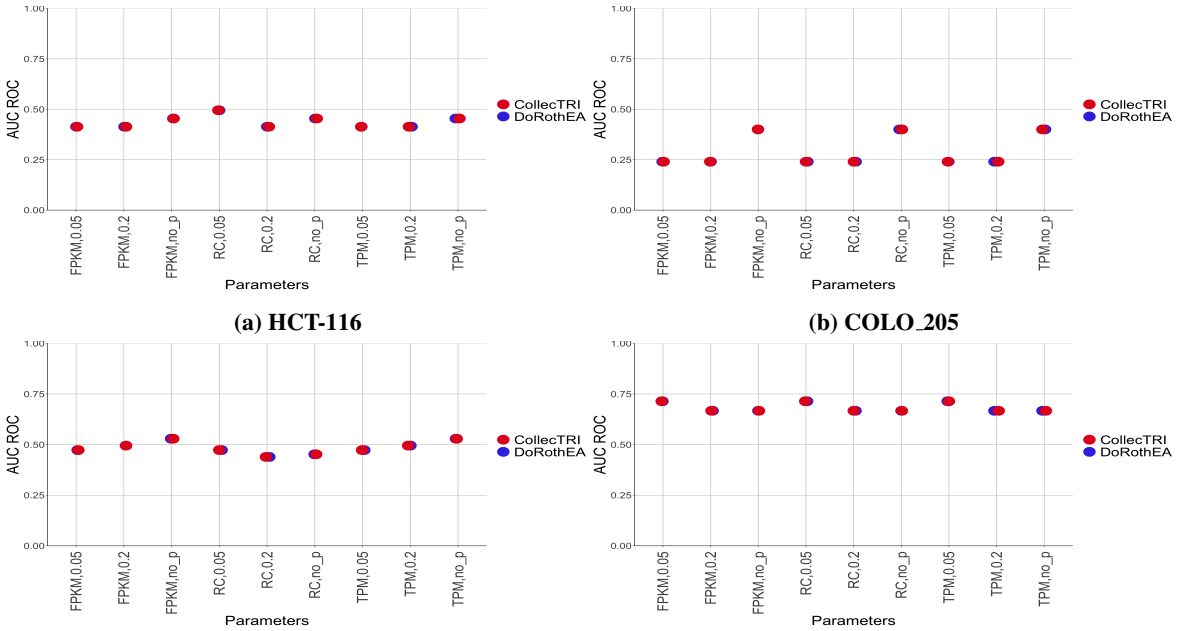
Gene Regulatory Networks



(a) HCT-116 **(b) COLO_205**

(c) SW48 **(d) SW620**

Figure E.3: Synergy results of HCT-116, COLO_205, SW48 and SW620 with the Park model.

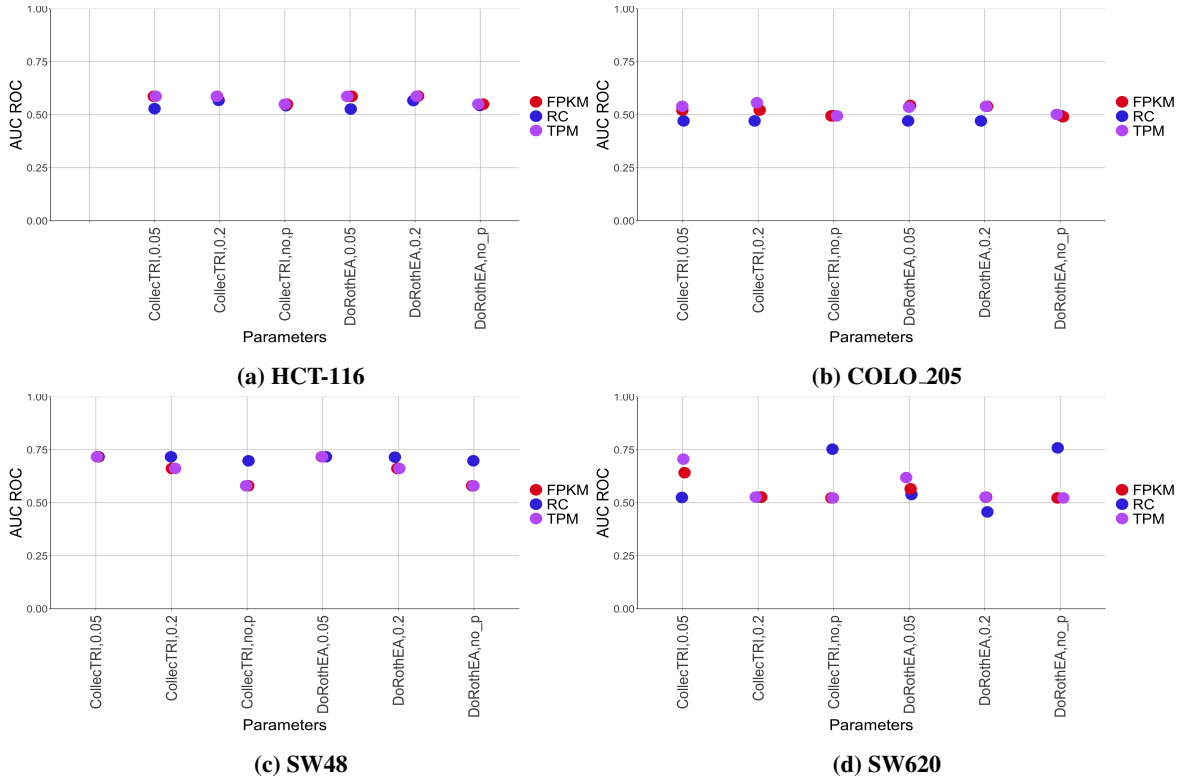


(a) HCT-116 **(b) COLO_205**

(c) SW48 **(d) SW620**

Figure E.4: Synergy results of HCT-116, COLO_205, SW48 and SW620 with the Lu model.

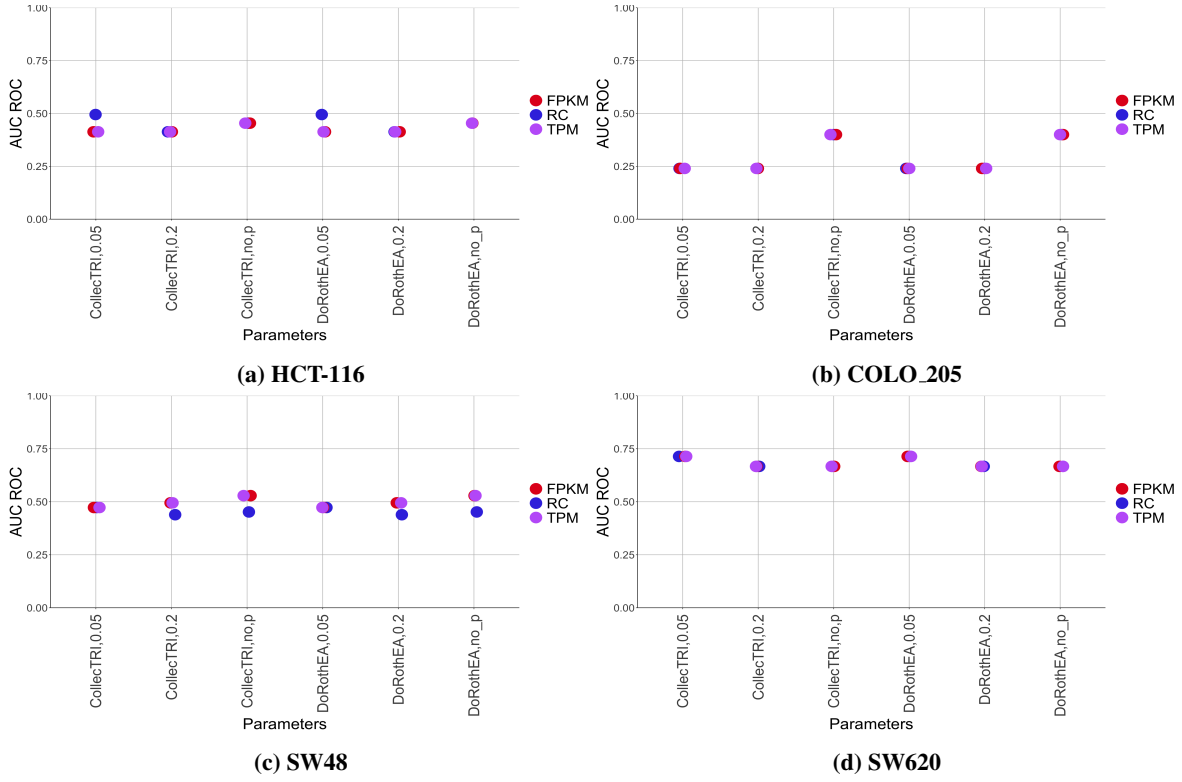
Expression Count Measures



(a) HCT-116 **(b) COLO_205**

(c) SW48 **(d) SW620**

Figure E.5: Synergy results of HCT-116, COLO_205, SW48 and SW620 with the Park model.

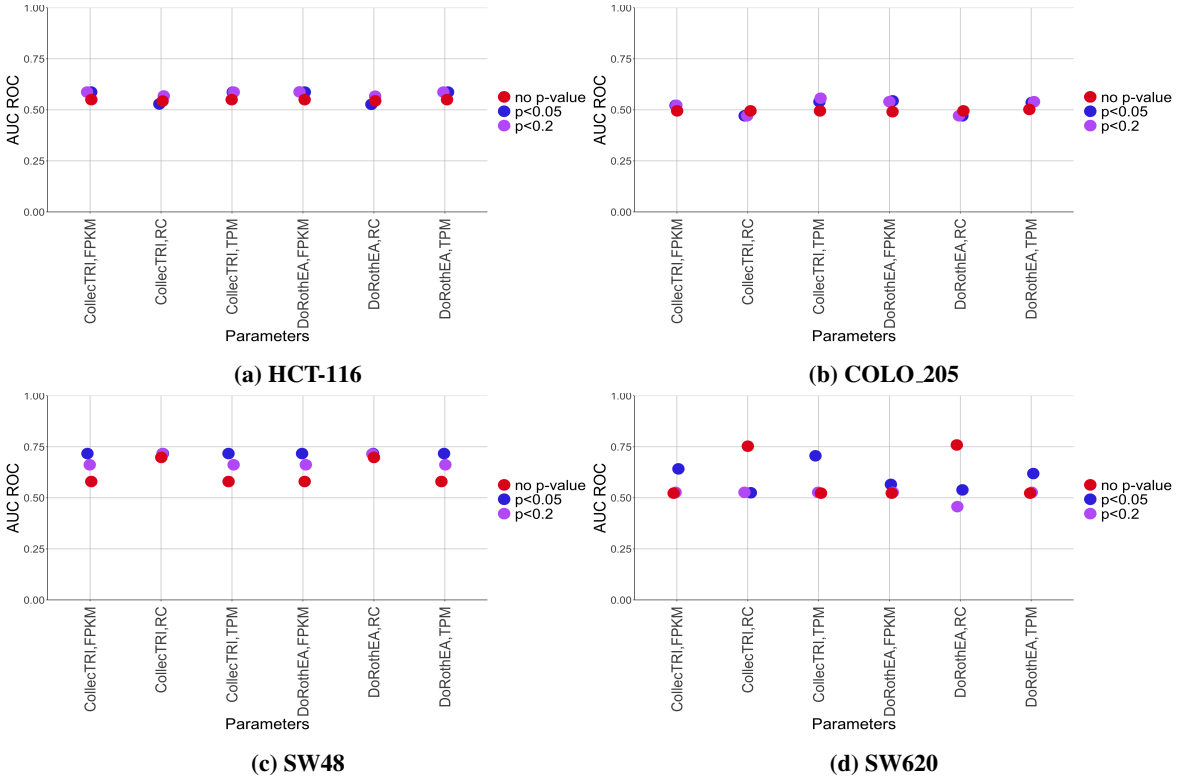


(a) HCT-116 **(b) COLO_205**

(c) SW48 **(d) SW620**

Figure E.6: Synergy results of HCT-116, COLO_205, SW48 and SW620 with the Lu model.

