7-2024

Data-driven Drug Repurposing for Immune System-related Diseases

Sabyasachi Mohanty
University of Nebraska-Lincoln, smohanty12@huskers.unl.edu

Data-driven Drug repurposing for immune system-related diseases

By

Sabyasachi Mohanty

A Thesis

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfilment of Requirements

For the Degree of Master of Science

Major: Biochemistry

Under the Supervision of Professor

Tomas Helikar

Lincoln Nebraska

July, 2024

Data-driven Drug repurposing for immune system-related diseases

Sabyasachi Mohanty, M.S.

University of Nebraska, 2024

Advisor: Tomas Helikar

In this research, we explore a data-driven approach for drug repurposing to enhance the immunomodulatory effect by integrating pattern-based search and genome-scale metabolic modeling. This research helps to find a solution to the preexisting problems of drug discovery which includes expenses and a high amount of time consumption. By leveraging the pre-existing data of the approved drugs to identify the major metabolic pathway changes, we can find new drugs with similar effects with less off-target effects.

The focus on the immune system modulating drugs is due to the high prevalence of immune system-involved diseases and the growing demand for effective treatments to cater to specific conditions. Our approach captures the complex interaction of cellular metabolism and handles the high-dimensional data. The research combines comprehensive data from different databases to build a curated dataset of immunomodulatory drugs. Disease models were built using the COMO pipeline, which can integrate transcriptomic, proteomic, and metabolomic data. The pattern recognition method finds the common off-target effects and metabolic pathway changes in the disease and the effects of the drug.

The results identified high-confidence drugs for repurposing, specifically dimethyl fumarate (DMF), a drug approved for multiple sclerosis that can potentially be used to treat rheumatoid arthritis due to the sharing of common targets. Metformin, a diabetes drug that affects lipid metabolism, can also be repurposed for the treatment of systemic lupus erythematosus.

These repurposed drugs can be effective due to the common targets and mechanisms involved. Similar metabolic flux patterns show the method's potential in repurposing the drugs.

In conclusion, the research contributes to accelerating the process of drug discovery by having the potential to find faster and more cost-effective treatment options and to move toward personalized medicine.

Dedication

To the almighty, my mentor, family, and friends.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest appreciation to Dr. Tomas Helikar, my major advisor, for providing me with such a great opportunity and assistance to complete my graduate work. I am grateful to him for his wonderful support throughout my stay at the University of Nebraska-Lincoln, very encouraging words, and intensive perusal to make my thesis presentable.

I am very grateful to all my committee members Dr. Massimiliano Pierobon and Toshihiro Obata for advising and supporting my research efforts. This acknowledgment would not be complete without mentioning my Lab mates. It was a great pleasure working with them and I appreciate their ideas and help during the time of my research work. I also thank the Department of Biochemistry staff for their support and cooperation.

Finally, I must express my very profound gratitude to my parents, my siblings, and my friends for providing me with unfailing support and continuous encouragement throughout my years of study. This accomplishment would not have been possible without them.

Table of contents

Table of Figures

Chapter 1

Introduction

In biotechnology and drug discovery, the constant evolution of computational methods has led to better and faster innovation methods, specifically in drug repurposing (Amiri et al., 2023). The combination of techniques such as natural language processing and classical bioinformatics tools would reshape the pipeline for drug discovery (Imami et al. 2021). This thesis focuses on harnessing the power of pattern recognition and natural language processing along with Genome-Scale Metabolic models to refine and increase the process of drug repurposing with a primary focus on diseases related to the immune system (Cong et al. 2022). Figure 1.1 overviews the entire drug repurposing approach, showing the workflow from diverse data mining to combining and predicting the novel use.

