UNIVERSITY OF MURCIA
Faculty of Biology

Detection and analysis of potential therapeutic targets for Multiple Myeloma lines

MASTER'S THESIS

By
Francisco Javier López Carbonell

Advised by
Francisco de Asís Guil Asensio
José Manuel García Carrasco

Murcia, June 2024

Acknowledgements

This work, as part of my Master's thesis in Bioinformatics, has been supported by the University of Murcia and the Faculty of Biology. I would like to express my gratitude to my supervisors Francisco de Asis Guil Asensio (Professor of Applied Mathematics for Science and Engineering in the Department of Computer Engineering at the University of Murcia, Spain) and José Manuel García Carrasco (Professor of Computer Architecture in the Department of Computer Engineering and also Head of the High Performance Computer Architecture Research Group (GACOP) at the University of Murcia, Spain) for their extraordinary attention and guidance throughout the process of preparing this thesis.

Declaration of Authorship

I, Francisco Javier López Carbonell, declare that this master's thesis, titled "*Detection and analysis of potential therapeutic targets for Multiple Myeloma lines*", and the work presented in it are on my own. I confirm that:

- All of the work contained in this thesis was completed primarily during my enrolment as a master's student at this University.
- In the event that any portion of this thesis was previously submitted for a degree or any other qualification at this University or any other institution, I have made it explicitly known.
- I have always acknowledged the original sources when referring to the published work of others.
- All quotations from the work of others are accompanied by the appropriate citations. Furthermore, the entirety of this dissertation is the result of my own efforts, apart from the quotations.
- In instances where this thesis is based on collaborative work, I have clearly delineated the specific contributions made by others and the contributions I have made myself.

Francisco Javier López Carbonell
Murcia, June 2024

Contents

Contents ii

List of Figures iv

List of Tables iv

Abstract v

1 Introduction and Goals 1
- 1.1 Metabolism from a systems biology perspective . . . 1
- 1.2 Genomic-scale metabolic modeling . . . 1
  - 1.2.1 Clinical approach using GEM specific models . . . 2
  - 1.2.2 Key metabolic tasks in cell survival . . . 2
  - 1.2.3 Metabolic subsystems in GEM models . . . 3
- 1.3 Goals and scope . . . 3

2 Materials and Methods 4
- 2.1 Contraint-Based Modeling and Flux Base Analysis . . . 4
  - 2.1.1 Constraint Based Modeling . . . 4
  - 2.1.2 Flux Base Analysis: modes and cutsets . . . 5
- 2.2 Workflow Overview . . . 5
- 2.3 Constructing models . . . 6
  - 2.3.1 The *Human-GEM* model . . . 6
  - 2.3.2 Context-specific models and metabolic tasks . . . 7
- 2.4 Computing and checking MCS . . . 7
- 2.5 Analyzing MCSs . . . 7
  - 2.5.1 Metabolic Atlas . . . 8
  - 2.5.2 Accessing REST Web Services for biosciences . . . 8
- 2.6 Software . . . 8

3 Results 9
- 3.1 MCSs obtained from each model . . . 9
- 3.2 Toxicities and toxicity levels . . . 11
- 3.3 Potential therapeutic targets for the MM lines studied . . . 11
  - 3.3.1 Therapeutic targets encountered for each line . . . 11
  - 3.3.2 Common targets for several MM lines . . . 12

4 Discussion 14

5 Conclusion 20

Bibliography 22

Appendices **26**

A Abbreviations **26**

B Supplementary material 1 **27**

C Supplementary material 2 **29**

D Supplementary material 3 **30**

E Supplementary material 4 **31**

List of Figures

2.1 **Workflow** 5

3.1 **Healthy tissues toxicities classification.** Image, which represents the number of reactions that can cause the death of different healthy cells. 11
3.2 **MM lines MCS classification.** Image which contains the number of common reactions considered MCS classified by the number of MM lines whose they affect. 12

4.1 **PPP ER route scheme.** Image extracted from *Metabolic Atlas*, showing the reticular segment of the pentose phosphate pathway. 16
4.2 **Retinol metabolism ER routes scheme.** Schematic representation of the information extracted from the model that does not match the version shown in *Metabolic Atlas* as it presents updated reactions. 17

List of Tables

3.1 **Summary of total and specific MCSs for different models** 10
3.2 **MCS from each MM line with low and medium toxicity** 12
3.3 **MCSs summary for several lines with low toxicity** 13
3.4 **Association of pairs of reactions that kill all lines** 13

4.1 **MCSs tasks and toxicities summary** 15
4.2 **Interesting MCS genes associated** 18
4.3 **chEMBL information coincidences with diseases** 19

B.1 **Metabolic tasks and their order.** 28

C.1 **Healthy tissue models** 29
C.2 **MM cell lines** 29

D.1 **Promising MCSs with medium toxicity** 30

Abstract

Genome-scale metabolic modeling is a rapidly expanding field within the discipline of systems biology. This growth has been facilitated, among other factors, by the capacity of these models to perform in silico simulations and predictions on models associated with specific contexts. This has made them a highly valuable tool in precision medicine.

The main aim of this study is to find potential therapeutic targets for eliminating multiple myeloma cancer cells while minimizing potential toxic effects on healthy tissues. Subsequently, the context of these reactions in the models will be characterised through metabolic subsystems. These reactions are essential for cell survival and are required to perform at least an essential metabolic tasks.

After their identification, these potential targets have been analyzed. This analysis provide a more in-depth comprehension of the specific characteristics of both groups of models, which will enable the identification of particular traits associated with the respective pathologies and potential treatment strategies.

Los modelos metabólicos a escala genómica están en pleno crecimiento en el área de la biología de sistemas. Este auge ha sido promovido, entre otras funciones, por su capacidad para realizar simulaciones y predicciones in silico sobre modelos asociados a contextos específicos, siendo un herramienta muy útil en medicina de precisión.

En el presente estudio, mediante el análisis de un extenso número de modelos asociados a tejidos sanos concretos y líneas cancerosas, se ha tratado de buscar reacciones esenciales para llevar a cabo 57 tareas metabólicas indispensables para la supervivencia celular.

La finalidad principal de este trabajo es, primeramente tratar de encontrar dianas terapéuticas que consigan eliminar líneas cancerosas, tanto de formal individual como un amplio espectro de las mismas y posteriormente tratar de caracterizar el contexto de esas reacciones en los modelos a través de subsistemas metabólicos.

Este análisis proporcionará una comprensión más profunda sobre las particularidades de ambos grupos de modelos, facilitando la detección de rasgos particulares asociados con las respectivas afecciones y sus posibles tratamientos.

Chapter 1

Introduction and Goals

1.1 Metabolism from a systems biology perspective

Metabolism can be defined as a series of interrelated biochemical processes that ensure the proper functioning, maintenance, and regulation of cells. Metabolic pathways serve as a bridge between the genetic content of cells and external environmental factors. To gain a comprehensive understanding of cellular metabolism, it is essential to conduct a detailed analysis of the biological system, integrating the functions of the entire metabolic network and its interactions [38].

The field of systems biology, a discipline that adopts this vision, has experienced a significant expansion over the past two decades, largely due to the advent of technologies that facilitate the generation of vast quantities of omics data. These data sets are capable of capturing the intricate functioning of living organisms (genomes, transcriptomes, proteomes, metabolomes...), and the development of computational tools has enabled the processing and integration of these data to construct interconnected global biological systems [11].

1.2 Genomic-scale metabolic modeling

The arrival of this new wave of information has permitted the creation of genome-scale metabolic networks for a multitude of organisms, including humans. These networks can be converted into a mathematical format for predictive *in silico* simulations, which are commonly referred to as genome-scale metabolic models (GEMs). The models encompass all the known biochemical reactions that occur in cells, along with their stoichiometric relationships, through the integration of omics data and logical rules that relate genes to proteins (GPR associations).

GEMs have been demonstrated to be effective in several fields. In metabolic engineering, they permit the design of strains that specialize in the production of specific metabolites. In the clinical setting, these models are valuable for identifying therapeutic targets, simulating multicellular interactions, and studying the molecular basis of various diseases.

These computational models are accessible through various software packages, which facilitate the conduct of simulations and predictions across a range of contexts.

A common method to perform these simulations is through constraint-based modeling (CBM) techniques, which permit rapid calculations in complex algebraic models. In these models, a steady state is assumed, whereby the concentration of internal metabolites is considered constant over time. Due to the majority of kinetic parameters associated with metabolic reactions remaining unknown or challenging to calculate experimentally, this approach is essential to obtain quantitative predictions [14].

The programming languages MATLAB and Python offer a suite of tools for the computational analysis of metabolic pathways, with a particular focus on Constraint-Based Reconstruction and Analysis (COBRA). The COBRA toolbox, originally developed in MATLAB, has been migrated to Python with COBRApy in recent years due to several advantages.

One of the advantages of Python is that it is an open-source language. This means that it is accessible to the entire community without the need for a license, unlike MATLAB, which is commercial software. Additionally, COBRApy employs object-oriented programming, which allows for the representation of more complex biological models. Nevertheless, it is essential to recognize that although most of the COBRA Toolbox functions have been implemented in COBRApy, not all are available [26].

The application of linear programming (LP) enables the exploration of flux distributions. This can be achieved by defining an objective function to be optimized [16]. This approach, known as flux balance analysis (FBA), typically employs biomass production as the objective function, as it is representative of cell growth. However, more specific biological functions can also be optimized. This objective function is usually a pseudo-reaction expressed by linear combinations of reactions involved in the biological function to be optimized. Nonetheless, previous research has shown the occurrence of bias [7], so the objective function must be chosen carefully depending on the study being carried out [4].