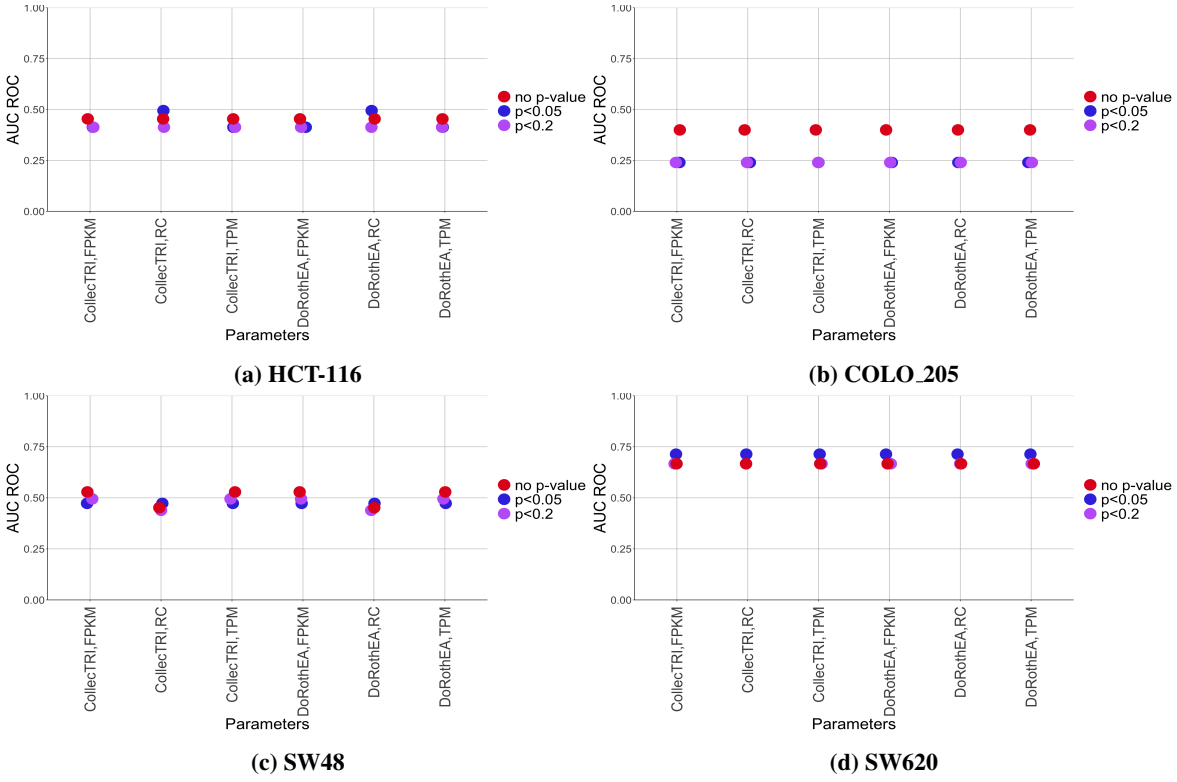
P-values



(a) HCT-116 **(b) COLO_205**

(c) SW48 **(d) SW620**

Figure E.7: Synergy results of HCT-116, COLO_205, SW48 and SW620 with the Park model.



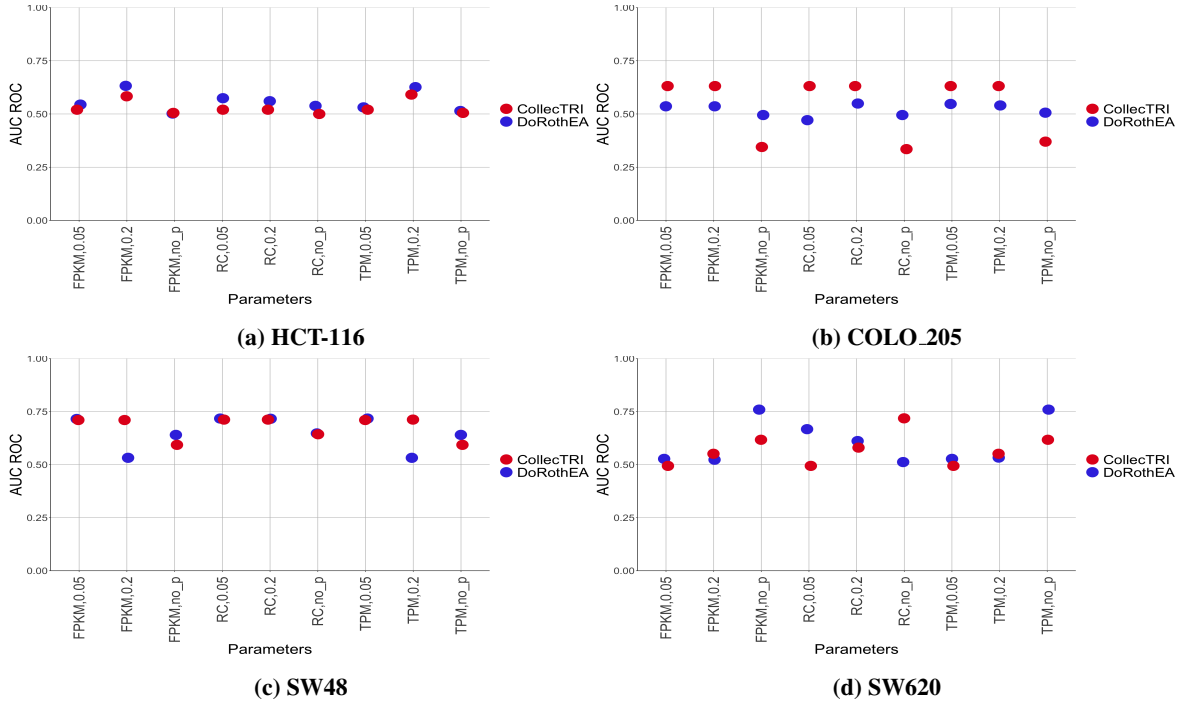
(a) HCT-116 **(b) COLO_205**

(c) SW48 **(d) SW620**

Figure E.8: Synergy results of HCT-116, COLO_205, SW48 and SW620 with the Lu model.

E.2.2 ULM

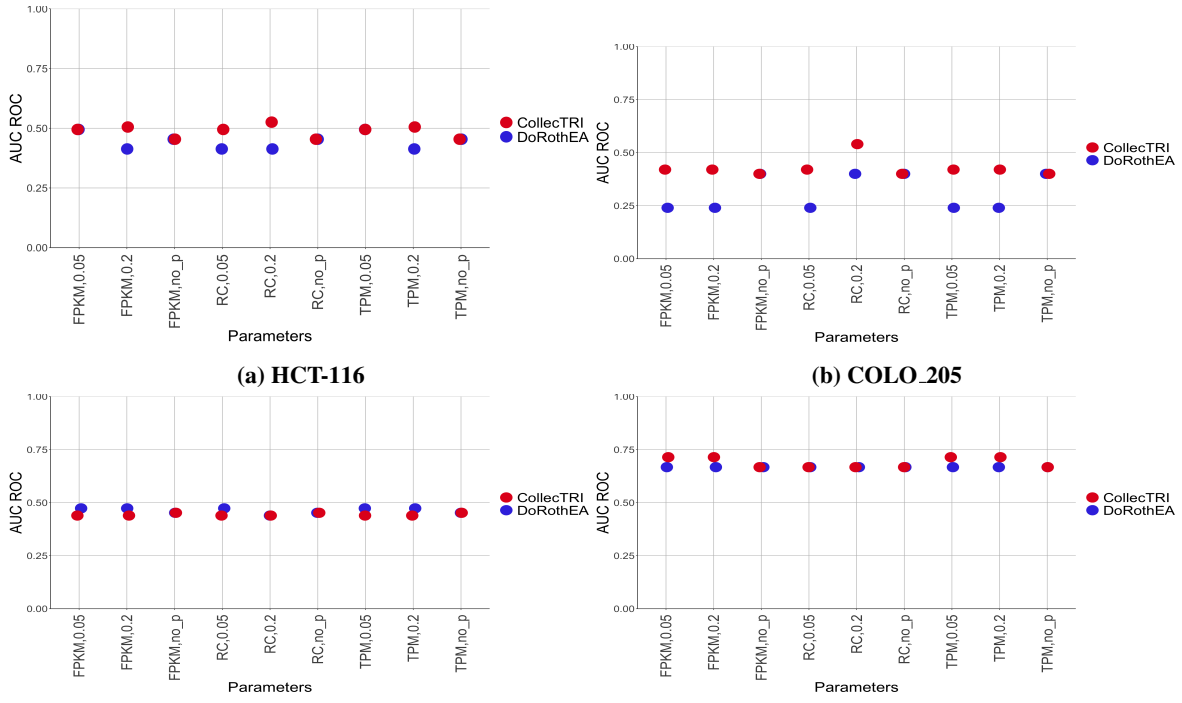
Gene Regulatory Networks



(a) HCT-116 **(b) COLO_205**

(c) SW48 **(d) SW620**

Figure E.9: Synergy results of HCT-116, COLO_205, SW48 and SW620 with the Park model.

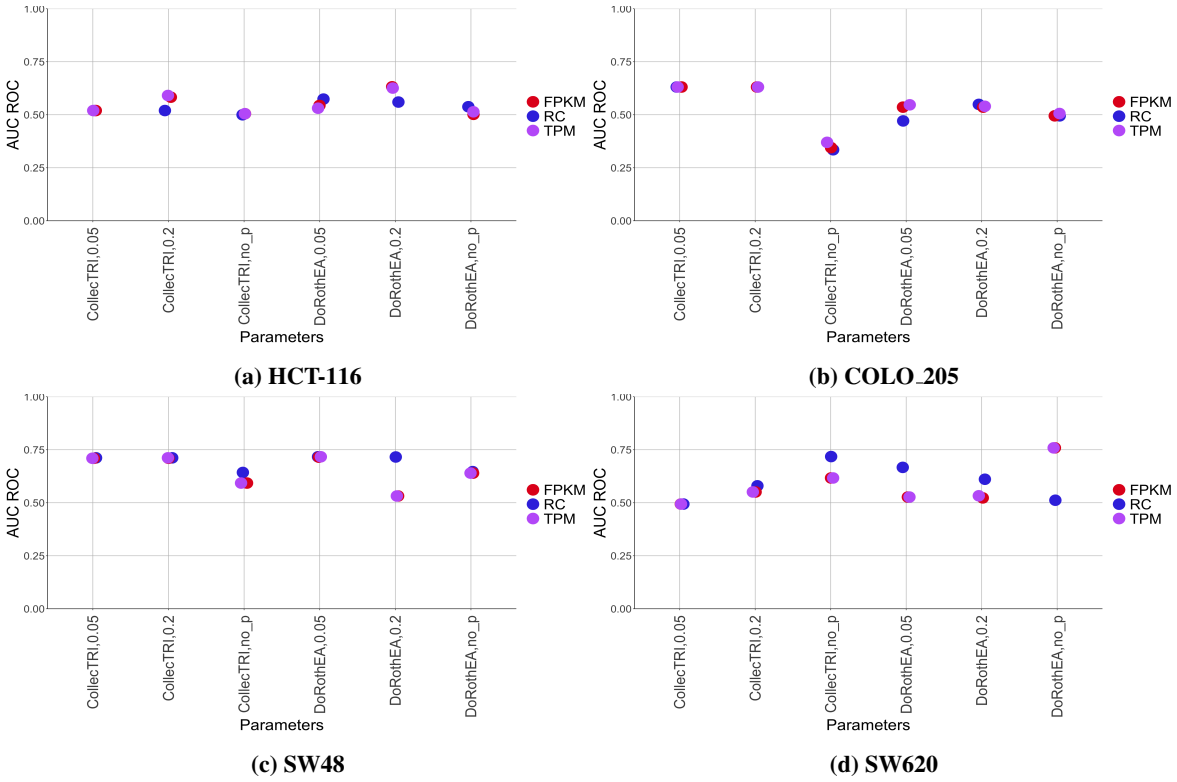


(a) HCT-116 **(b) COLO_205**

(c) SW48 **(d) SW620**

Figure E.10: Synergy results of HCT-116, COLO_205, SW48 and SW620 with the Lu model.

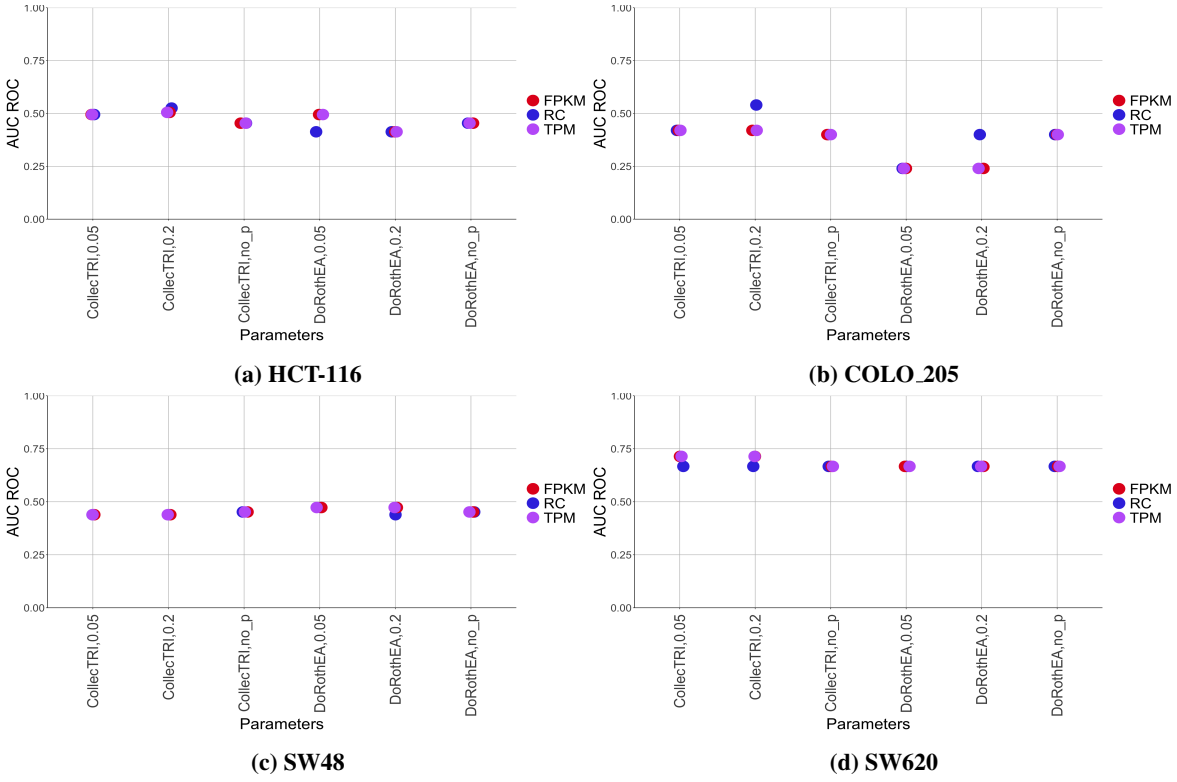
Expression Count Measures



(a) HCT-116 **(b) COLO_205**

(c) SW48 **(d) SW620**

Figure E.11: Synergy results of HCT-116, COLO_205, SW48 and SW620 with the Park model.