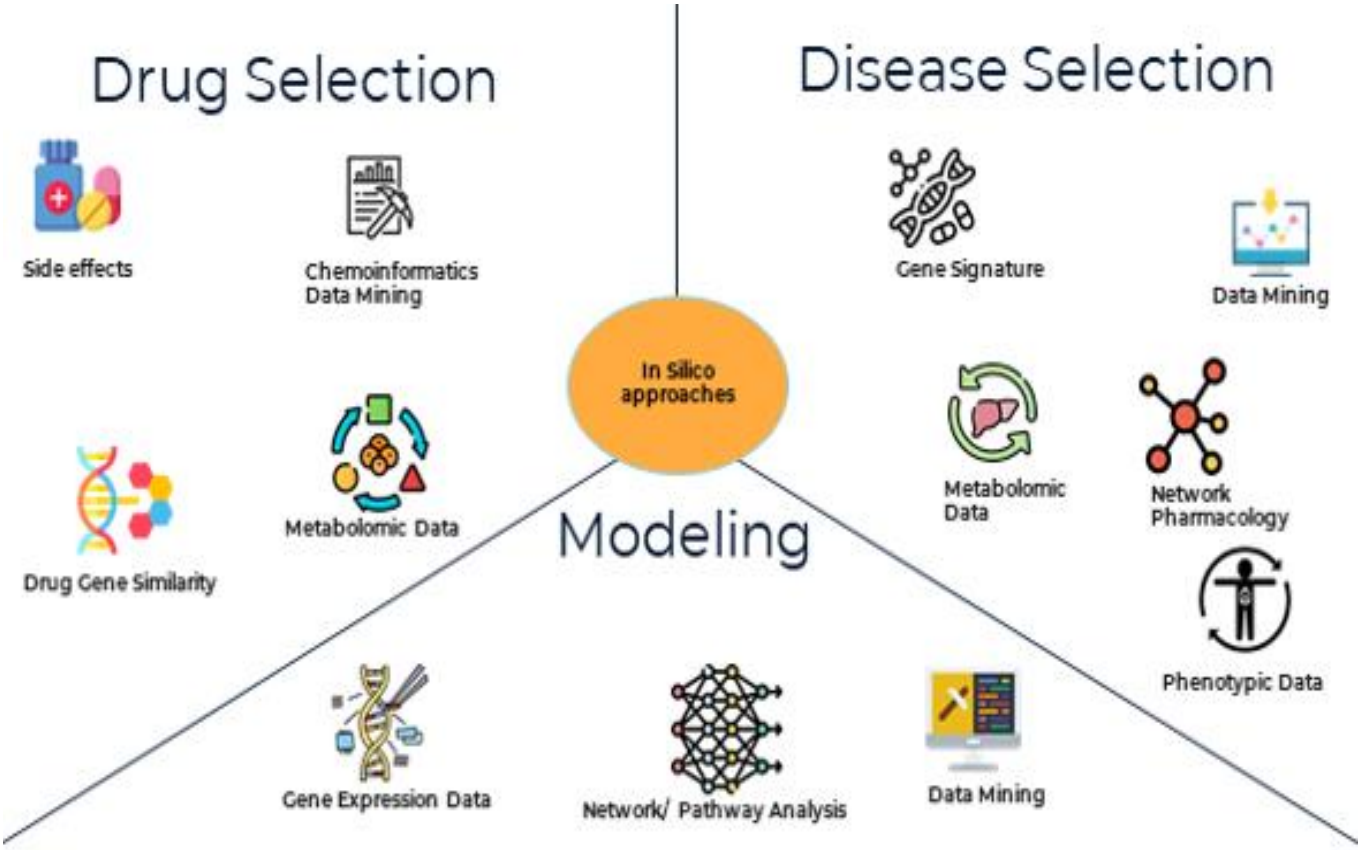

Figure 1.1: Overview of drug Repurposing Approach

The primary importance of this research is to address the pressing issues faced by the pharmaceutical industry (Wang et al. 2018). The traditional drug discovery pipeline is too time-consuming and costly, the process often takes more than a decade and costs multiple billions of dollars just to get one new drug approved to be sold in the market (Juhn & Liu, 2020). Drug repurposing, which includes recognizing the novel uses of the preexisting medicines, gives a highly confident alternative candidate which ends up reducing both cost and the time for the drug development process (Payne et al, 2015). This approach leverages the massive amount of preexisting data on the approved drugs, including all the absorption, distribution, metabolism,

and excretion(ADME) profiles and chemical properties, which leads to the discovery of new treatments.