FBA has been successfully used in the past to assess the phenotypes of wild-type organisms despite its recognized limitations, such as the exclusion of gene regulation, signaling processes, and metabolic regulation [5].

1.2.1 Clinical approach using GEM specific models

Until now, we have been describing models at the genomic level in general terms. However, cell types with different but interconnected metabolisms are found in complex organisms. Systems biology aims to simulate and analyze any metabolic phenotype, which requires the integration of diverse cellular metabolic information. Advances in omics technologies and algorithms allow large volumes of data to be obtained and assembled into genomic reference models (GEMs), enabling the creation of specific models for different biological contexts [8].

This approach is particularly useful in the study of various diseases resulting from metabolic, genetic, or environmental changes, including cancer, diabetes, and hypertension. Moreover, the complexity of metabolic networks and their regulation, which means that different molecular abnormalities can cause a similar phenotypic change, is one of the major challenges in identifying therapeutic targets for these diseases [1].

For this reason, it has been proposed to move towards personalized medicine. This means developing specific strategies for each patient based on their unique profile of genetic, phenotypic, and environmental factors. In this sense, the generation of specialised models is essential to distinguish and compare models of different types of healthy and diseased cells in order to find biomarkers or therapeutic targets to attack with drugs. [9].

Based on this premise, one commonly used strategy is to explore vulnerable points in various healthy human cell types and different diseased cell lines. This approach will be employed in the research of this thesis to minimize damage to healthy tissue and increase the precision of the treatment for the disease under consideration, in this case, multiple myeloma (MM).

1.2.2 Key metabolic tasks in cell survival

Previously, section 1.2 mentioned biomass production as the most commonly chosen objective function in FBA and how this choice can affect the analysis's results. This becomes even more important when we move into models of multicellular organisms, where it seems difficult for the maximisation of a single objective function to capture the totality of all the combinations of functions used by the human cell.

In addition, cancer cells are known to have metabolic patterns that differ from those of healthy cells and those of malignant cells. In response to this challenge, Agren et al. developed a set of 57 essential metabolic tasks that every human cell must perform in order to be considered viable.

It, therefore, seems reasonable to take this set of metabolic tasks into account in our analyses, as we will be working with a large number of healthy tissue models and cancer lines. The full list, as provided by Agren *et al.* [15], can be found in B.1.

1.2.3 Metabolic subsystems in GEM models

To date, most studies of GEMs have focused on analysis at the genetic level. This is particularly relevant for diseases such as cancer, which often involve genetic mutations, epigenetic changes, and variations in protein expression. However, in genome-scale models, reactions offer more flexibility to investigate the network, therefore it is easier to perform comprehensive analyses of reactions than genes.

Recently, some available models have introduced a classification of reactions into metabolic groups (subsystems), which enriches the information obtained and highlights the interrelationships between these reactions, bringing us closer to understanding how they affect possible metabolic changes and ultimately determine phenotypic changes [15].

1.3 Goals and scope

Main goal

The overarching objective of this research is to identify and analyze the most promising therapeutic targets in multiple myeloma (MM) cell lines.

Specific goals

The specific objectives pursued in the present work are:

- Calculate and check minimal cut sets (MCS) of reactions on a large number of models.
- Analyse MCSs with high MM cancer reach and low toxicities.
- Examine the subsystems to which the analyzed MCSs belong to clarify the results obtained.

Chapter 2

Materials and Methods

2.1 Contraint-Based Modeling and Flux Base Analysis

2.1.1 Constraint Based Modeling

Given the scarcity of information inherent to biological networks, it is advantageous to employ an approach that leverages data-driven physicochemical and biological constraints to identify potential phenotypic states of the network. This is precisely the objective of constraint-based modeling, which employs conservation of mass, thermodynamic directionality, and the steady-state assumption as fundamental constraints [17].

A metabolic model is defined by a set of reactions, R, a set of metabolites, M, and a stoichiometric matrix S of dimensions m x n, where rows and columns represent metabolites and reactions:

$$S_{m\times n} = (s_{i,j})_{m\times n} \in M_{m\times n}(\mathbf{R}) \tag{2.1}$$

The stoichiometric coefficient ($s_{i,j}$) of a metabolite i may result in a variety of states in reaction j. If the coefficient is zero, the metabolite i does not participate in the reaction. Conversely, if the coefficient is negative, the metabolite i is consumed and occupies the left side of the reaction. In contrast, if the coefficient is positive, the metabolite i is produced and occupies the right side of the reaction.

In a given network, each state is described by a vector of variables whose length is equal to the number of reactions involved. This vector, known as the flow rate vector, v, provides information about the performance of each reaction. Each variable represents the rate at which the corresponding reaction occurs in that particular state.

When a sufficiently short time interval is considered, it is common to assume that the concentrations of internal metabolites remain constant. This assumption leads to the steady-state constraint, which refers to an equilibrium in which no metabolites are produced or consumed. This equilibrium can be summarized as follows:

$$S \cdot v = 0 \tag{2.2}$$

A reaction $r_i \in R$ will be irreversible if it carries flow in one possible direction. Irreversible reactions belong to a subset $Irr \subset R$. It will be assumed that all irreversible reactions proceed in the positive direction. Reversible reactions will be those where flux can be directed in both directions. The reversibility or irreversibility of the reactions will be determined by the Gibbs free energy (ΔG), as thermodynamics is another constraint considered in CBM [13]. Equation 2.3 is known as the thermodynamic constraint:

$$r_i \geq 0 \; \forall r_i \in Irr \tag{2.3}$$

2.1.2 Flux Base Analysis: modes and cutsets

Any vector satisfying equations 2.2 and 2.3 is known as a mode (or state) of the network, representing a possible phenotype of the system. All these flux vectors constitute a solution space, denoted C, and known as the flux cone of the metabolic network.

$$C = \{v \in \mathbb{R}^n \mid S \cdot v = 0, v_i \geq 0 \; \forall r_i \in Irr\}$$

Flux balance analysis, FBA, represents the principal CBM method employed in the flux analysis of metabolic models. In this method, an objective function (typically a linear combination of fluxes) is introduced. By optimizing it by using linear programming techniques, the corresponding solution provides different solutions of the respective equations 2.2 and 2.3.

The support of a mode, $supp(v)$, is defined as the set of reactions that appear in the mode with a non-zero rate. It should be noted that the set of modes is infinite for any non-trivial network and contains the trivial mode, that is, the mode in which all fluxes are null.

Given a target set of nodes T, a cut set for T is a set of reactions C such the simultaneous blockage of all reactions in C renders infeasible all modes in T. A cut set for T is called a minimal cut set, MCS, if it does not contain any proper subset that is also a cut set for T.

2.2 Workflow Overview

The purpose of this section is to describe the methods and the stages into which this work has been divided. The workflow that has been followed is the following:

- Generate all the models needed in our study
- Compute and verify the corresponding MCS for each model
- Analyze the obtained MCSs

This workflow is summarized in figure 2.1.

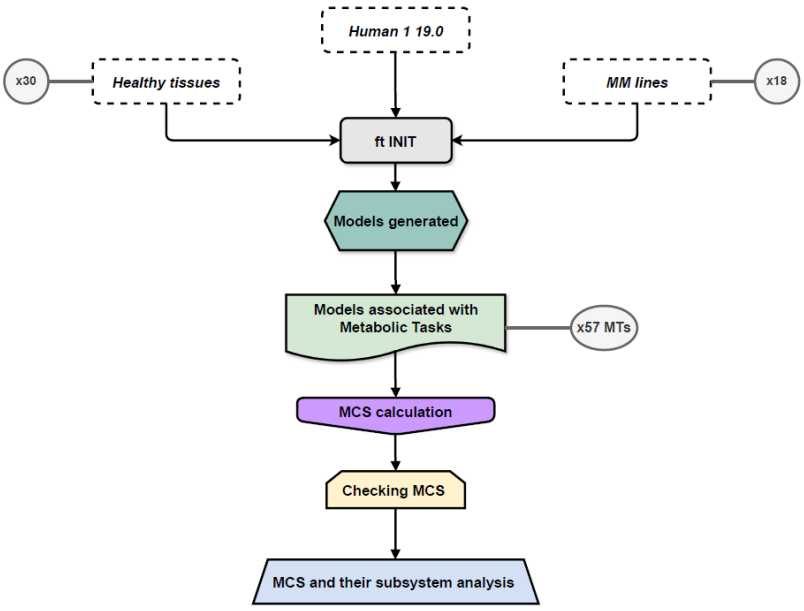

Figure 2.1: **Workflow**

The remaining sections are devoted to explaining the tools used in each step of the proposed workflow.

2.3 Constructing models

The first step for our study is to obtain all the necessary models. We need to differentiate between the generic model (*Human-GEM*) which represents a generic human cell and models representing healthy tissues and multiple myeloma cells.

2.3.1 The *Human-GEM* model

Human-GEM is a generic genome-scale metabolic model of *Homo sapiens*, which was reconstructed in a version-controlled GitHub repository by the Systems and Synthetic Biology department at Chalmers University of Technology (Gothenburg, Sweden).

Human-GEM is the result of the integration of information gathered from previous human metabolic models: *Recon3D* [2], *HMR2* [12] and *iHsa* [10]. *Recon3D* is the most recent version of the Recon series, while *HMR2* is a model from the Human Metabolic Reaction series. These lineages of cellular models were acquiring updates by mutually benefiting from each other. Finally, *iHsa* is a GEM that draws from *HMR2*, *Recon1* and *Recon2* (two previous versions of *Recon3D*) and other specialized human models.

The purpose of Robinson *et al.* for generating *Human1* was to unify all the human metabolism information gathered in the aforementioned models in order to standardize identifiers, limit the propagation of errors and make model updates and changes accessible to the whole community.