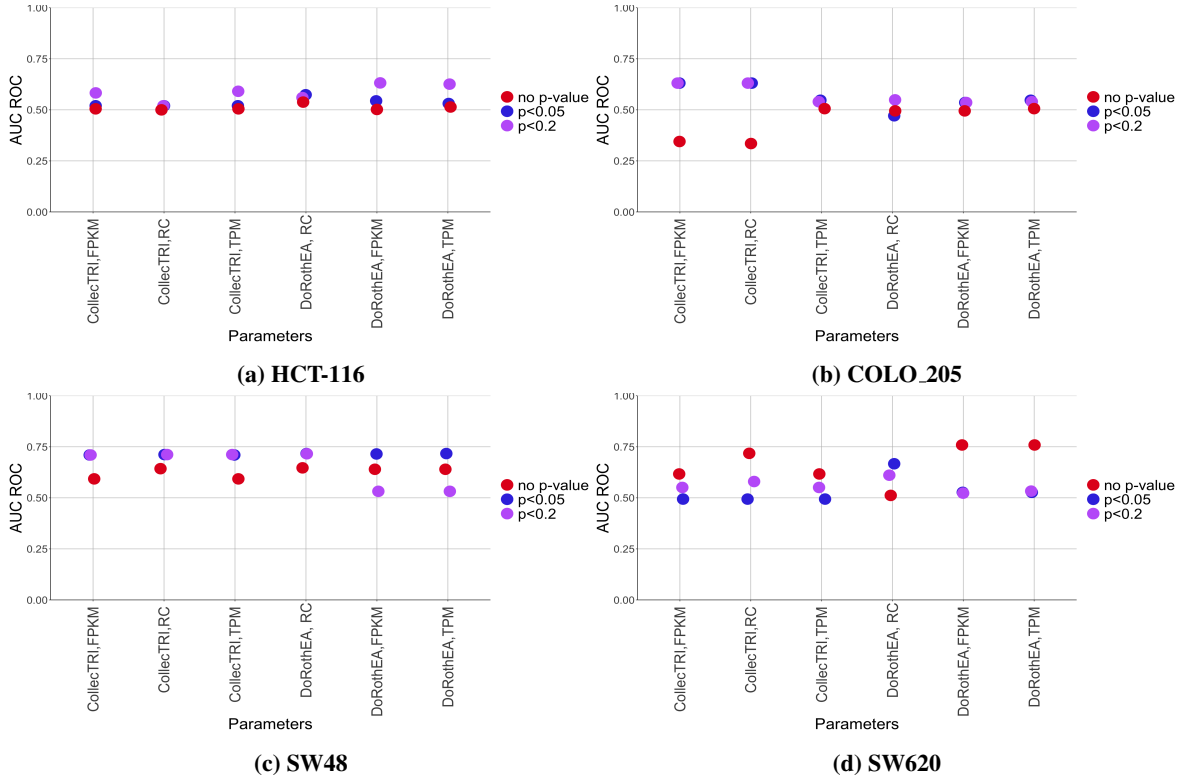


(a) HCT-116 **(b) COLO_205**

(c) SW48 **(d) SW620**

Figure E.12: Synergy results of HCT-116, COLO_205, SW48 and SW620 with the Lu model.

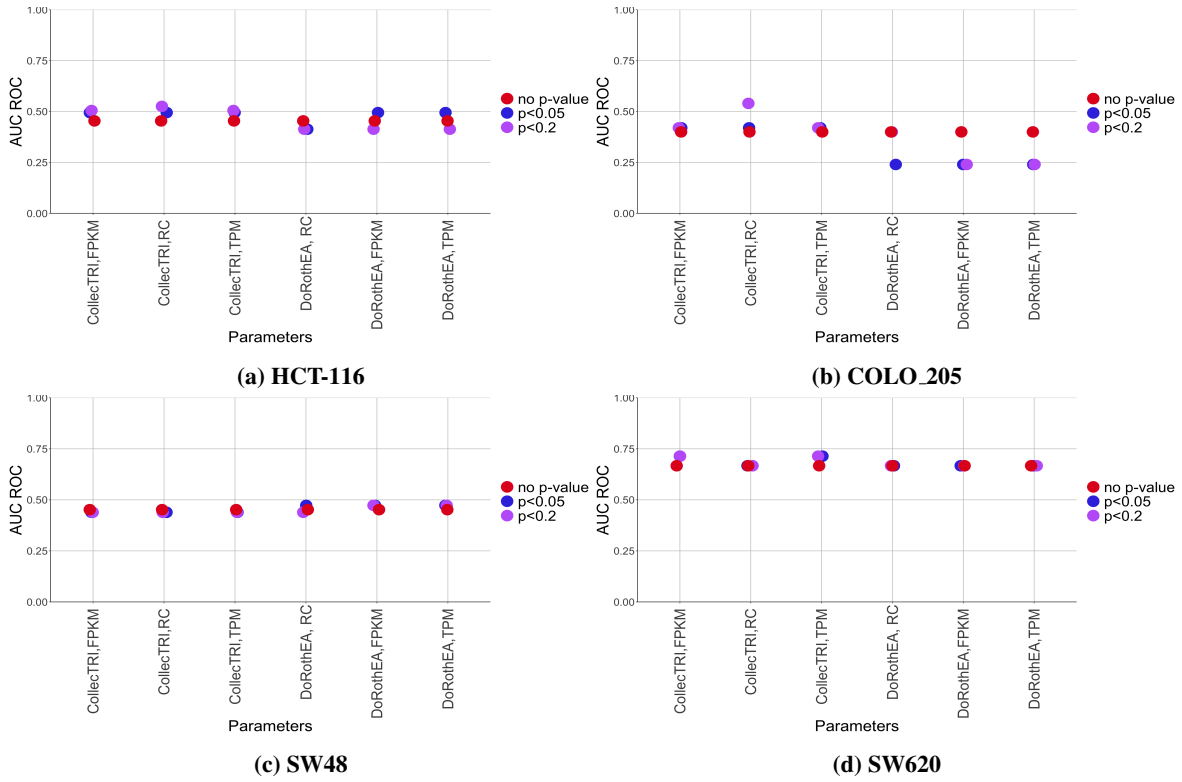
P-values



(a) HCT-116 **(b) COLO_205**

(c) SW48 **(d) SW620**

Figure E.13: Synergy results of HCT-116, COLO_205, SW48 and SW620 with the Park model.



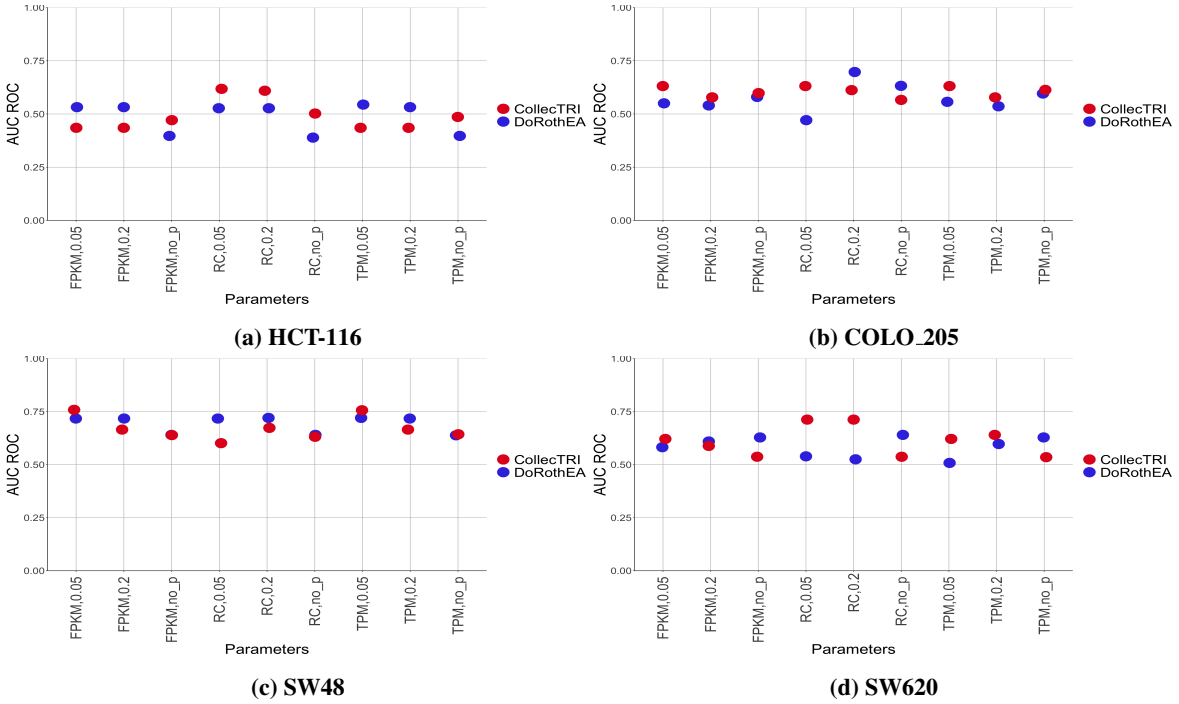
(a) HCT-116 **(b) COLO_205**

(c) SW48 **(d) SW620**

Figure E.14: Synergy results of HCT-116, COLO_205, SW48 and SW620 with the Lu model.

E.2.3 MLM

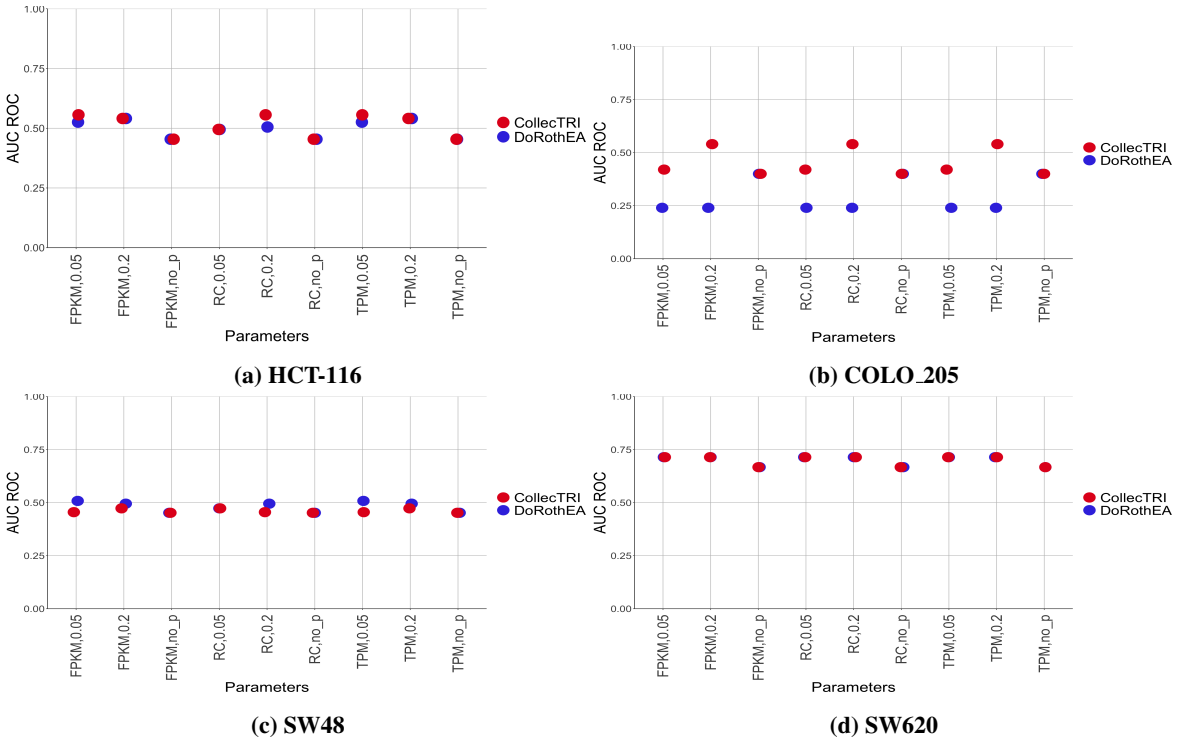
Gene Regulatory Networks



(a) HCT-116 **(b) COLO_205**

(c) SW48 **(d) SW620**

Figure E.15: Synergy results of HCT-116, COLO_205, SW48 and SW620 with the Park model.