The primary interest in immunomodulatory drugs is critical, as the number of people being affected by immune-related diseases on a global scale has been growing substantially (Khoo et al. 2014). Autoimmune diseases affect 5-7% of the world's population, and allergies and immunodeficiency impact multiple millions of people more (Holmes et al. 2008). Further, the recent COVID-19 pandemic has also highlighted the role of the subject of immunology in public health, showing the urgent need for effective immunomodulatory therapies (Korbee et al. 2018). By focusing on this domain, our research addresses significant unanswered medical questions, while also helping in the broader field of immunology.

Through this research project, I delved into the application of pattern-based drug selection along with simulating and predicting the effect on the metabolic pathways(Jung et al. 2019). The metabolic pathways are the basic mechanism for understanding the drug effect and the interaction with the human body(Dong et al, 2023), the complex nature of the metabolic networks includes the multiple thousands of interconnected reactions, which helps in making them an ideal form of network for pattern-based search(Wang et al. 2024). Unlike the traditional modeling methods, the pattern based search can effectively handle highly dimensional data and capture every relationship making it well suited to modeling the details of the cellular metabolism.

The final goal of this research is to build a robust pipeline that will be able to identify the potential novelties for the existing drugs with high confidence and efficacity(Dakshanamurthy et al. 2012), By doing so, the goal is to increase the speed of drug discovery and also shed light on

the areas where treatments are scarce(Rahman et al. 2020), This research has the potential to improve the drug discovery by:

1. Hugely reducing the time and cost associated with the selection of the drug candidates.
2. Improving the confidence of the drug candidates being selected.
3. Contributing to the rapid response to the new emerging diseases.
4. Helping in personalized medicine by helping in the selection of drugs based on individual metabolic profiles.