The model curation process was carried out using a workflow available in the *Human-GEM* GitHub repository. This procedure eliminated a large number of reactions and duplicates, corrected many metabolite formulas, rebalanced biochemical reaction equations, corrected the thermodynamic directionality of some reactions, and eliminated or blocked reactions that did not comply with rules such as conservation of mass and energy or were unnecessary.

In addition, a new biomass reaction was introduced with the aim of improving the accuracy of the analyses dependent on this function. For this purpose, its composition was based on experimentally obtained macromolecular stoichiometry of several human tissues and cells.

To evaluate the quality of the model, *Memote*, a community-maintained framework that attempts to evaluate GEMs using a series of statistical tests, was used. Furthermore, to increase the percentage of standardized annotation, a semi-automatic curation process was performed employing the *MetaNetX* database.

Human1 also contains improvements in gene-reaction associations necessary for high-throughput data integration. In addition, to facilitate transcriptome and proteome integration, transcript- and protein-reaction rules were also defined [15].

All of the aforementioned information, in addition to the evolution of the model since its inception, is published and accessible in the *Human-GEM* repository on GitHub from the account *SysBioChalmers*. This repository enables the entire community to access and review the updates and their rationale, as well as reproduce the experiments, as the functions and programs used are also publicly available.

The generic human model employed in this project was the last version *Human1* 19.0, derived from the Human-GEM GitHub repository. This version comprises 12,995 reactions, 8,456 metabolites, and 2,889 genes.

The creators of this model also propose a list of 57 known metabolic tasks whose blockage would prevent the survival of any cell.

2.3.2 Context-specific models and metabolic tasks

This generic cell model can be used to construct specific models of cells present in healthy tissues or malignant cells. To this end, several algorithms have been proposed. In this work, we use context-specific models computed using the *ftINIT* (fast Integrative Network Inference for Tissues) algorithm, which consists of constructing models using cell type-specific information on protein abundances contained in the Human Proteome Atlas as the primary source of evidence. This algorithm enables the incorporation of information on cell lines associated with diseases such as cancer, thereby facilitating a specific description (accessible online) of the metabolism of different types of human cells [8]. This will allow simulations at the tissue level, which will contribute to a better understanding of complex diseases.

In our case, 30 specific models of healthy tissues (table C.1) and also 18 models of MM cell lines (table C.2) have been developed. These models were provided by the Parallel Computing Architecture Research Group (GACOP) of the University of Murcia.

Moreover, for each context-specific model, and in consideration of the 57 known metabolic tasks whose blockage would prevent the survival of any cell (B.1), additional models have been developed (one for each specific model and task). Any resulting model is a modification of the context-specific one so that the viability of the aforementioned tasks is crucial for the model. That is, the task can be fulfilled if and only if the new model is feasible. This gives a total amount of 2,736 models to be analyzed.

2.4 Computing and checking MCS

In our case, the main interest is in preventing the proliferation of malignant cells. To do so, it is sufficient to ensure that at least one essential metabolic task cannot be fulfilled in the models associated with these cells. So, we must find cut sets for any of them, considering as target set all the possible modes of the model. In order to detect possible toxicities, we must do the same process for models associated with healthy tissues.

Computing MCS is a time-consuming effort. So, considering the high number of models to be analyzed, we have restricted our work to only compute MCS of length 1, also known as essential reactions. This limitation also takes into account the recommendations given by Hematopoietic Transplant and Cell Therapy group at H.C.U. Virgen de la Arrixaca (which the GACOP research group collaborates), that point to the complexity of in vitro and in vivo blocking reactions without adversely affecting the entire metabolic network. To calculate these MCSs, we applied a Python program (developed by Francisco Guil) to the constructed models.

The calculated minimal cut sets (MCSs) were subsequently reviewed and recalculated if any errors were detected during verification. The verification process was conducted using a script that followed these steps:

1. Calculated MCS loading for each specific metabolic task model.
2. Blocking each MCS using the COBRApy *bounds* method.
3. Find an optimal solution to the problem using the function *slim_optimize()*.
4. Incorrect MCS were reported if the function did not yield an expected value.

All the used methods are available in the COBRApy documentation [35].

2.5 Analyzing MCSs

After computing these MCSs, it is important to examine them carefully. We have utilized two tools for analysis. The first tool analyzes the behavior of the proposed MCSs, while the second tool searches for potential drugs to target them.

2.5.1 Metabolic Atlas

Metabolic Atlas is an advanced web platform that integrates all the GEMs developed for various organisms, providing users with access to them through a robust search engine. The platform provides detailed information on the molecules, with links to external databases to expand the available content. In addition, Metabolic Atlas includes 2D and 3D metabolic maps, making it easy to visualize entire metabolic pathways and specific cellular compartments. This tool is invaluable for contextualizing simulations or predictions related to cellular metabolism and for studying in-depth the metabolic pathways that constitute it.

However, the version included in Human Atlas (Human1 14.0) does not align with the version present in the Human-GEM model employed (Human1 19.0), resulting in a discrepancy in the number of reactions, metabolites, and genes. Consequently, these discrepancies should be taken into account when comparing the two versions of the analysis.

2.5.2 Accessing REST Web Services for biosciences

In the majority of cases, drugs in databases are associated with genes or proteins rather than reactions. Consequently, it is necessary to search for genes associated with our MCS before searching for associated drugs. Queries can be conducted through web-based data application programming interfaces (APIs) or by using data access libraries for Python. *MyGene.info* database will be used for genes and *chEMBL* database for drugs, as they are open-source and continuously updated.

- ***MyGene.Info* Python package:** MyGene.Info is a resource dedicated to querying gene annotation data. Specifically, we utilize the MyGene wrapper, a user-friendly Python interface for accessing MyGene.Info services. This resource enables us to obtain the names and various IDs of genes associated with reactions of interest, as Genome-Scale Metabolic (GEM) models exclusively use Ensembl IDs to define genes. The code employed is a variant of what is provided by biothings/mygene.py on their GitHub repository [37].
- ***ChEMBL* webresource client:** ChEMBL is a well-established database in the fields of drug discovery and medicinal chemistry research. It provides access to standardized data on bioactivity, molecules, targets, and drugs extracted from multiple sources, including primary medicinal chemistry literature. Additionally, it is possible to access this data along with commonly used cheminformatics methods through the open-source RESTful service it has implemented. Using this resource, we have been able to consult all drugs related to a certain disease (in our case, multiple myeloma) or retrieve molecules that exhibit binding activity to specific targets of interest. The code used is a variant of what The ChEMBL Group has available on their GitHub account [25].

2.6 Software

The High-Performance Computing Architecture Research Group (GACOP) of the University of Murcia, Spain, provided all the necessary resources for this work. Specifically, we used the GACOP cluster, which consists of several x86_64 architecture computer nodes. An internal Gigabit Ethernet network connects these nodes and runs CentOS GNU/Linux 8.2.

In addition, the language used to carry out the tasks was Python 3.6, using the Jupyter Notebook platform, which includes the COBRApy package used to perform the analysis and exploration of the models under consideration. Finally, Gurobi version 10.0.1 was used to solve the associated linear programming (LP) problems.

Chapter 3

Results

As stated before, we have computed MCSs of length 1 for all models associated with healthy tissues and multiple myeloma lines.

The interpretation of the MCS obtained for each model varies depending on whether it is associated with essential metabolic tasks for healthy tissues or MM lines. In the first case, they represent a potential toxicity. That is their blockage damages a healthy cell. On the other hand, for MM lines, they represent potential therapeutic targets.

In that sense, MCS in healthy tissues will be reactions to preserve when searching for targets for MM. On the other hand, those associated with cancer cells are reactions that we can attack since they are exclusive to MM lines and, in theory, they will minimize the toxic effects of possible therapies.

This approach has been explored both individually in any MM line or in all lines considered as a whole.

All scripts and results obtained can be consulted at https://github.com/fjavierlopezc/MScBioinformaticsThesis.

3.1 MCSs obtained from each model

Firstly, we have studied those reactions whose blockage involves the death of one of the different tissues analyzed. We have designated these reactions as toxicities. As a preliminary step, we have examined the toxicities associated with the generic model, as these are also toxicities of the other models. In total, there are 441 generic toxicities (MCS from the generic model).

The next stage of the process involves calculating MCS from each model in order to identify differences between the healthy tissue and MM-line models. A total of 912 and 972 reactions were identified as toxicities for one or more tissues and MM lines, respectively. Avoiding generic MCS, we obtained 471 tissue-specific MCS and 531 MM line-specific MCS.

The results of this calculation are presented in 3.1. It can be seen that the IDs starting with "ACH" correspond to MM line models (see correspondences with line names in C.2).

Model Name	Total MCSs	Specific MCS
Human-GEM	441	-
Adipose tissue	638	197
Adrenal gland	612	171
Bladder	664	223
Blood	637	196
Blood vessel	682	241
Brain	646	205
Breast	715	274
Colon	623	182
Esophagus	642	201
Heart	626	185
Lung	671	230
Kidney	657	216
Muscle	717	276
Nerve	680	239
Pancreas	598	157
Pituitary	628	187
Small intestine	595	154
Stomach	612	171
Vagina	698	257
Prostate	635	194
Skin	688	247
Testis	612	171
Thyroid	667	226
Uterus	729	288
Cervix uteri	654	213
Fallopian tube	716	275
Liver	626	185
Ovary	655	214
Salivary gland	664	223
Spleen	687	246
ACH_000838	784	343
ACH_000889	702	261
ACH_000512	710	269
ACH_000426	619	178
ACH_000436	677	236
ACH_000363	683	242
ACH_000183	711	270
ACH_000204	636	195
ACH_000541	670	229
ACH_000576	647	206
ACH_000588	676	235
ACH_000653	709	268
ACH_000714	643	202
ACH_000658	670	229
ACH_000763	680	239
ACH_000817	713	272
ACH_000821	620	179
ACH_000829	663	222

Table 3.1: **Summary of total and specific MCSs for different models**

3.2 Toxicities and toxicity levels

Once the MCSs have been compiled individually for each model, we can inspect those that are common to healthy tissues to produce a level ranking based on the number of different models they affect, considering each level as a toxicity level.