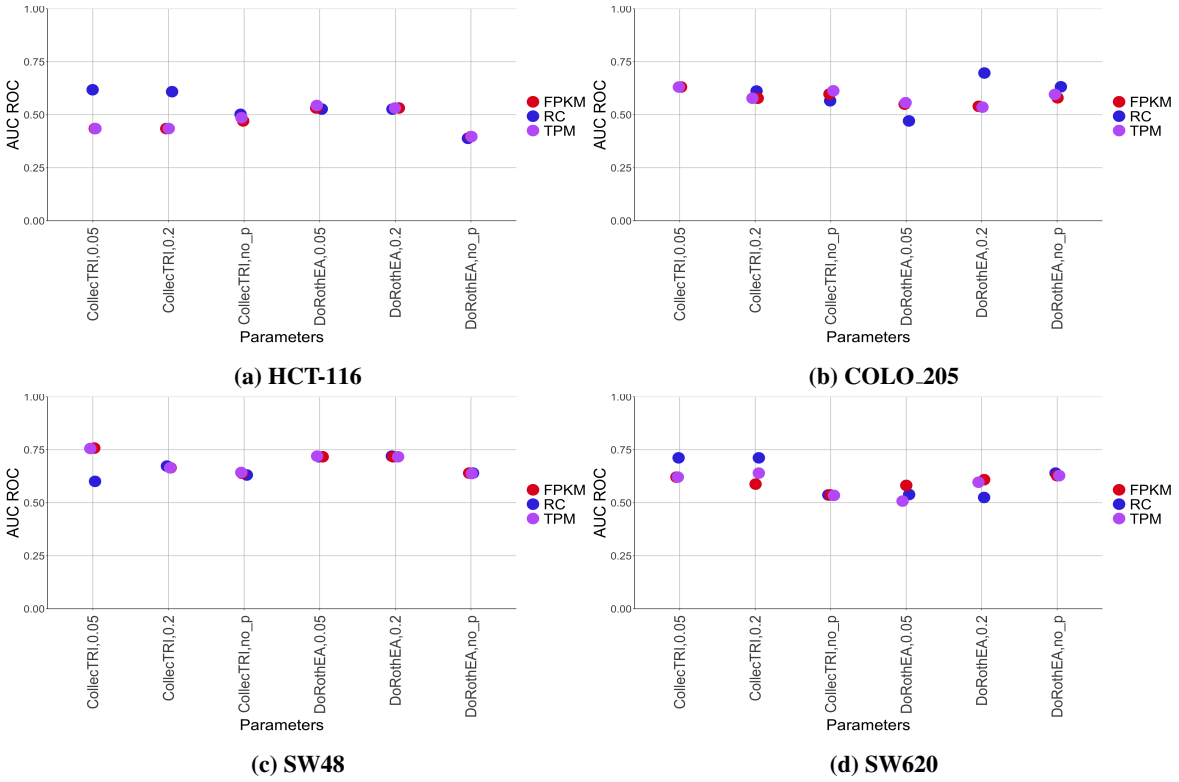


(a) HCT-116 **(b) COLO_205**

(c) SW48 **(d) SW620**

Figure E.16: Synergy results of HCT-116, COLO_205, SW48 and SW620 with the Lu model.

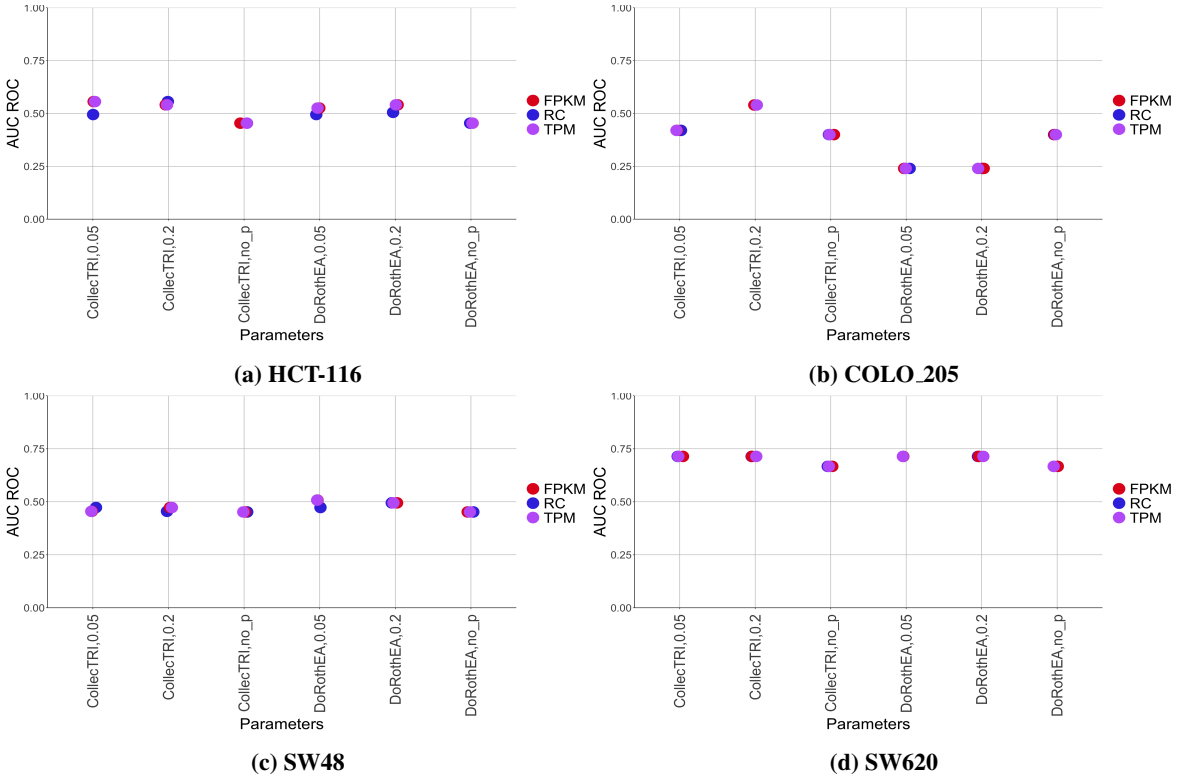
Expression Count Measures



(a) HCT-116 **(b) COLO_205**

(c) SW48 **(d) SW620**

Figure E.17: Synergy results of HCT-116, COLO_205, SW48 and SW620 with the Park model.

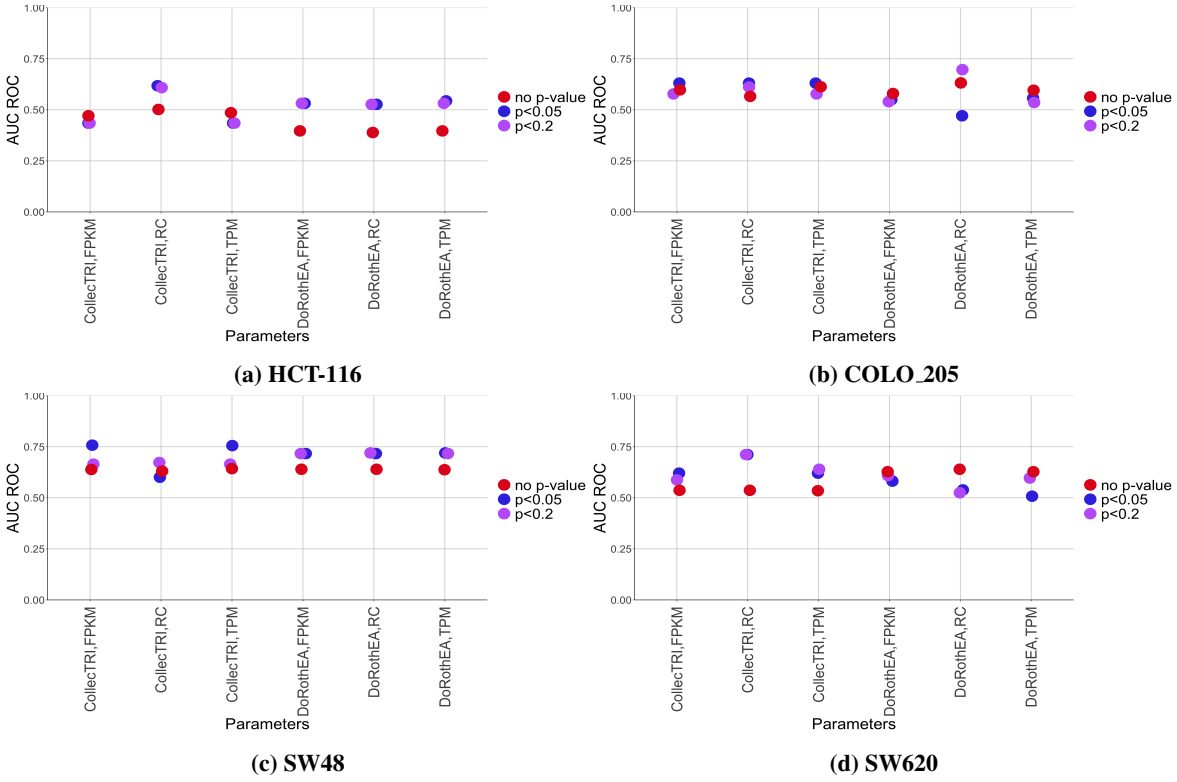


(a) HCT-116 **(b) COLO_205**

(c) SW48 **(d) SW620**

Figure E.18: Synergy results of HCT-116, COLO_205, SW48 and SW620 with the Lu model.

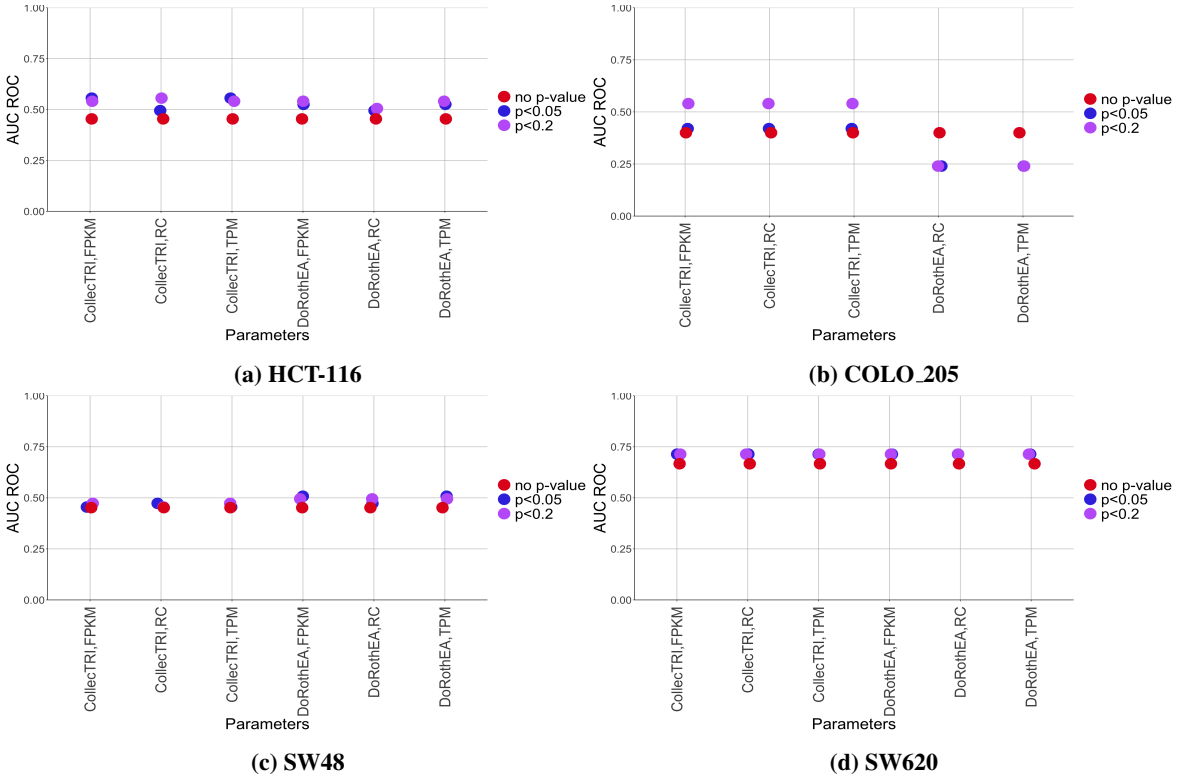
P-values



(a) HCT-116 **(b) COLO_205**

(c) SW48 **(d) SW620**

Figure E.19: Synergy results of HCT-116, COLO_205, SW48 and SW620 with the Park model.



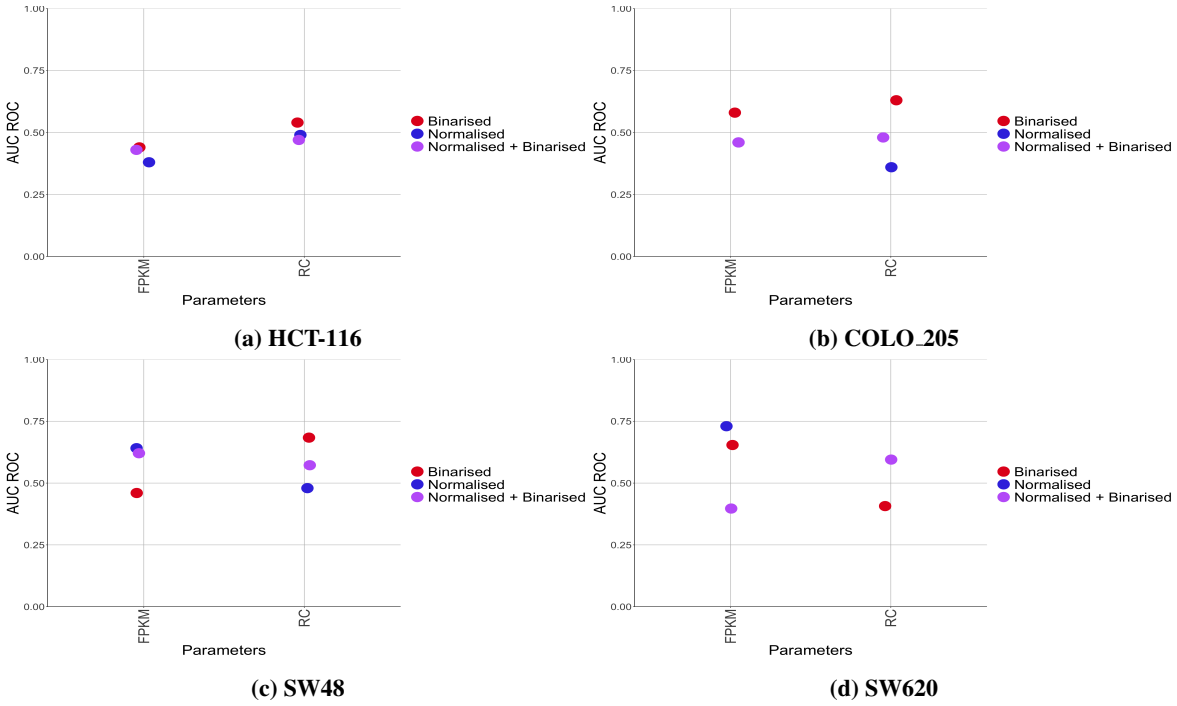
(a) HCT-116 **(b) COLO_205**

(c) SW48 **(d) SW620**

Figure E.20: Synergy results of HCT-116, COLO_205, SW48 and SW620 with the Lu model.

PROFILE

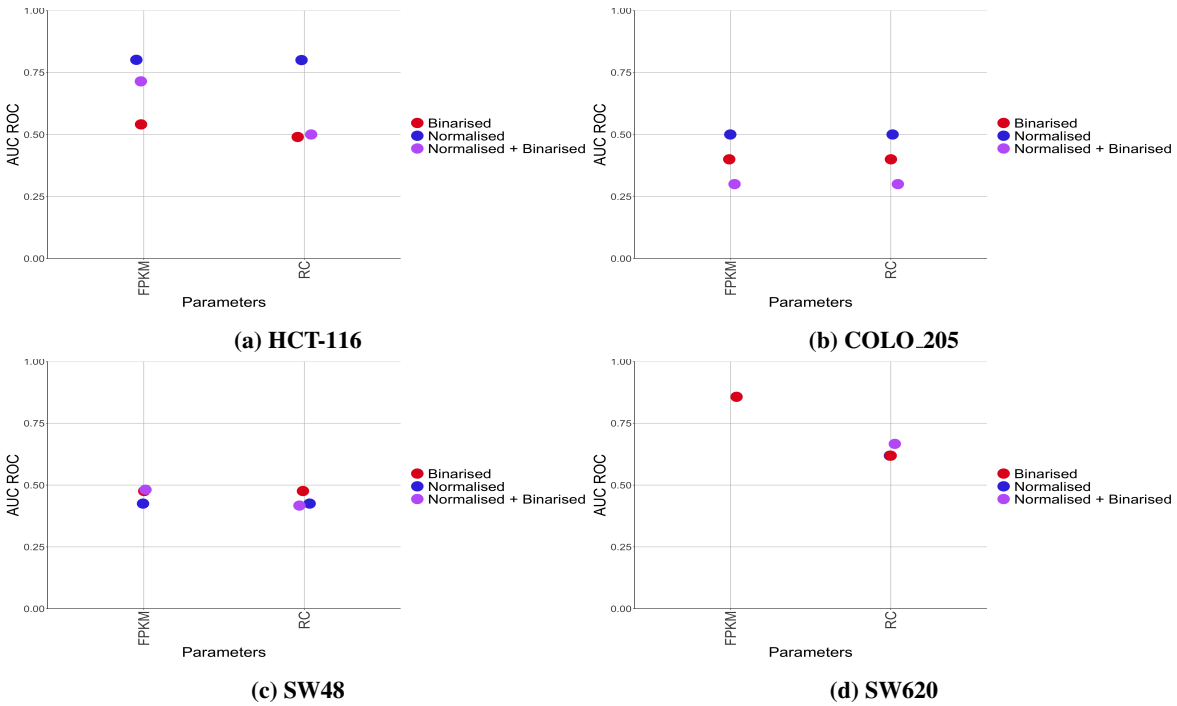
Normalised/Binarised Output



(a) HCT-116 **(b) COLO_205**

(c) SW48 **(d) SW620**

Figure E.21: Synergy results of HCT-116, COLO_205, SW48 and SW620 with the Park model.

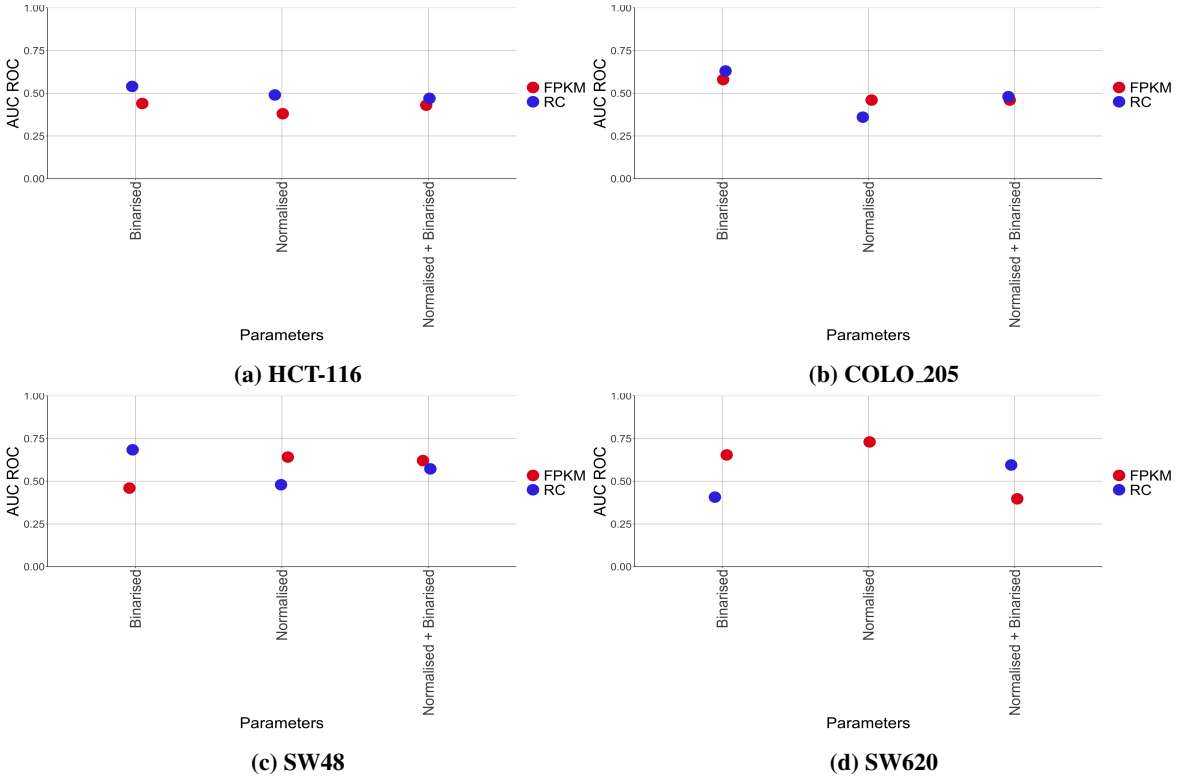


(a) HCT-116 **(b) COLO_205**

(c) SW48 **(d) SW620**

Figure E.22: Synergy results of HCT-116, COLO_205, SW48 and SW620 with the Lu model.

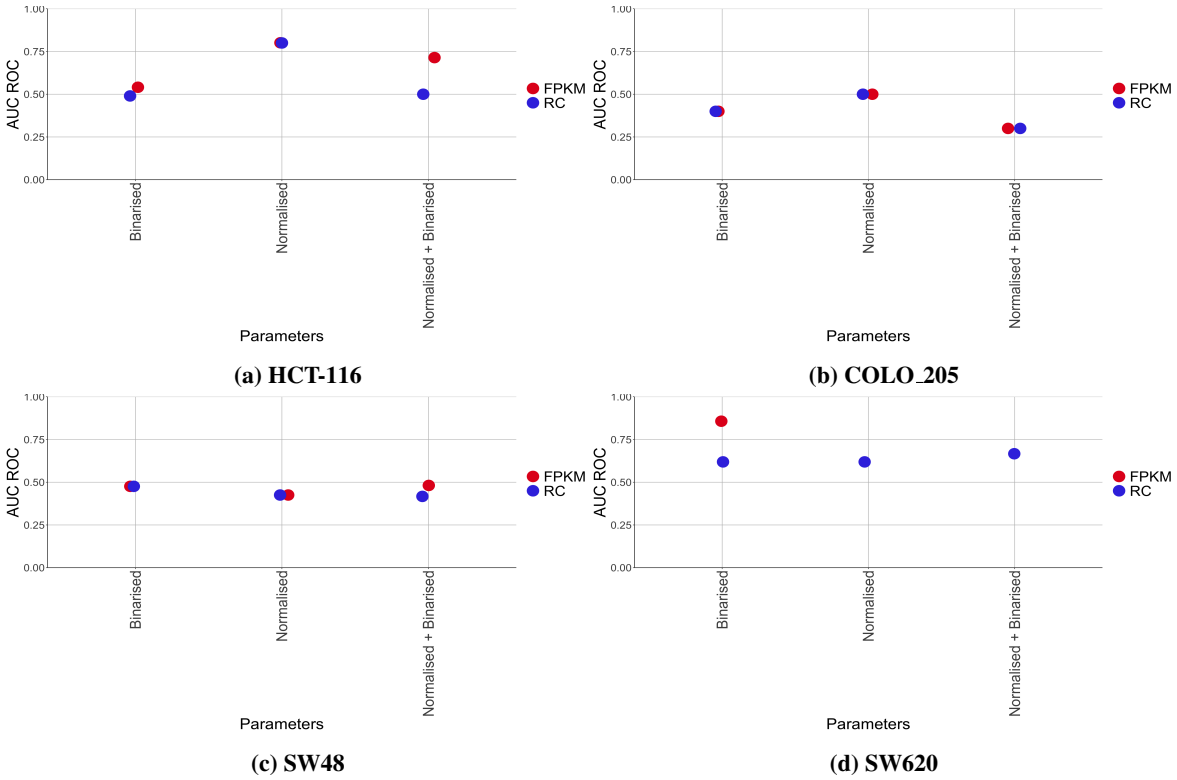
Expression Count Measures



(a) HCT-116 **(b) COLO_205**

(c) SW48 **(d) SW620**

Figure E.23: Synergy results of HCT-116, COLO_205, SW48 and SW620 with the Park model.



(a) HCT-116 **(b) COLO_205**

(c) SW48 **(d) SW620**

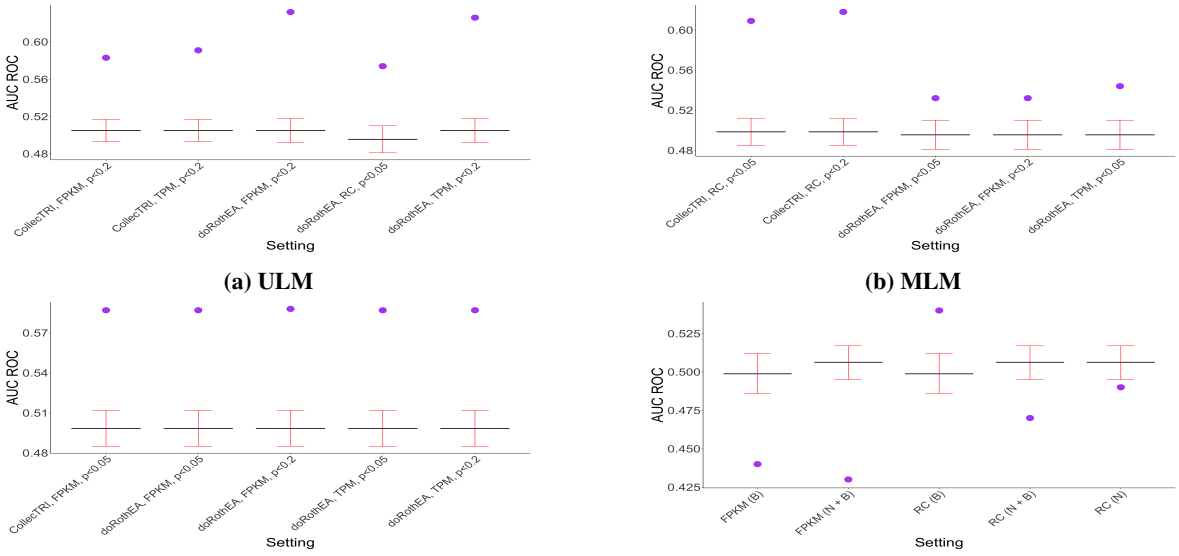
Figure E.24: Synergy results of HCT-116, COLO_205, SW48 and SW620 with the Lu model.