Firstly, we have 12093 reactions that are nontoxic for any healthy tissue. In the opposite direction, we have reactions that are generic toxicities and MCSs for all models. The remaining non-generic toxicities have been classified into five different levels according to the number of tissues affected. Figure 3.1 summarizes the number of specific toxicities in each group.

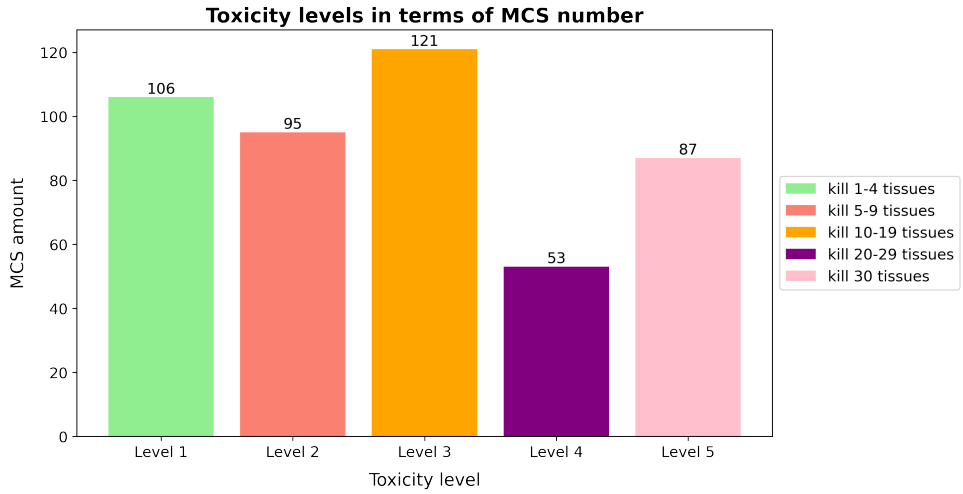

Figure 3.1: **Healthy tissues toxicities classification.** Image, which represents the number of reactions that can cause the death of different healthy cells.

For this study, we have focused on those reactions with no or few toxicities (toxicity levels 0, 1). We have grouped them into two large blocks: low toxicity (0 to 4 toxicities) and medium toxicity (5 to 9 toxicities). The analysis of MCSs with medium toxicity can be found in Appendix D.

3.3 Potential therapeutic targets for the MM lines studied

3.3.1 Therapeutic targets encountered for each line

Despite the high variability between lines, we can inspect each line's MCS individually and their possible associated toxicities for healthy tissues.

The number of targets found for each line is summarized in table 3.2. The columns indicate the name of the line, the number of total targets found, and the number of targets with low (tox. level 1) or medium (tox. level 2) toxicity associated with each line. Medium toxicity refers to a number of MCS, which affects a number between 5 and 9 healthy tissues.

MM lines	Total MCS	Tox. level 1	Tox. level 2
KMM1	702	15	28
AMO1	784	26	40
HUNS1	663	32	47
EJM	620	5	23
RPMI8226	713	26	43
MM1S	680	8	27
KMS11	643	19	32
KMS18	670	9	23
JJN3	709	16	43
KMS26	676	8	22
KMS27	647	15	33
KMS34	670	6	19
INA6	710	36	55
OCIMY7	677	10	24
KMS20	619	13	27
SKMM2	683	26	76
LP1	636	13	29
L363	711	33	51

Table 3.2: **MCS from each MM line with low and medium toxicity**

3.3.2 Common targets for several MM lines

Subsequently, we have also classified reactions that result in the demise of at least one of the MM lines under analysis. The targets identified have been categorized according to the number of MM lines they affect (see Figure 3.2).

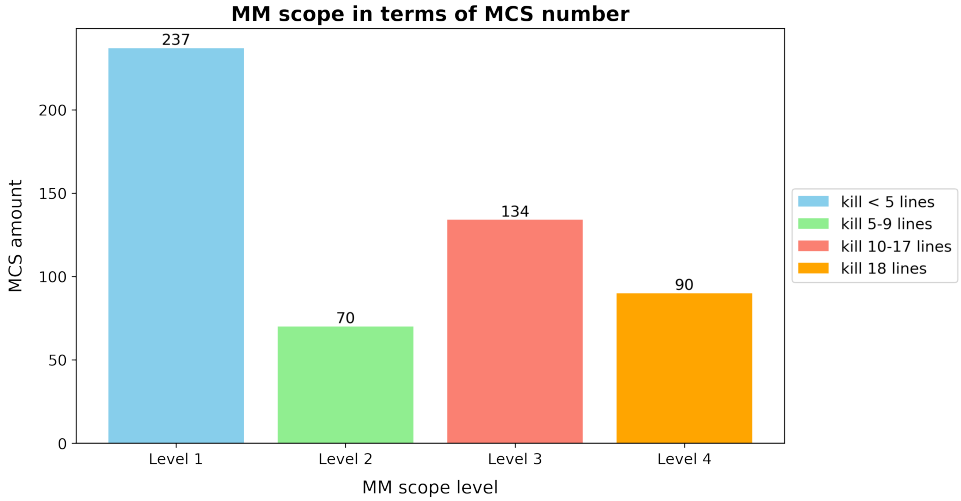

Figure 3.2: **MM lines MCS classification.** Image which contains the number of common reactions considered MCS classified by the number of MM lines whose they affect.

We have examined those targets that are common to all MM lines (level 4 in figure 3.2), but 74 of these 90 reactions are toxic to all healthy tissues, and the other 16 reactions are toxic to a significant number of healthy tissues (with a minimum number of toxicities to 17 of the 30 tissues), so they have been dismissed.

Instead, we have focused on three related analyses: low-toxicity targets in several MM lines, combinations of these MCSs, and medium-toxicity targets in several MM lines.

Common low-toxicity targets in several MM lines

Due to the high variability among the different MM lines, the subsequent analysis will focus on the set of reactions whose inactivation induces the death of at least ten multiple myeloma lines. With these restrictions, we have identified seven reactions. Table 3.3 presents reaction IDs, along with the metabolites that comprise the reaction and the subsystem to which they belong.

ID MCS	MM lines	Metabolite associated names	Subsystem
MAR06660	17	9-cis-retinoate <=> retinoate	Retinol metabolism
MAR06644	15	9-cis-retinal + H+ + NADH <=> 9-cis-retinol + NAD+	Retinol metabolism
MAR08702	15	9-cis-retinal <=> retinal	Retinol metabolism
MAR04623	16	glucono-1,5-lactone-6-phosphate + H2O $->$ 6-phospho-D-gluconate + H+	Pentose phosphate pathway
MAR04473	14	6-phospho-D-gluconate + NADP+ $->$ CO2 + NADPH + ribulose-5-phosphate	Pentose phosphate pathway
MAR02598	12	CoA + O-propanoylcarnitine <=> L-carnitine + propanoyl-CoA	Carnitine shuttle (M)
MAR06386	12	glucose-6-phosphate + Pi $->$ glucose-6-phosphate + Pi	Transport reactions

Table 3.3: **MCSs summary for several lines with low toxicity**

Combining more than one low-toxicity target

It was of interest to explore the possibility of **combining pairs of these reactions** in order to kill the 18 MM lines.

The results of 3.4 indicate that there are four possible combinations. In each case, one of the reactions in the combination is involved in the retinol metabolism (with toxicity in the blood model), and the other some related to the pentose phosphate pathway (no toxicity).

Reaction Pair	Subsystem	MM lines number	Toxicities
MAR06644 MAR04623	Retinol metabolism, Pentose phosphate pathway	18	Blood
MAR06660 MAR04473	Retinol metabolism, Pentose phosphate pathway	18	Blood
MAR06660 MAR04623	Retinol metabolism, Pentose phosphate pathway	18	Blood
MAR08702 MAR04623	Retinol metabolism, Pentose phosphate pathway	18	Blood

Table 3.4: **Association of pairs of reactions that kill all lines**

In order to enlarge the number of possible targets, we have also studied MCSs in multiple MM lines with toxicities in 5 to 9 healthy tissues. This study has ended with the identification of 12 additional targets that kill between 11 and 17 lines. A summary of these targets is presented in table D.1.

Chapter 4

Discussion

As a final step, we proceeded to interpret the obtained results. In our case, we will start by discussing the number of MCSs computed for each model and then their classification according to the number of toxicities and MM lines that affect them.

We will conduct a brief analysis of the issue of studying low-toxicity MCS on a line-by-line basis. However, the majority of the discussion will be centered around the study of the reactions that form our most promising MCSs. These are the ones that killed more than 10 lines and had less than 5 associated toxicities. To accomplish this, we will use *COBRApy* and *Metabolic atlas*.

We have also begun the search for possible treatments against them. This has been possible by accessing the web services of different drug databases that have implemented open-source APIs (*chEMBL* and *myGene.info*). To carry out this analysis, we first access the genes that control the reactions of interest, store them and search for matches to possible drugs that target one of the proteins encoded by these genes. Of course, we have to check if there really is a link to multiple myeloma and cancer in general. We have done this by accessing approved drugs for both terms in the database and then cross-checking with experimental studies published in the literature.

Based on the above, we screened for individual lineage targets and multi-lineage targets.

Toxicities and targets for each MM line

Table 3.1 shows a great number of MCS for each model. The third column will be our focus as it excludes the MCSs from the generic model. A first examination shows that the models from cancer lines exhibit a higher number of MCSs. This indicates that we may have identified potential pathways of attack that do not affect healthy tissues.