E.3 Statistical Analysis Results

Figure E.25 - E.32 displays the top five synergy results from each software tool compared to bootstrapping of random training data for the Lu and Park models. Purple dots represent tool-specific AUC ROC values, black lines represent the mean of bootstrapped random training data, and red lines indicate the associated confidence intervals.

E.3.1 Park Model

HCT-116

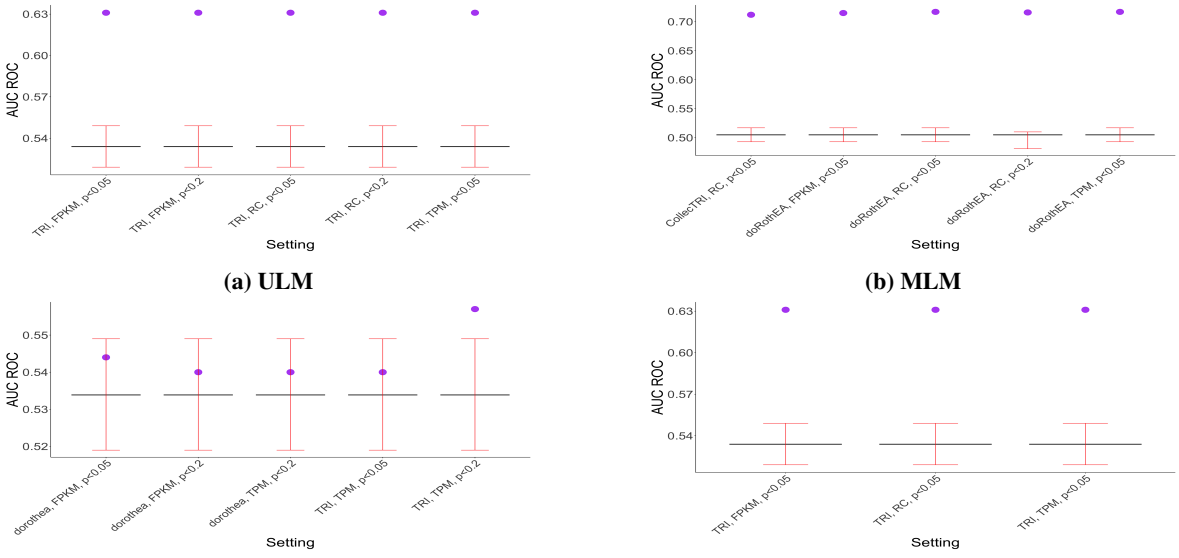


(a) ULM **(b) MLM**

(c) CONSENSUS **(d) PROFILE**

Figure E.25: Bootstrap results with the HCT-116 cell line and the Park model.

COLO_205

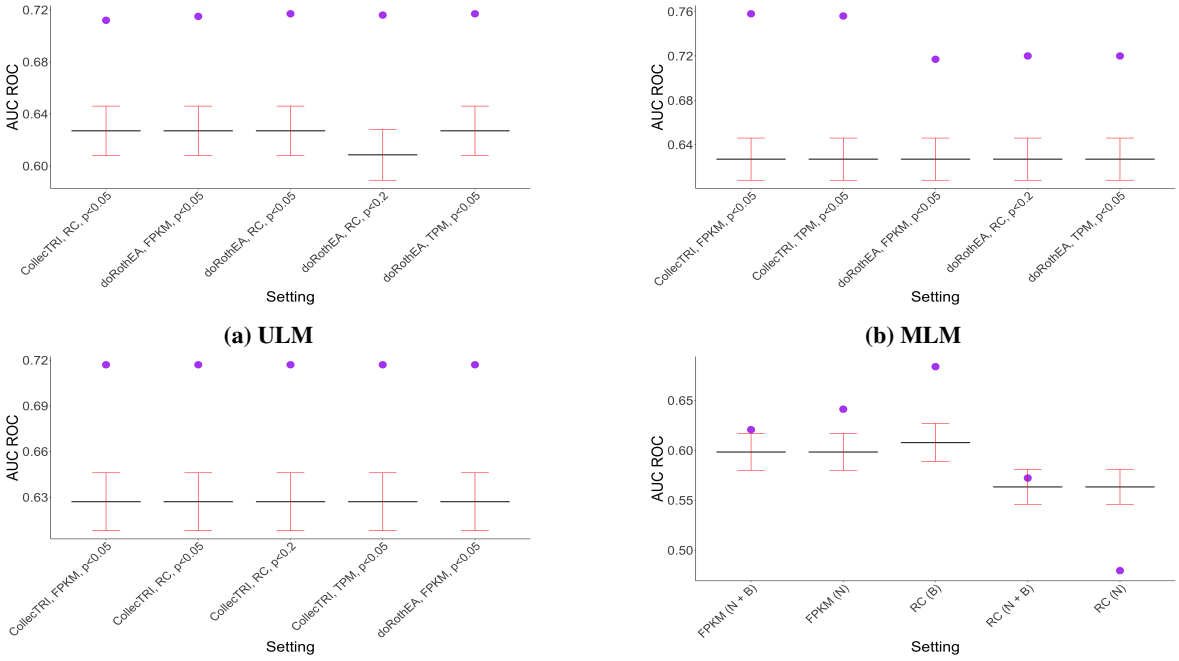


(a) ULM **(b) MLM**

(c) CONSENSUS **(d) PROFILE**

Figure E.26: Bootstrap results with the COLO_205 cell line and the Park model.

SW48

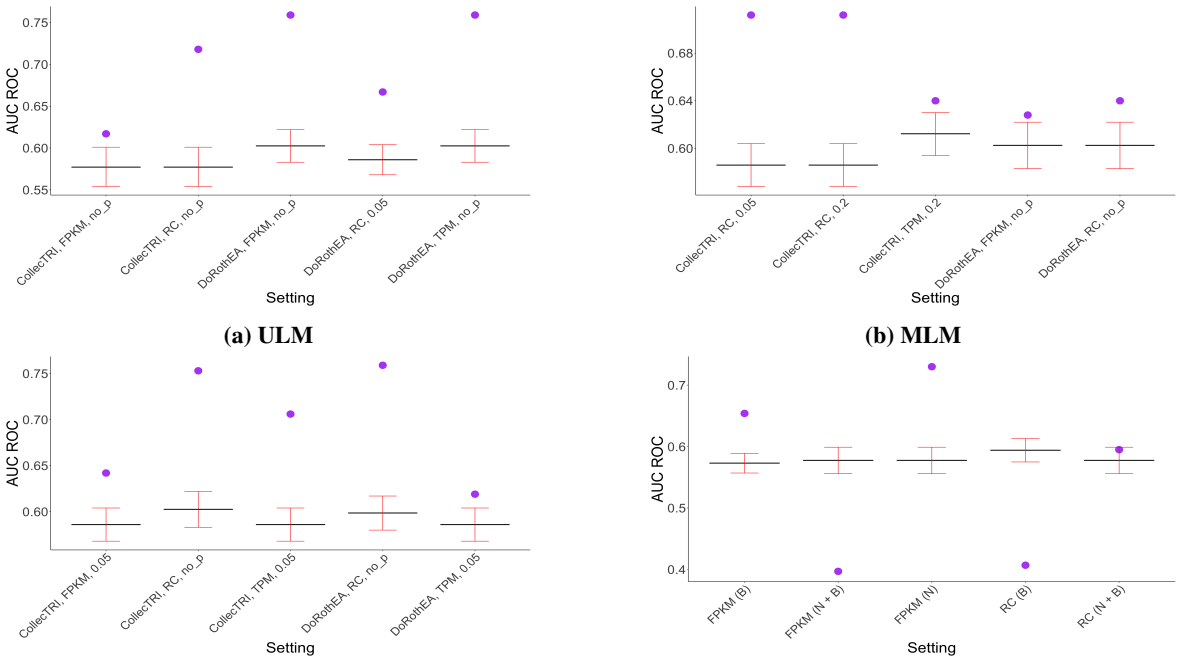


(a) ULM **(b) MLM**

(c) CONSENSUS **(d) PROFILE**

Figure E.27: Bootstrap results with the SW48 cell line and the Park model.

SW620



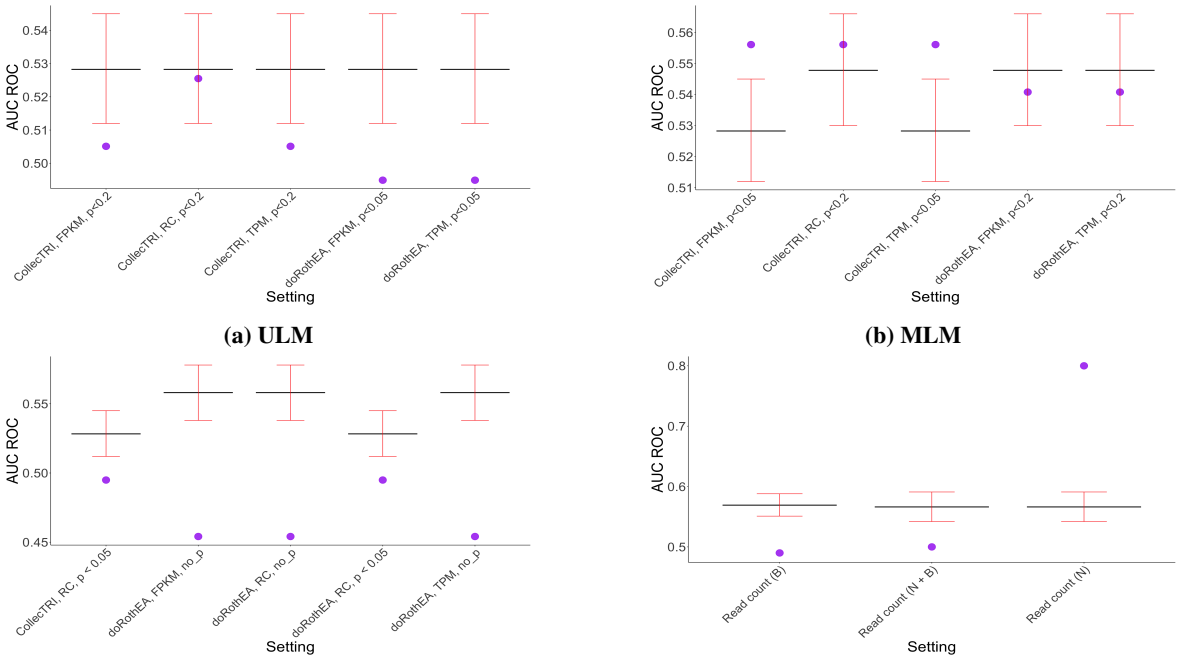
(a) ULM **(b) MLM**

(c) CONSENSUS **(d) PROFILE**

Figure E.28: Bootstrap results with the SW620 cell line and the Park model.

E.3.2 Lu Model

HCT-116

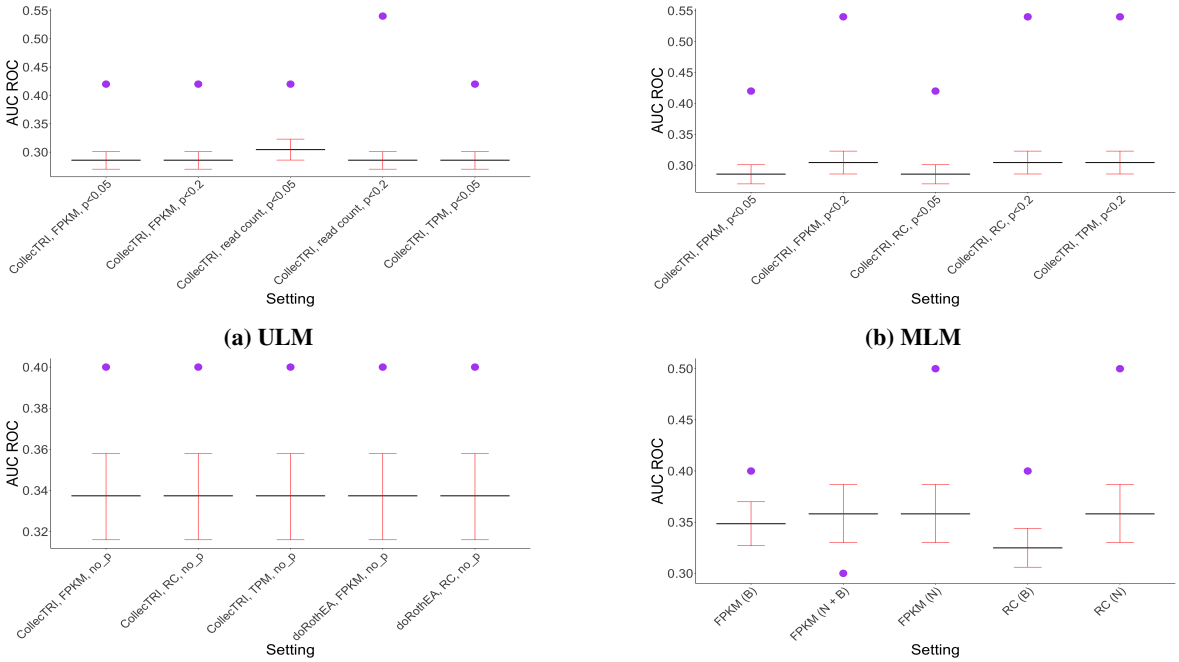


(a) ULM **(b) MLM**

(c) CONSENSUS **(d) PROFILE**

Figure E.29: Bootstrap results with the HCT-116 cell line and the Lu model.

COLO_205

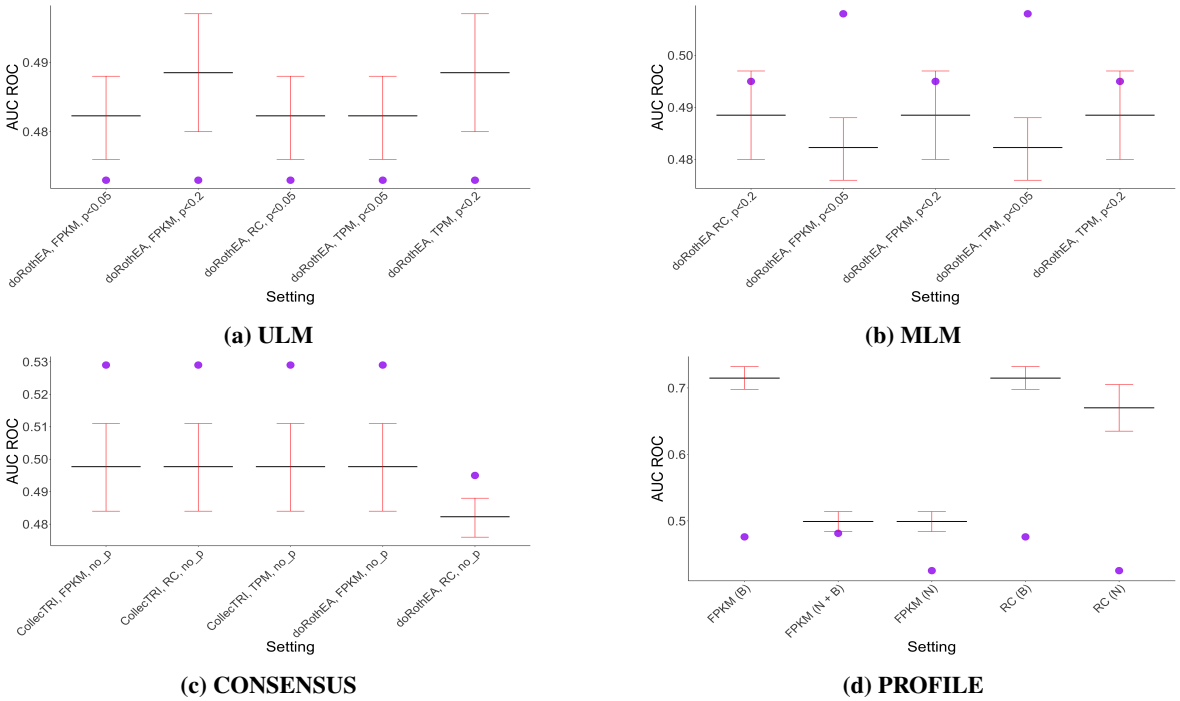


(a) ULM **(b) MLM**

(c) CONSENSUS **(d) PROFILE**

Figure E.30: Bootstrap results with the COLO_205 cell line and the Lu model.

SW48

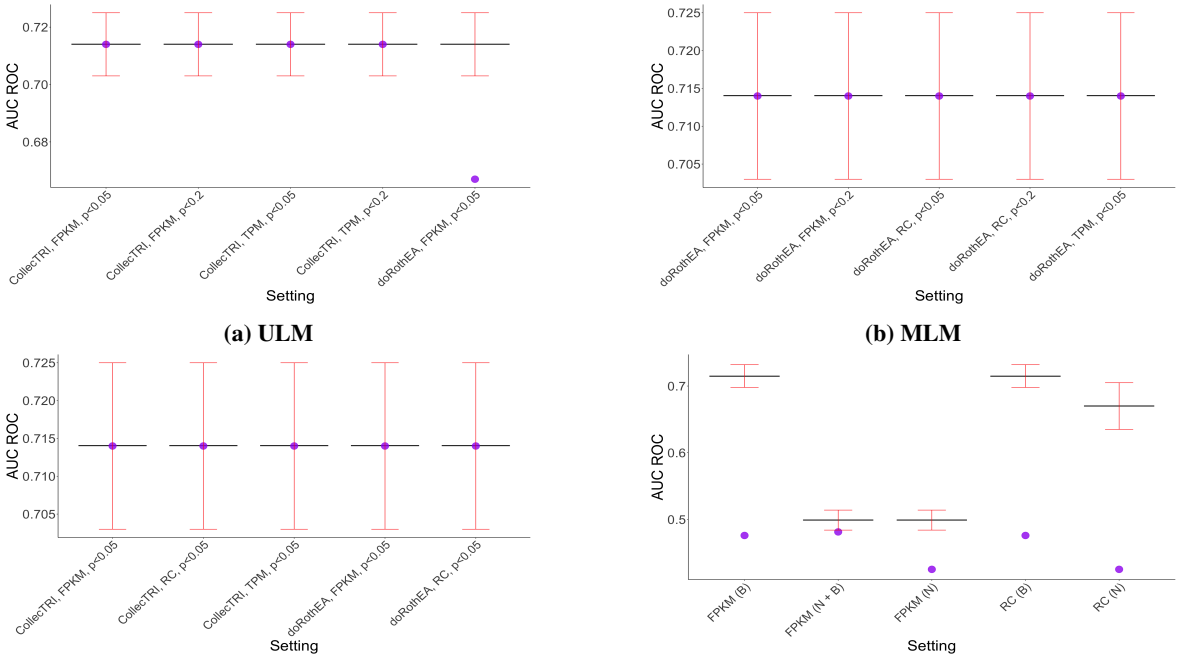


(a) ULM **(b) MLM**

(c) CONSENSUS **(d) PROFILE**

Figure E.31: Bootstrap results with the SW48 cell line and the Lu model.

SW620



(a) ULM **(b) MLM**

(c) CONSENSUS **(d) PROFILE**

Figure E.32: Bootstrap results with the SW620 cell line and the Lu model.

F Curated AGS Biomarker Activities

Table F.1 and Table F.2 displays the reports on conflicting findings for well sustained and less sustained AGS steady state activity observations of Flobak et al. [30]. Only entities in the CASCADE 1.0 model are included in the tables. References for these observations can be found in the supplementary material of Flobak et al. [30].

- Pro: Number of scientific publications in favor of interpretation
- Con: Number of scientifc publications opposing interpretation
- Questionmarks: Indicate probable interpretations

Table F.1: Well substantiated observations of AGS biomarker activities reported by Flobak et al. [30].

Protein	State	Pro	Con
AKT_f	On	21	2
ERK_f	On	16	13
MMP_f	On	10	4
JNK_f	Off	9	4
MAPK14	Off	7	2
BCL2	On	6	0
GSK3_f	Off	6	0
KRAS	On	6	0
NFκB_f	On	6	1
TP53	Off	5	0
S6K_f	On	5	1
BAX	Off	4	0
CASP3	Off	4	0
PIK3CA	On	4	0
PTEN_g	Off	4	0
RAC_f	On	4	0
MYC	On	4	1
CCND1	On	4	1
CTNNB1	ON	3	0
CASP8	Off	3	0
TCF7_f	On	3	0

Table F.2: Less well substantiated observations of AGS biomarker activities reported by Flobak et al. [30].

Protein	State	Pro	Con
LEF	On	2	0
SOS1	On	2	0
GRB2	On	2	1
IKBKB	Off	2	1
MAP2K7	On	2	1
RAF_f	On	2	1
MAP3K5	Off	1	0
BAD	Off	1	0
CASP9	Off	1	0

Table F.2 – *Continued from previous page*

Protein	State	Pro	Con
cFLIP	On	1	0
EGFR	On/off	2	3
EGR1	Off	1	0
FOXO_f	On?	1	0
FZD_f	On	1	0
LRP_f	On	1	0
MAP2K4	Off	1	0
MAP3K11	Off?	1	0
PDPK1	On	1	0
SFRP1	Off	1	0
SHC1	Off	1	0
PTPN11	On	1	0
Tab1/2	Off	1	0
AXIN1	On/off	1	1
CHUCK	On	1	1
MDM2	On	1	1
MEK_f	On/off	1	1
mTOR	On/off	1	1
MAP3K7	Off	1	1
MAP3K8	Off	0	1

No experimental observations were reported for:

- betaTrCP
- CK1_f
- Cytochrome C
- DKKf_f
- DUSP1
- DUSP6
- DVL_f
- GAB_f
- GRAP2
- IRS1
- ITCH
- MAP3K4
- MAP2K3

- MSK_f
- mTORC1_c
- mTORC2_c
- NLK
- AKT1S1
- RHEB
- RSK_f
- TSC2/1

G AGS CASCADE 1.0 Synergy Results

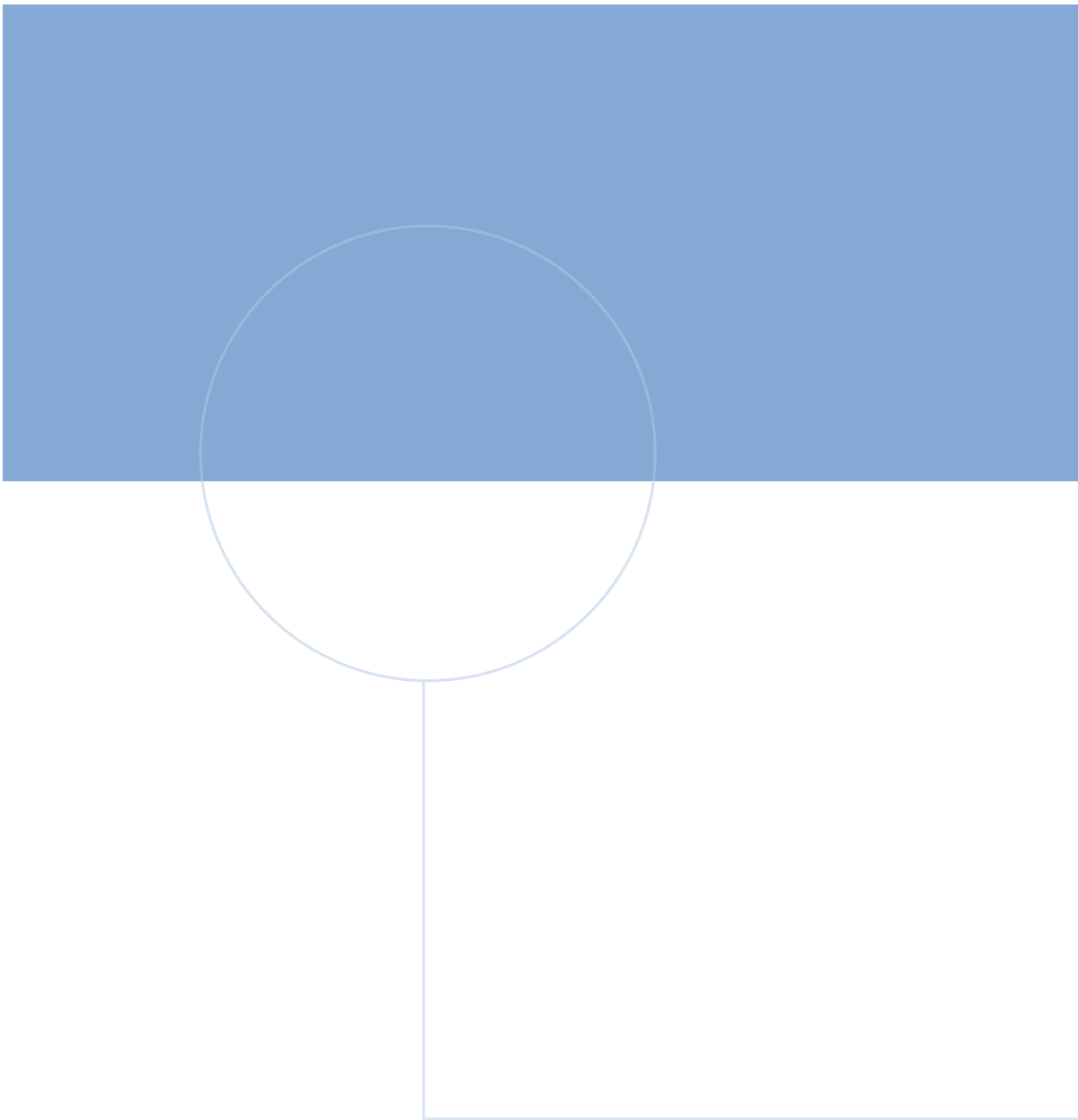
Table G.1 displays the AUC ROC values for AGS CASCADE 1.0, together with the inferred biomarker activities and the respective inference tools and parameter settings.