On the other hand, looking at the ranking of toxicity levels in figure 3.1, we observe 53 reactions affecting between 20-29 tissues and 87 reactions affecting all 30 tissues. Clearly, those reactions can't be used as therapeutic targets due to their high toxicity.

We conclude that it is really important to study toxicities before proposing targets for multiple myeloma lines.

We can perform analyses by inspecting the MCS of each line and their associated toxicities. Looking at table 3.2, we can see that there is quite some variability in terms of the number of targets for the different MM lines. Focusing on the number of targets with low toxicity, we range from 5 targets for the EJM line to 36 for INA6. This indicates that there should be a high variability in the number of possible pathways of attack depending on the line. Moreover, usually, MM patients present more than a line at once, so the most relevant MCSs will be those that are common to several lines.

Clearly, it would also be desirable to explore each line separately to broaden the range of treatment opportunities, even if they are only effective in a specific MM line. After having tested this analysis individually with several MM lines, we can say that in addition to the variability observed in the number of possible therapeutic targets, there is also variability in the identity of the MCS and in the subsystems to which they belong, depending on the line selected.

Metabolic tasks and subsystems for targets that are common to several MM lines

We have already seen that targets that affect all the tissues considered are extremely toxic. So, we must focus on MCS, which are common to several MM lines while having low toxicities.

All those promising MCSs affected between 12 and 17 lines. It should be noted from table 3.3 that the majority of these targets are located within the retinol metabolism and pentose phosphate pathway subsystems. Furthermore, the only targets that exhibit toxicity are linked to retinol metabolism, and toxicity is exclusively observed in the blood tissue.

An **inspection of the affected tasks** revealed that, depending on the type of line and the subsystem to which the MCS belongs, the same task is not always affected (see Table 4.1). Four MCSs (those associated with the retinol metabolism and carnitine shuttle) only affect metabolic task 57, i.e., biomass growth under Ham's medium. In the remaining cases, we also find other tasks affected. Specifically, sometimes the MCS also affects tasks 5 to 12 (related to de novo ribonucleotide and deoxyribonucleotide synthesis) and tasks 49 to 56 (related to de novo synthesis of phospholipids, as well as the metabolism of vitamins and important cofactors).

ID MCS	Toxicities	Task affected
MAR06660	Blood	Growth
MAR06644	Blood	Growth
MAR08702	Blood	Growth
MAR04473	Nothing	Growth or others
MAR04623	Nothing	Growth or others
MAR02598	Nothing	Growth
MAR06386	Nothing	Growth or others

Table 4.1: **MCSs tasks and toxicities summary**

The a posteriori analysis proceeded with **the identification of the subsystem** to which these MCSs belong. This was achieved through the utilization of the *Metabolic Atlas* portal, in conjunction with command-line exploration. As can be observed, 4 subsystems have been identified:

- The **transport reactions** subsystem: it is responsible for the exchange of substances between cellular compartments and the external environment. In our case, we identified one transport MCS (MAR06386), which appears to be essential for the transport of inorganic phosphorus (Pi) towards the endoplasmic reticulum (ER).

  In the *Human-GEM model*, six reactions (all transport reactions) are capable of carrying out this function. In healthy tissue models, there is typically more than one of these active reactions (but not always), whereas in MM lines, there is only one. This can be either MAR01621 (in six cases) or MAR06386 (in the other 12). In instances where MAR01621 is active, this reaction serves as an MCS for MM lines. This MCS exhibits toxicities in blood and ovaries, the only two tissues that include this reaction as the sole Pi transport between the reticulum and cytoplasm. Its occurrence is contingent upon the SLC37A4 gene (which regulates this reaction), which is associated with MM. In cases where MAR06386 is active, this reaction is an MCS for those MM lines without toxicities. However, the MAR06386 reaction also exchanges glucose-6-phosphate (G6P), which is where the pentose phosphate cycle comes into play.

- The **Pentose Phosphate Pathway (PPP)** subsystem: is a fundamental pathway in cellular metabolism that has both biosynthetic and oxidative functions [33]. It represents the main source of NADPH for cancer cells through its oxidative branch and the supply of ribose-5-phosphate for nucleotide synthesis [6].

Furthermore, studies have indicated that PPP becomes a crucial metabolic mechanism of MM cell survival when MM cells adapt to the suppression of the EGFR protein, which is involved in cell signaling pathways that control cell multiplication and survival [3].

Normally, this pathway occurs in the cytosol, although it has recently been shown that it can also occur in the endoplasmic reticulum (ER). In our case, it is precisely in the ER where we obtain two MCS that belong to this pathway. The two pathways in question are contiguous in the pathway (MAR04623 and MAR04473 in figure 4.1), and their blockage would prevent the formation of ribulose-5-phosphate, a metabolite involved in nucleotide synthesis that seems to be essential in most MM lines when generated in the ER. This is consistent with recent studies that have explored this reticular pathway via gene silencing of hexose-6-phosphate dehydrogenase (H6PD), an enzyme that is very similar to G6PD and which controls one of the two MCSs mentioned (MAR04623 in figure 4.1) [31].

In both cases, silencing has been shown to decrease the content of D-ribose and NADPH intermediates and considerably alter cell proliferation. Therefore, our analysis supports the hypothesis that there seems to be an inevitable interference between reticular and cytosolic PPP. This probably implies the presence of reciprocal control mechanisms that cannot be defined, according to current data [24].

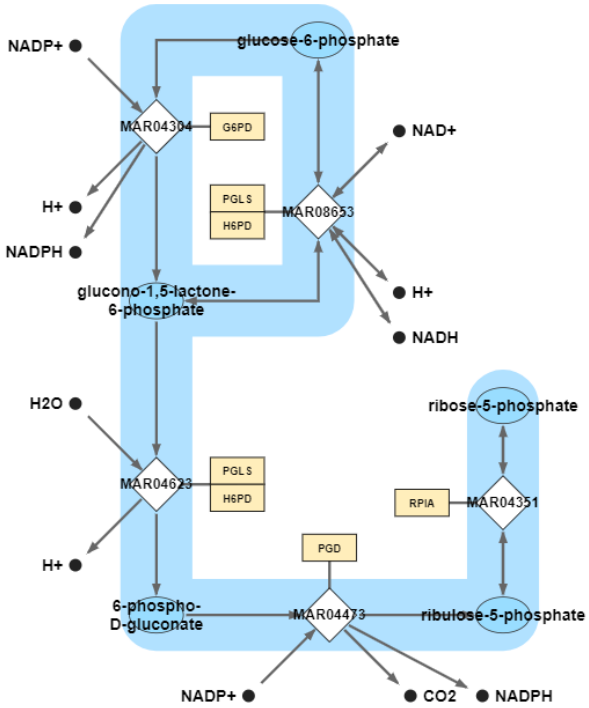

Figure 4.1: **PPP ER route scheme.** Image extracted from *Metabolic Atlas*, showing the reticular segment of the pentose phosphate pathway.

- The **Retinol metabolism** subsystem: Retinoids are a class of compounds structurally related to vitamin A. Retinoic acid, the active metabolite of retinol, regulates a wide range of biological processes, including development, differentiation, proliferation, and apoptosis. These compounds exert their effects through several binding proteins, including cellular retinol-binding protein (CRBP), retinol-binding proteins (RBP), cellular retinoic acid-binding protein (CRABP), and nuclear receptors: retinoic acid receptor (RAR) and retinoid X receptor (RXR) [22].

In our study, we identified three MCSs affecting between 15 and 17 MM lines related to retinol metabolism (two in the cytosol and one in the ER). Retinol metabolism in the cytosol is an extremely complex subsystem, making it difficult to analyze, particularly in the cytosol. However, we have made significant observations on the MCS in the ER, illustrated in the accompanying figure 4.2.

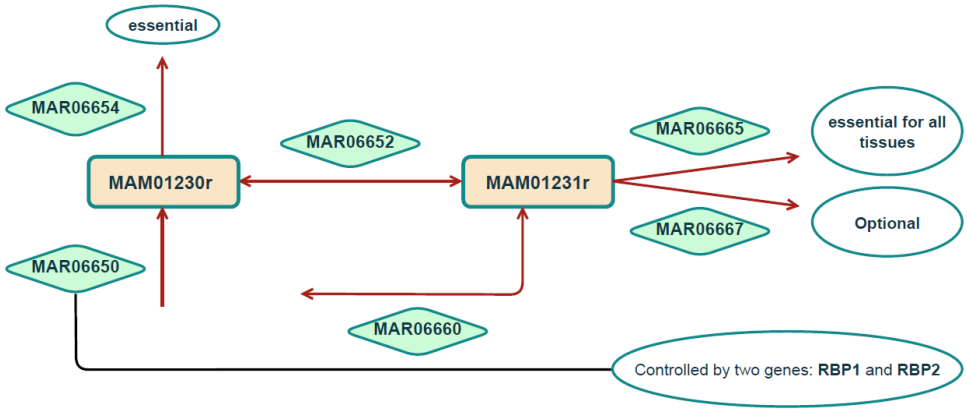

Figure 4.2: **Retinol metabolism ER routes scheme.** Schematic representation of the information extracted from the model that does not match the version shown in *Metabolic Atlas* as it presents updated reactions.

9-cis-retinal (MAM01230r) and 9-cis-retinoate (MAM01231r) are two essential metabolites. 9-cis-retinal is produced by the MAR06650 reaction, which is controlled by the RBP1 and RBP2 genes, or by the MAR06652 reaction in a reversible manner, which also produces 9-cis-retinoate. The removal of 9-cis-retinal occurs via the MAR06654 reaction, while the removal of 9-cis-retinoate occurs via the MAR06665 reaction (both of which are essential) or via the MAR06667 reaction (which is non-essential).