Table G.1: AUC ROC values of synergy results from the DrugLogics pipeline using the CASCADE 1.0 model, along with the inferred biomarker activities from the AGS cell line, inference tools, and parameter settings. The AUC ROC values are ranked from highest to lowest.

Tools (parameters)	Biomarker activities	AUC ROC
Consensus (CollecTRI, FPKM, p<0.2)	MYC:1, TP53:0	0,824
Consensus (CollecTRI,TPM,p<0.2)	MYC:1, TP53:0	0,824
Consensus (doRothEA, FPKM, p<0.2)	MYC:1, TP53:0	0,824
Consensus (doRothEA, TPM, p<0.2)	MYC:1, TP53:0	0,824
ULM (CollecTRI, FPKM, no p-value)	MYC:1, TP53:0, EGR1:0, NFKB_f:1, FOXO_f:0, TCF7_f:1, CTNNB1:1	0,779
ULM (CollecTRI, TPM, no p-value)	MYC:1, TP53:0, EGR1:0, NFKB_f:1, FOXO_f:0, TCF7_f:1, CTNNB1:1	0,779
Consensus (CollecTRI, RC, p<0.2)	MYC:1, TCF7_f:1, TP53:0	0,765
Consensus (CollecTRI, RC, no p-value)	EGR1:0, FOXO_f:0, MYC:1, TP53:0	0,765
Consensus (doRothEA, RC, p<0.2)	MYC:1, TCF7_f:1, TP53:0	0,765
ULM (doRothEA, read,count, p<0.05)	MYC:1, TCF7_f:1, TP53:0	0,765
ULM (doRothEA, RC, p<0.2)	FOXO_f:0, MYC:1, NFKB_f:1, TCF7_f:1, TP53:0	0,765
ULM (doRothEA, RC, no p-value)	EGR1:0, FOXO_f:0, MYC:1, TP53:0	0,765
ULM (doRothEA, TPM, no p-value)	EGR1:0, MYC:1, TP53:0	0,765
ULM (doRothEA, FPKM, p<0.2)	MYC:1 TCF7_f:1 TP53:0	0,765
ULM (doRothEA, TPM, p<0.2)	MYC:1, TCF7_f:1, TP53:0	0,765
PROFILE (FPKM, normalized + binarized)	AKT_f:1, MAP3K5:0, AXIN1:0, BAD:0, BAX:1, CTNNB1:1, BTRC:0, CASP3:1, CASP8:1, CASP9:0, CCND1:1, CFLAR:0, CK1_f:0, MYC:0, CYCS:0, DKK_f:1, JNK_f, DVL_f:0, EGR1:0, ERK_f:1, FOXO_f:0, FZD_f:1, GSK3_f:1, IKBKB:0, IRS1:1, ITCH:1, JNK_f:1, LRP_f:1, MAP3K8:1, MDM2:1, MEK_f:0, MAP3K4:0, MAP2K3:1, MAP2K4:1, MAP2K7:0, MAP3K11:0, MSK_f:1, mTORC1_c:1, mTORC2_c:0, NFKB_f:0, NLK:0, MAPK14:1, TP53:0, PDPK1:0, PIK3CA:0, AKT1S1:1, PTEN:0, RAC_f:1, RAF_f:1, KRAS:1, RHEB:0, RSK_f:1, RTPK_f:1, S6K_f:1, SHC1:1, PTPN11:1, SOS1:0, TAB_f:1, MAP3K7:1, TCF7_f:1, TSC_f:1	0.76
ULM (CollecTRI, FPKM, p<0.2)	MYC:1, FOXO_f:0, TCF7_f:1	0,721
ULM (CollecTRI, TPM, p<0.2)	MYC:1, FOXO_f:0, TCF7_f:1	0,721
PROFILE (RC, binarized)	FZD_f:0, RTPK_f:1	0,69
Consensus (CollecTRI, FPKM, p<0.05)	MYC:1	0,676
Consensus (CollecTRI, RC, p<0.05)	MYC:1	0,676
Consensus (CollecTRI, TPM, p<0.05)	MYC:1	0,676
Consensus (doRothEA, FPKM, p<0.05)	MYC:1	0,676
Consensus (doRothEA, RC, p<0.05)	MYC:1	0,676
Consensus (doRothEA, TPM, p<0.05)	MYC:1	0,676
MLM (doRothEA, FPKM, p<0.2)	MYC:1	0,676
MLM (doRothEA, FPKM, no p-value)	EGR1:0, FOXO_f:0, MYC:1, NFKB_f:0, TCF7_f:0, TP53:0	0,676
MLM (doRothEA, RC, p<0.05)	MYC:1	0,676
MLM (doRothEA, RC, p<0.2)	MYC:1	0,676
MLM (doRothEA,RC, no p-value)	EGR1:0, FOXO_f:0, MYC:1, NFKB_f:0, TCF7_f:0, TP53:0	0,676
MLM (doRothEA, TPM, p<0.2)	MYC:1	0,676
MLM (doRothEA, TPM, no p-value)	EGR1:0, FOXO_f:0, MYC:1, NFKB_f:0, TCF7_f:0, TP53:0	0,676
MLM (CollecTRI, FPKM, p<0.05)	MYC:1	0,676
MLM (CollecTRI, RC, p<0.05)	MYC:1	0,676
MLM (CollecTRI, TPM, p<0.05)	MYC:1	0,676
ULM (doRothEA, FPKM, p<0.05)	TCF7_f:1	0,676
ULM (doRothEA, TPM, p<0.05)	TCF7_f:1	0,676
ULM (CollecTRI, FPKM, p<0.05)	MYC:1	0,676
ULM (CollecTRI, RC, p<0.05)	MYC:1	0,676

Table G.1 – *Continued from previous page*

Tools (parameters)	**Biomarker activities**	**AUC ROC**
ULM (CollecTRI, TPM, p<0.05)	MYC:1	0,676
Consensus (doRothEA, RC, no p-value)	EGR1:0, FOXO_f:0, MYC:1, NFKB_f:0, TP53:0	0,662
ULM (doRothEA, FPKM, no p-value)	EGR1:0, FOXO_f:0, MYC:1, NFKB_f:0, TP53:0	0,662
MLM (CollecTRI, RC, p<0.2)	MYC:1, TP53:1, FOXO_f:0	0,603
PROFILE (RC, normalized + binarised)	AKT_f:1, MAP3K5:0, AXIN1:0, BAD:0, BAX:1, BCL2:0, CTNNB1:1, BTRC:0, CASP3:1, CASP8:0, CASP9:1, CCND1:1, CFLAR:0, CK1_f:0, MYC:0, CYCS:0, DKK_f:1, JNK_f, DVL_f:0, EGR1:0, ERK_f:1, FOXO_f:0, FZD_f:0, GAB_f:1, GRB2:0, GSK3_f:1, IKBKB:0, IRS1:1, ITCH:1, JNK_f:1, LEF1:0, LRP_f:1, MAP3K8:1, MDM2:1, MEK_f:0, MAP3K4:0, MAP2K3:1, MAP2K4:1, MAP2K7:0, MAP3K11:1, MMP_f:0, MSK_f:0, mTORC1_c:0, mTORC2_c:0, NFKB_f:0, NLK:0, MAPK14:1, TP53:0, PDPK1:1, PIK3CA:0, AKT1S1:1, PTEN:1, RAC_f:0, RAF_f:1, KRAS:1, RHEB:0, RSK_f:1, RTPK_f:1, RTPK_g:1, S6K_f:1, SFRP1:0, SHC1:1, PTPN11:1, SOS1:0, TAB_f:1, MAP3K7:0, TCF7_f:0	0,6
MLM (doRothEA, FPKM, p<0.05)	no inferred biomarkers	0,588
MLM (doRothEA, TPM, p<0.05)	no inferred biomarkers	0,588
PROFILE, RC (normalized)	AKT_f:0.54, MAP3K5:0.41, AXIN1:0.40, BAD:0.26, BAX:0.80, BCL2:0.14, CTNNB1:0.84, BTRC:0.48, CASP3:0.74, CASP8:0.48, CASP9:0.60, CCND1:0.72, CFLAR:0.46, CK1_f:0.49, MYC:0.32, CYCS:0.26, DKK_f:0.68, JNK_f:0.44, DVL_f:0.46, EGR1:0.40, ERK_f:0.54, FOXO_f:0.35, FZD_f:0.42, GAB_f:0.57, GRB2:0.49, GSK3_f:0.60, IKBKB:0.31, IRS1:0.67, ITCH:0.55, JNK_f:0.57, LEF1:0.31, LRP_f:0.57, MAP3K8:0.62, MDM2:0.74, MEK_f:0.43, MAP3K4:0.43, MAP2K3:0.67, MAP2K4:0.89, MAP2K7:0.33, MAP3K11:0.50, MMP_f:0.50, MSK_f:0.47, mTORC1_c:0.48, mTORC2_c:0.48, NFKB_f:0.42, NLK:0.38, MAPK14:0.74, TP53:0.34, PDPK1:0.54, PIK3CA:0.46, AKT1S1:0.53, PTEN:0.51, RAC_f:0.47, RAF_f:0.56, KRAS:0.60, RHEB:0.42, RSK_f:0.53, RTPK_f:0.51, RTPK_g:0.51, S6K_f:0.52, SFRP1:0.26, SFRP1:0.26, SHC1:0.61, PTPN11:0.56, SOS1:0.39, TAB_f:0.51, MAP3K7:0.46, TCF7_f:0.47, TSC_f:0.53	0,57
PROFILE, FPKM (normalized)	AKT_f:0.61, MAP3K5:0.32, AXIN1:0.25, BAD:0.20, BAX:0.94, CTNNB1:0.71, BTRC:0.50, CASP3:0.88, CASP8:0.62, CASP9:0.50, CCND1:0.71, CFLAR:0.50, CK1_f:0.48, MYC:0.09, CYCS:0.04, DKK_f:1.00, DUSP1:0.40, DUSP6:0.41, DVL_f:0.45, EGR1:0.33, ERK_f:0.70, FOXO_f:0.37, FZD_f:0.55, GAB1:0.68, GAB2:0.33, GSK3_f:0.67, CHUK:0.69, IKBKB:0.32, IRS1:0.71, ITCH:0.69, JNK_f:0.70, LRP_f:0.65, MAP3K8:0.75, MDM2:0.94, MEK_f:0.42, MAP3K4:0.39, MAP2K3:0.61, MAP2K4:0.91, MAP2K7:0.20, MAP3K11:0.40, MSK_f:0.68, mTORC1_c:0.52, mTORC2_c:0.46, NFKB_f:0.40, NLK:0.36, MAPK14:0.84, TP53:0.21, PDPK1:0.50, PIK3CA:0.32, AKT1S1:0.56, PTEN:0.44, RAC_f:0.55, RAF_f:0.66, KRAS:0.87, RHEB:0.19, RSK_f:0.55, RTPK_f:0.55, S6K_f:0.53, SHC1:0.70, PTPN11:0.79, SOS1:0.42, TAB_f:0.59, MAP3K7:0.60, TCF7_f:0.67, TSC_f:0.63	0,53
MLM (CollecTRI, RC, no p-value)	MYC:1, TP53:1, EGR1:0, TCF7_f:1, CTNNB1:1	0,485
PROFILE (FPKM, binarized)	BAD:0, BAX:1, CASP3:1, MYC:0, MAP2K4:1, REL_f:0, RTPK_f:1	0,48
MLM (CollecTRI, FPKM, p<0.2)	MYC:1, TP53:1	0,471
MLM (CollecTRI, TPM, p<0.2)	MYC:1, TP53:1	0,471
ULM (CollecTRI, RC, p<0.2)	MYC:1, FOXO_f:0, TCF7_f:0	0,412
MLM (CollecTRI, FPKM, no p-value)	MYC:1, TP53:1, EGR1:0, TCF7_f:1, CTNNB1:0	0,390
MLM (CollecTRI, TPM, no p-value)	MYC:1, TP53:1, EGR1:0, TCF7_f:1, CTNNB1:0	0,390
Consensus (CollecTRI, FPKM, no p-value)	EGR1:1, FOXO_f:0, MYC:1, NFKB_f:0, TP53:0	0,294
Consensus (CollecTRI, TPM, no p-value)	EGR1:1, FOXO_f:0, MYC:1, NFKB_f:0, TP53:0	0,294
Consensus (doRothEA, FPKM, no p-value)	EGR1:1, FOXO_f:0, MYC:1, NFKB_f:0, TP53:0	0,294
Consensus (doRothEA, TPM, no p-value)	EGR1:1, FOXO_f:0, MYC:1, NFKB_f:0, TP53:0	0,294
ULM (CollecTRI, RC, no p-value)	MYC:1, TP53:0, EGR1:0, NFKB_f:0, FOXO_f:0, TCF7_f:0, CTNNB1:0	0,294



NTNU

Norwegian University of
Science and Technology