In the absence of RBP1 and RBP2, the MAR06652 reaction is forced to produce 9-cis-retinal, activating the MAR06660 reaction, which converts another substrate to 9-cis-retinoate. This process becomes essential (MCS) in all 15 MM lines and in blood tissue, the only tissue affected by the toxicity of this MCS.

Our analysis indicates that genes RBP1 and RBP2 seem to be important in retinol metabolism. However, the absence of these genes in certain models is not clearly understood.

Furthermore, the literature on this metabolic subsystem focuses on the use of retinoids as a cancer treatment rather than as a therapeutic target [32]. In particular, retinoids have been shown to suppress carcinogenesis in various animal models and, in humans, can reverse premalignant epithelial lesions, induce myeloid cell differentiation, and prevent lung, liver, and breast cancer [20].

- The **Carnitine shuttle** subsystem: Fatty acid oxidation utilizes L-carnitine to facilitate the transport of acyl residues to the mitochondria via the carnitine shuttle [29]. Our study identified one MCS related to this subsystem, which is essential for 12 of the 18 MM lines and is linked to propanoyl-CoA production.

  In addition to alterations in glucose and amino acid metabolism, there is evidence that cancer cells can modify lipid metabolism. Carnitine is crucial for the transport of long-chain fatty acids into the mitochondria for Beta-oxidation. The concerted function of carnitine transporters creates a collaborative network relevant to metabolic reprogramming in cancer cells [27].

Studies have proposed that a reduction in propionyl-CoA metabolism contributes to metabolic remodeling and facilitates hepatocarcinogenesis. This may provide potential targets for the development of new therapeutic strategies against hepatocellular carcinoma (HCC) [34].

Genes and potential drugs associated with low toxicity targets

In order to gain insight into the underlying **genetic mechanisms of these reactions**, we have conducted a comprehensive analysis of the genes that regulate them and are present in the model. The table below provides a comprehensive overview of the IDs from different databases for each of the reactions, along with the genes that control them.

ID MCS	Ensembl	Entrez	Name	Symbol
MAR06660	-	-	-	-
MAR06644	ENSG00000157326	10901	dehydrogenase/reductase 4	DHRS4
	ENSG00000135437	5959	retinol dehydrogenase 5	RDH5
	ENSG00000265203	5949	retinol binding protein 3	RBP3
	ENSG00000187630	317749	dehydrogenase/reductase 4 like 2	DHRS4L2
	ENSG00000139547	8608	retinol dehydrogenase 16	RDH16
	ENSG00000170786	195814	short chain dehydrogenase/reductase family 16C5	SDR16C5
	ENSG00000198099	127	alcohol dehydrogenase 4 (class II), pi polypeptide	ADH4
	ENSG00000025423	8630	hydroxysteroid 17-beta dehydrogenase 6	HSD17B6
	ENSG00000140522	6017	retinaldehyde binding protein 1	RLBP1
MAR08702	-	-	-	-
MAR04473	ENSG00000142657	5226	phosphogluconate dehydrogenase	PGD
MAR04623	ENSG00000049239	9563	hexose-6-phosphate dehydrogenase/glucose 1-dehydrogenase	H6PD
MAR02598	ENSG00000157184	1376	carnitine palmitoyltransferase 2	CPT2
	ENSG00000095321	1384	carnitine O-acetyltransferase	CRAT
MAR06386	ENSG00000281500	2542	solute carrier family 37 member 4	SLC37A4

Table 4.2: **Interesting MCS genes associated**

Table 4.2 illustrates that the reactions linked to retinol metabolism are challenging to study at the gene level. Two of the reactions lack associated genes, while the third reaction is controlled by eight genes (isoenzymes). Consequently, to block it from a genetic perspective, it is necessary to inhibit all eight genes simultaneously, which would likely result in a high number of toxicities.

Conversely, we also considered it worthwhile to conduct a **search for drugs** that may have activity against the genes of interest. To this end, we accessed the chEMBL database API and retrieved the results in our analysis script, which is available in the indicated Github repository. Upon querying by disease, we observed that there are a total of 1734 approved drugs for cancer in general and 436 for multiple myeloma. Potential drugs were identified for some of the genes of interest. Specifically, we have found 9 drugs with binding activity to the protein encoded by the CPT2 gene: 6 for PGD, 4 for SLC37A4, 3 for ADH4, and 1 for CRAT.

The proceedings were then directed towards the identification of correlations between the outcomes of the pharmacological agents that have been approved for the treatment of both cancer and MM and those observed for the genes that have been identified as being of interest. The objective is to identify pharmaceutical agents that have been approved for either cancer or MM and that are capable of modulating the expression of the aforementioned genes. In our case, we identified matches for three genes of interest summarised in table 4.3: CPT2, PGD, and SLC37A4.

Target name	Gen symbol	Target chEMBL ID	Molecule drug name
Carnitine palmitoyltransferase 2	CPT2	CHEMBL3238	PIOGLITAZONE ROSIGLITAZONE
Prostaglandin-H2 D-isomerase	PGD	CHEMBL4334	INDOMETHACIN KETOROLAC IBUPROFEN NAPROXEN
Glucose-6-phosphate translocase	SLC37A4	CHEMBL3217398	CHLOROGENIC ACID

Table 4.3: **chEMBL information coincidences with diseases**

Obviously, these coincidences need to be checked against the experimental literature, and that was our next step.

The thiazolidinediones group, which includes pioglitazone, rosiglitazone, and troglitazone, is indicated in type 2 diabetes mellitus. This is due to their ability to promote insulin sensitivity and improve blood glucose uptake through agonism of peroxisome proliferator-activated receptor gamma (PPAR-gamma). PPARs are ligand-activated transcription factors that are involved in the expression of more than 100 genes and affect numerous metabolic processes, particularly lipid and glucose homeostasis [28].

The CPT2 gene plays a role in the carnitine shuttle, a metabolic pathway that facilitates the transport of long-chain fatty acids into the mitochondria for subsequent oxidation [29]. With regard to their relationship to cancer, they have been subjected to rigorous testing and dismissed as a consequence of the occurrence of numerous adverse effects in the context of bladder cancer (see, for example, [27]).

One of the biological functions of prostaglandin-H2 D-isomerase (PGD) is to facilitate the recruitment of inflammatory cells [30]. This property has led to the development of a class of non-steroidal anti-inflammatory drugs (NSAIDs) that are related to PGD [18]. The use of these drugs against cancer has been the subject of research. Indomethacin has been described as having anticancer properties through the activation of protein kinase R (PKR) and the downstream phosphorylation of eIF2-alpha, which inhibits protein synthesis [19]. Furthermore, ibuprofen has been considered a potential prophylactic agent against Alzheimer's disease, Parkinson's disease, and breast cancer [21]

Finally, chlorogenic acid (CGA) is a drug that targets the SLC37A4 gene. This gene regulates the transport of glucose-6-phosphate from the cytoplasm to the lumen of the endoplasmic reticulum (ER), thereby maintaining glucose homeostasis. It also plays a role in ATP-mediated calcium sequestration in the ER lumina. Furthermore, in cancer, it has been associated with glioblastoma progression and regulation of metabolic reprogramming [36].

In contrast, G6PT (G6P transporter) deficiency is responsible for glycogen storage disease type Ib (GSD-Ib), an autosomal recessive disorder associated with defective metabolic and myeloid phenotypes. Several types of mutations have been identified in the SLC37A4 gene that affects G6PT function, thus necessitating consideration of potential toxicity when blocking it as a therapeutic target [23].

Indeed, CGA has been demonstrated to possess the potential to induce GSD1b manifestations by inhibiting glycogenolysis and gluconeogenesis [23].

Target combination and medium toxicity targets

According to our analysis, it is possible to block all 18 MM lines by combining pairs of promising reactions 3.4. The combinations always resulted in a pair composed of one reaction belonging to retinol metabolism and one associated with the pentose phosphate pathway. A single toxicity was associated with these combinations, which always came from the reaction belonging to retinol metabolism.

However, as discussed above, the practical implementation of blocking more than one reaction is very complex, as it would lead to unfeasible changes in the network.

Finally, the analysis of average toxicities was not discussed in depth, but is available in table D.1.

Chapter 5

Conclusion

In recent years, genome-scale models have revolutionized the field of systems biology, proving to be a very useful tool in the study of cellular metabolism and disease through predictions and *in silico* simulations.

On the other hand, they bring us closer to personalized medicine thanks to the implementation of algorithms capable of generating models for specific contexts, such as different cell types or malignant cells. They also permit the analysis of metabolic tasks necessary to maintain the viability of any cell and try to compute sets of reactions, known as Minimal Cut Sets, whose inhibition blocks these tasks.

In our study, we have attempted to exploit this approach by working with a large number of models. In particular, we have worked on individual reactions (MCS of length 1), as this allows us to study the network in more depth and obtain more realistic results.

Given the large number of models analyzed, the main objective was to find specific therapeutic targets in multiple myeloma, i.e. targets whose inhibition would cause the least damage to the various healthy tissues and the maximum damage to the cancer lines. This approach aims to identify and analyze not only very specific targets in each multiple myeloma line but also look for common targets that could affect a large number of lines.

This work has shown the importance of considering toxicity while looking for targets in multiple myeloma. In addition, the cancer lines analyzed appear to have very specific pathways of attack.

On the other hand, if we focus exclusively on MCS with low toxicity, we conclude that the number of computed MCS is drastically reduced, and there is a large variability depending on the line studied. This variability suggests the presence of effective targets only for the selected cell line.

If the same approach is continued for the MCS common to several lines, the possibilities are greatly reduced. The endoplasmic reticulum seems to be an essential organelle for cancer cells since the majority of the most promising reactions take place there. In our analysis, these reactions are distributed over 4 cancer-related subsystems that could be pairwise combined to reach all the lines considered, without increasing their toxicity, but taking into account the complexity of this process.

According to our study, some of the genes that control these promising reactions are highly relevant in cancer, such as H6PD or SLC37A4, which is consistent with the existing literature.

Finally, we have turned our attention to suggesting possible treatments for the targets under consideration. Specifically, cancer-related targets with potential treatments include CPT2, PGD and SLC37A4. The most promising treatment is chlorogenic acid, which acts against SLC37A4 and has already been shown to be effective for multiple myeloma.

It should be noted that such predictions need to be verified experimentally, as the models are, to some degree, limited in terms of biological information due to their experimental nature. In particular, they omit regulatory mechanisms, cannot predict the adaptive mechanisms of tumour cells and neglect changes in the environment, which play an important role in the predictions. These considerations affect the reliability of the results obtained to some extent, and there is still much room for improvement in modelling studies.

Some suggestions in the short term could be to extend the search to minimal cut reaction sets of length 2 or more, to repeat the same study for genetic MCS and contrast it with the reactions, or to include many more different lines showing the multiple myeloma phenotype.

In the long term, it is of obvious interest to model cells of the immune system, modelling the connections and regulations between cells to create a supermodel organism and trying to model the environment in some way.

Nevertheless, GEMs have proven to be an effective first step in the search for drug targets, which can be complemented by other genetic screening techniques to obtain more reliable results. Thus, despite its limitations, in silico identification of metabolic vulnerabilities in cancer cells to find therapeutic targets promises to be an efficient and useful technique in the field of personalized medicine. By using advanced technological methods and tools, such as computer simulations and virtual experiments, we can significantly reduce the need for traditional wet lab procedures, saving a lot of time and valuable resources.

The framework developed in this work has been published to make it accessible to the community.

Bibliography

[1] J. Nielsen A. Mardinoglu. "Systems medicine and metabolic modelling". In: *Journal of Internal Medicine* 271.2 (2012), pp. 142–154. DOI: 10.1111/j.1365-2796.2011.02493.x.

[2] Brunk et al. "Recon3D enables a three-dimensional view of gene variation in human metabolism". In: *Nature Biotechnology* 36 (2018), pp. 272–281. DOI: 10.1038/nbt.4072. URL: https://pubmed.ncbi.nlm.nih.gov/29457794/.

[3] Chen et al. "Multiple myeloma acquires resistance to EGFR inhibitor via induction of pentose phosphate pathway". In: *Scientific Reports* 5.9925 (2015). DOI: 10.1096/fj.201700870RR. URL: https://www.nature.com/articles/srep09925.

[4] Ng et al. "Constraint-Based Reconstruction and Analyses of Metabolic Models: Open-Source Python Tools and Applications to Cancer". In: 12 (2022). DOI: 10.3389/fonc.2022.914594. URL: https://doi.org/10.3389/fonc.2022.914594.

[5] Orth et al. "What is flux balance analysis?" In: *Nature Biotechnology* 28 (2010), pp. 245–248. DOI: 10.1371/journal.pcbi.1009294. URL: https://www.nature.com/articles/nbt.1614.

[6] Zhao et al. "Integrative Analysis of Bulk RNA-Seq and Single-Cell RNA-Seq Unveils Novel Prognostic Biomarkers in Multiple Myeloma". In: *Biomolecules* 12.12 (2022). DOI: 10.3390/biom12121855. URL: https://www.mdpi.com/2218-273X/12/12/1855#B34-biomolecules-12-01855.

[7] Agren et al. "Metabolic Network-Based Identification and Prioritization of Anticancer Targets Based on Expression Data in Hepatocellular Carcinoma". In: *Frontiers in Physiology* 9 (2018). DOI: doi.org/10.3389/fphys.2018.00916.

[8] Agren et al. "Reconstruction of Genome-Scale Active Metabolic Networks for 69 Human Cell Types and 16 Cancer Types Using INIT". In: *PLoS Computational Biology* 8.5 (2012). DOI: doi.org/10.1371/journal.pcbi.1002518.

[9] Alam et al. "Bioinformatics and network-based screening and discovery of potential molecular targets and small molecular drugs for breast cancer". In: *Frontiers in Physiology* 13 (2022). DOI: doi.org/10.3389/fphar.2022.942126.

[10] Blais et al. "Reconciled rat and human metabolic networks for comparative toxicogenomics and biomarker predictions". In: *Nat Commun* 8.8 (2017). DOI: 10.1038/ncomms14250.

[11] Kim et al. "Recent advances in reconstruction and applications of genome-scale metabolic models". In: *Current Opinion in Biotechnology* 23.4 (2012), pp. 617–623. DOI: 10.1016/j.copbio.2011.10.007.

[12] Mardinoglu A. et al. "Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease". In: *Nat Commun* 5.3083 (2014). DOI: 10.1038/ncomms4083.

[13] Moulin et al. "Combining Kinetic and Constraint-Based Modelling to Better Understand Metabolism Dynamics". In: *Processes* 9.10 (2021). DOI: doi.org/10.3390/pr9101701.

[14] Riemer et al. "A metabolite-centric view on flux distributions in genome-scale metabolic models". In: *BMC Systems Biology* 7.33 (2013). DOI: 10.1186/1752-0509-7-33.

[15] Robinson et al. "An atlas of human metabolism". In: *Science Signaling* 13.624 (2020). DOI: doi.org/10.1371/journal.pcbi.1002518.

[16] Vieira et al. “A pipeline for the reconstruction and evaluation of context-specific human metabolic models at a large-scale”. In: *PLoS Computational Biology* 18.16 (2022). DOI: 10.1371/journal.pcbi.1009294.

[17] Yasemi et al. “Modelling Cell Metabolism: A Review on Constraint-Based Steady-State and Kinetic Approaches”. In: *Processes* 9.2 (2021). DOI: doi.org/10.3390/pr9020322.

[18] Carla Amici et al. “Inhibition of viral protein translation by indomethacin in vesicular stomatitis virus infection: role of eIF2α kinase PKR”. eng. In: *Cellular Microbiology* 17.9 (Sept. 2015), pp. 1391–1404. ISSN: 1462-5822. DOI: 10.1111/cmi.12446.

[19] Claudia Brunelli et al. “The non-steroidal anti-inflammatory drug indomethacin activates the eIF2α kinase PKR, causing a translational block in human colorectal cancer cells”. eng. In: *The Biochemical Journal* 443.2 (Apr. 2012), pp. 379–386. ISSN: 1470-8728. DOI: 10.1042/BJ20111236.

[20] Mulazim H. Bukhari et al. “Chemotherapeutic/chemopreventive role of retinoids in chemically induced skin carcinogenesis in albino mice”. eng. In: *International Journal of Dermatology* 46.11 (Nov. 2007), pp. 1160–1165. ISSN: 0011-9059. DOI: 10.1111/j.1365-4632.2007.03425.x.

[21] Rabia Bushra and Nousheen Aslam. “An overview of clinical pharmacology of Ibuprofen”. eng. In: *Oman Medical Journal* 25.3 (July 2010), pp. 155–1661. ISSN: 2070-5204. DOI: 10.5001/omj.2010.49.

[22] Nathan Bushue and Yu-Jui Yvonne Wan. “Retinoid Pathway and Cancer Therapeutics”. In: *Advanced drug delivery reviews* 62.13 (Oct. 2010), pp. 1285–1298. ISSN: 0169-409X. DOI: 10.1016/j.addr.2010.07.003. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2991380/ (visited on 06/11/2024).

[23] Anna Rita Cappello et al. “The Physiopathological Role of the Exchangers Belonging to the SLC37 Family”. eng. In: *Frontiers in Chemistry* 6 (2018), p. 122. ISSN: 2296-2646. DOI: 10.3389/fchem.2018.00122.

[24] Vanessa Cossu et al. “Two high-rate pentose-phosphate pathways in cancer cells”. en. In: *Scientific Reports* 10.1 (Dec. 2020). Publisher: Nature Publishing Group, p. 22111. ISSN: 2045-2322. DOI: 10.1038/s41598-020-79185-2. URL: https://www.nature.com/articles/s41598-020-79185-2 (visited on 06/11/2024).

[25] Mark Davies et al. “ChEMBL web services: streamlining access to drug discovery data and utilities”. In: *Nucleic Acids Research* 43.Web Server issue (July 2015), W612–W620. ISSN: 0305-1048. DOI: 10.1093/nar/gkv352.

[26] Ali Ebrahim et al. “COBRApy: COnstraints-Based Reconstruction and Analysis for Python”. In: *BMC Systems Biology* 7.1 (Aug. 2013), p. 74. ISSN: 1752-0509. DOI: 10.1186/1752-0509-7-74. URL: https://doi.org/10.1186/1752-0509-7-74.

[27] Raheleh Farahzadi et al. “Clinical Significance of Carnitine in the Treatment of Cancer: From Traffic to the Regulation”. eng. In: *Oxidative Medicine and Cellular Longevity* 2023 (2023), p. 9328344. ISSN: 1942-0994. DOI: 10.1155/2023/9328344.

[28] C. Janani and B. D. Ranjitha Kumari. “PPAR gamma gene–a review”. eng. In: *Diabetes & Metabolic Syndrome* 9.1 (2015), pp. 46–50. ISSN: 1878-0334. DOI: 10.1016/j.dsx.2014.09.015.

[29] Barbara Juraszek and Katarzyna A. Nałęcz. “SLC22A5 (OCTN2) Carnitine Transporter-Indispensable for Cell Metabolism, a Jekyll and Hyde of Human Cancer”. eng. In: *Molecules (Basel, Switzerland)* 25.1 (Dec. 2019), p. 14. ISSN: 1420-3049. DOI: 10.3390/molecules25010014.

[30] Yoshihide Kanaoka and Yoshihiro Urade. “Hematopoietic prostaglandin D synthase”. eng. In: *Prostaglandins, Leukotrienes, and Essential Fatty Acids* 69.2-3 (2003), pp. 163–167. ISSN: 0952-3278. DOI: 10.1016/s0952-3278(03)00077-2.

[31] Cecilia Marini et al. “Discovery of a novel glucose metabolism in cancer: The role of endoplasmic reticulum beyond glycolysis and pentose phosphate shunt”. en. In: *Scientific Reports* 6.1 (Apr. 2016). Publisher: Nature Publishing Group, p. 25092. ISSN: 2045-2322. DOI: 10.1038/srep25092. URL: https://www.nature.com/articles/srep25092 (visited on 06/11/2024).

[32] Y. Muto, H. Moriwaki, and Y. Shiratori. “Prevention of second primary tumors by an acyclic retinoid, polyprenoic acid, in patients with hepatocellular carcinoma”. eng. In: *Digestion* 59 Suppl 2 (July 1998), pp. 89–91. ISSN: 0012-2823. DOI: 10.1159/000051435.

[33] Anna Stincone et al. “The return of metabolism: biochemistry and physiology of the pentose phosphate pathway”. en. In: *Biological Reviews* 90.3 (2015). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/brv.12140, pp. 927–963. ISSN: 1469-185X. DOI: 10.1111/brv.12140. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/brv.12140 (visited on 06/11/2024).

[34] Jiaqi Sun et al. “Decreased propionyl-CoA metabolism facilitates metabolic reprogramming and promotes hepatocellular carcinoma”. eng. In: *Journal of Hepatology* 78.3 (Mar. 2023), pp. 627–642. ISSN: 1600-0641. DOI: 10.1016/j.jhep.2022.11.017.

[35] The cobrapy core team. *Documentation for COBRApy*. 2022. URL: https://cobrapy.readthedocs.io/en/latest/index.html.

[36] Sima Torabidastgerdooei, Marie-Eve Roy, and Borhane Annabi. “A molecular signature for the G6PC3/SLC37A2/SLC37A4 interactors in glioblastoma disease progression and in the acquisition of a brain cancer stem cell phenotype”. eng. In: *Frontiers in Endocrinology* 14 (2023), p. 1265698. ISSN: 1664-2392. DOI: 10.3389/fendo.2023.1265698.

[37] Jiwen Xin et al. “High-performance web services for querying gene and variant annotation”. eng. In: *Genome Biology* 17.1 (May 2016), p. 91. ISSN: 1474-760X. DOI: 10.1186/s13059-016-0953-9.

[38] M Yasemi M.; Jolicoeur. “Modelling Cell Metabolism: A Review on Constraint-Based Steady-State and Kinetic Approaches”. In: *Processes* 9.2 (2021), p. 322. DOI: 10.3390/pr9020322.

Appendices

Appendix A

Abbreviations

CBM : Constraint-Based Modeling.

CGA : Chlorogenic acid.

COBRApy : Constraint-Based Reconstruction and Analysis in Python.

CS : Cut set.

EFM : Elementary Flux Mode.

ER : Endoplasmic reticulum.

FBA : Flux Balance Analysis.

GEM : Genome-scale metabolic model.

MCS : Minimal Cut Set.

GPR : Gene-Protein Rule.

LP : Linear Programming.

MM : Multiple Myeloma.

MT : Metabolic Task.

TR : Transport reactions.

Pi : Inorganic phosphorus.

PPP : Pentose phosphate pathway.

RM : Retinol metabolism.

Appendix **B**

Supplementary material 1

Metabolic tasks and their order.

Order	Description
MT-1	Aerobic rephosphorylation of ATP from glucose
MT-2	Aerobic rephosphorylation of GTP
MT-3	Aerobic rephosphorylation of CTP
MT-4	Aerobic rephosphorylation of UTP
MT-5	ATP de novo synthesis
MT-6	CTP de novo synthesis
MT-7	GTP de novo synthesis
MT-8	UTP de novo synthesis
MT-9	dATP de novo synthesis
MT-10	dCTP de novo synthesis
MT-11	dGTP de novo synthesis
MT-12	dTTP de novo synthesis
MT-13	Histidine uptake
MT-14	Isoleucine uptake
MT-15	Leucine uptake
MT-16	Lysine uptake
MT-17	Methionine Uptake
MT-18	Phenylalanine Uptake
MT-19	Threonine uptake
MT-20	Tryptophan uptake
MT-21	Valine uptake
MT-22	Glycerate 3-phosphate de novo synthesis
MT-23	Mitochondrial acetyl-CoA de novo synthesis
MT-24	Mitochondrial AKG de novo synthesis
MT-25	Erythrose 4-phosphate de novo synthesis
MT-26	Fructose 6-phosphate de novo synthesis
MT-27	Glyceraldehyde 3-phosphate de novo synthesis
MT-28	Glucose 6-phosphate de novo synthesis
MT-29	Mitochondrial oxaloacetate de novo synthesis
MT-30	Phosphoenolpyruvate de novo synthesis
MT-31	Pyruvate de novo synthesis
MT-32	Ribose 5-phosphate de novo synthesis
MT-33	Mitochondrial succinyl-CoA de novo synthesis
MT-34	Cholesterol de novo synthesis

MT-35	Protein synthesis from AAs
MT-36	Oxidative phosphorylation
MT-37	Oxidative decarboxylation
MT-38	Krebs cycle NADH
MT-39	Ubiquinol-to-proton
MT-40	Ubiquinol-to-ATP
MT-41	Beta oxidation of saturated FA
MT-42	Beta oxidation of long-chain FA
MT-43	Beta oxidation of odd-chain FA
MT-44	Beta oxidation of unsaturated fatty acid (n-9)
MT-45	Beta oxidation of unsaturated fatty acid (n-6)
MT-46	Beta oxidation of all NEFAs
MT-47	Choline uptake
MT-48	Inositol uptake
MT-49	Phosphatidylcholine de novo synthesis
MT-50	Phosphatidylethanolamine de novo synthesis
MT-51	Phosphatidylserine de novo synthesis
MT-52	Phosphatidylinositol de novo synthesis
MT-53	Thiamin phosphorylation to TPP
MT-54	Coenzyme A synthesis from pantothenate
MT-55	FAD synthesis from riboflavin
MT-56	Heme biosynthesis
MT-57	Biomass in Ham's medium

Table B.1: **Metabolic tasks and their order.**

Appendix C

Supplementary material 2

Healthy tissue models and MM cell lines

Healthy tissue models	
adipose_tissue	muscle
adrenal_gland	nerve
bladder	pancreas
blood	pituitary
blood_vessel	small_intestine
brain	stomach
breast	vagina
colon	prostate
esophagus	skin
heart	testis
lung	thyroid
kidney	uterus

Table C.1: **Healthy tissue models**

ID modelo	Cell line
ACH_000889	KMM1
ACH_000838	AMO1
ACH_000829	HUNS1
ACH_000821	EJM
ACH_000817	RPMI8226
ACH_000763	MM1S
ACH_000714	KMS11
ACH_000658	KMS18
ACH_000653	JJN3
ACH_000588	KMS26
ACH_000576	KMS27
ACH_000541	KMS34
ACH_000512	INA6
ACH_000436	OCIMY7
ACH_000426	KMS20
ACH_000363	SKMM2
ACH_000204	LP1
ACH_000183	L363

Table C.2: **MM cell lines**

Appendix **D**

Supplementary material 3

Common medium-toxicity targets

If we look at the table containing MCS with medium toxicity (table D.1, according to the subsystem, the majority of these novel targets are concentrated within the glycerophospholipid metabolism subsystem. Furthermore, we see that the number of different toxicities is too high, with 7 being the lowest value recorded. These targets are responsible for the death of between 16 and 17 lines, and they exhibit seven common toxicities: blood, blood vessel, esophagus, muscle, salivary gland, skin and small intestine.

ID MCS	Toxicities	MM lines	Subsystem
MAR02594	9	11	Carnitine shuttle
MAR04336	7	16	Folate metabolism
MAR00651	7	16	Glycerophospholipid metabolism
MAR00636	8	17	Glycerophospholipid metabolism
MAR00625	8	17	Glycerophospholipid metabolism
MAR00638	8	17	Glycerophospholipid metabolism
MAR00589	8	17	Glycerophospholipid metabolism
MAR00597	8	17	Glycerophospholipid metabolism
MAR04917	8	14	Transport reactions
MAR04898	6	17	Transport reactions
MAR05029	7	16	Transport reactions
MAR06505	7	17	Transport reactions

Table D.1: **Promising MCSs with medium toxicity**

Appendix E

Supplementary material 4

Relating the Master thesis to the Sustainable Development Goals (SDGs)

On 25 September 2015, world leaders adopted a set of global goals to eradicate poverty, protect the planet and ensure prosperity for all as part of a new sustainable development agenda. Specifically, 17 goals with specific targets were set.

Multiple myeloma is an extremely complex challenge for medicine because there are different types of myeloma and different levels of aggressiveness. In addition, although patients are now living longer and with a better quality of life, it is still considered an incurable disease and treatments need to be properly individualised. In this sense, our work contributes to progress in the search for effective treatment for patients, with the aim of ensuring a healthy life and promoting well-being for all at all ages (SGD 3).

On the other hand, this work aims to leverage investment in research to develop solutions to diseases and to further promote research and development in this area, which is reflected in SGD 9, which aims to promote innovation.

Finally, through this work we want to give visibility to the importance of fulfilling these goals and support the idea of raising awareness and achieving collaboration between governments, the private sector and civil society.