1.1 Background and importance

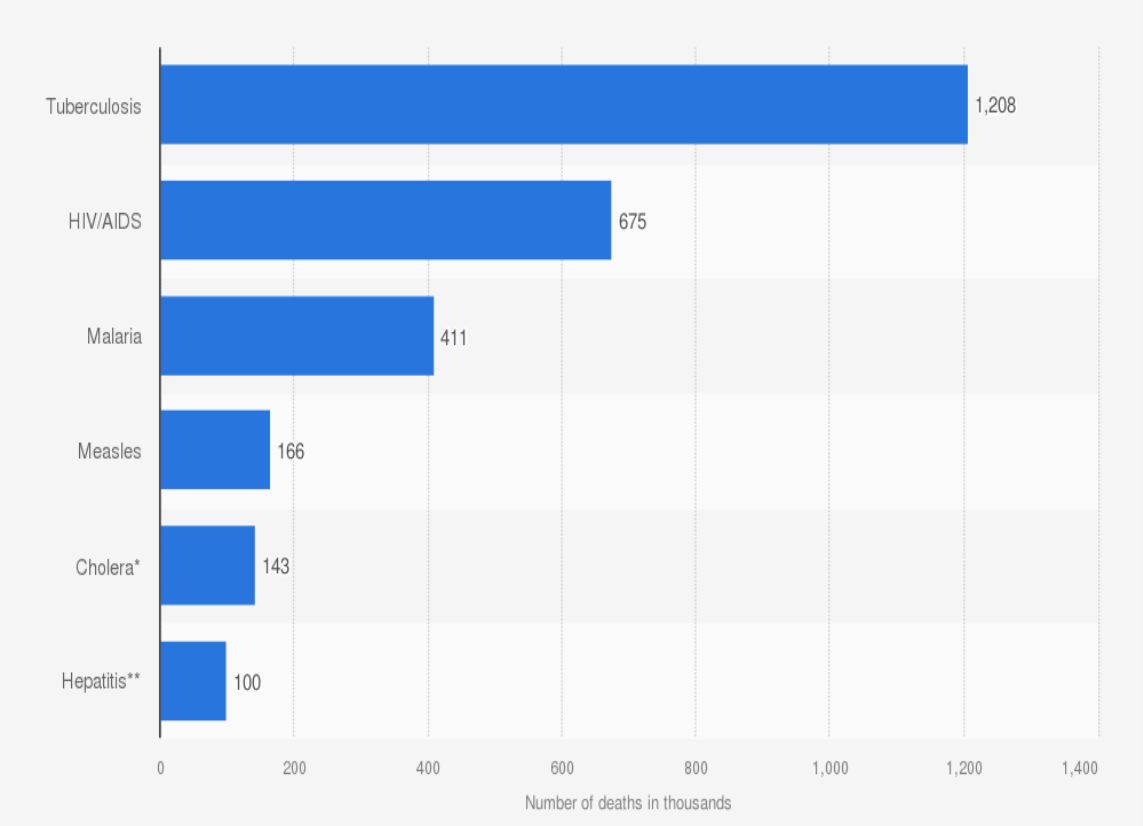

Figure 1.2: Global impact of immune system-related diseases(Deaths Communicable Diseases Annually Worldwide, 2019)

Figure 1.2 shows the widespread cases of people dying due to infection in which the immune system is directly involved, showing the importance of the research. The use of global deaths highlights the significance of the research on innovative approaches in drug repurposing.

The field of drug repurposing, also called drug repositioning, has attracted a significant amount of attention in recent rates as a strategy for the mitigation of the problems and the cost associated with novel drug discovery(Ayala-Ruano et al., 2022), While the idea of drug repurposing is not novel, the fortunate discovers have led to a lot of repurposed drugs throughout the history of medicine(Cichonska et al., 2017), the systems-based approach of the drug repurposing has evolved very fast with advancement in the method and technologies.

Genome-Scale Metabolic Models or the GEMMs have proven to be a powerful tool in the field of network studies(Ekins et al. 2019). This detailed representation of the metabolic network provides a systems-level overview of cellular metabolisms, which allows the researchers to simulate the effects of the environment on the genetic constraints which also include the drug reactions and interactions(McEntire et al. 2016). The Genome-Scale Metabolic Models usually contain thousands of metabolites and reactions, showing the complete functioning of the metabolism of the organism or the cell type. Regarding `drug repurposing, Genome-Scale Metabolic Models can be used to predict how drugs can interact with various metabolic pathways with the potential to reveal the novel effects of the off-target interaction that could also cause side effects.

The detailed and complex intricacies that make the Genome-Scale Metabolic Models also lead to major challenges. The computational power required to simulate these large-scale models

can also be one of the challenges, this is a major concern when we are trying to work with hundreds of potential drug-target combinations for the drug repurposing pipeline. The traditional methods for analyzing Genome-Scale Metabolic Models, such as the flux balance analysis (FBA), while powerful, can be expensive in terms of time and computation and may struggle to consider every single detail of the cellular response to the drug perturbations.

This is where the pattern-based approach presents a novel and promising solution(oh et al. 2007). The pattern-based approach is a subset of Natural Language processing characterized by the ability to recognize the repeating targets and the type of disease it is targeting(Zhang et al. 2019). The approach has several benefits in the context of Genome-Scale Metabolic Models and drug repurposing(Zhang & Hua, 2016):

1 **Effective pattern recognition:** the methods can recognize the repeating metabolic targets within the sub-networks of Genome-Scale Metabolic Models, which show the conserved metabolic functions in several types of diseases or the regulatory mechanisms.

2 **The drug perturbations:** By analyzing the effects of different drugs on the metabolic networks the pattern-based approach can identify the similar effects of the drug interaction which can be used to predict the effect of novel compounds or the proposed drugs,

3 **Disease-specific metabolic patterns:** The pipeline helps in identifying the changes linked to specific diseases leading to the search of the appropriate drug candidates for repurposing.

4 **Rapid discovery:** The pattern-based approach can search the enormous number of potential drug target combinations matching the effects on known metabolic patterns, reducing the dependency on computational time and cost as compared to traditional Genome-Scale Metabolic Models simulations.

5 **Multi-omics data integration:** the method can also include the data from various omics research to consider and include more comprehensive patterns of metabolic pathways and drug interactions.

6 **Pathway level analysis:** Rather than analyzing every reaction individually, the pattern-based method can concentrate on higher-level metabolic pathway patterns giving a more understandable and potentially more robust analysis of drug activity.

7 **Dynamic pattern recognition:** the approach is also capable of capturing the dynamic patterns of the metabolic response over time as it can include the time-dependent omics data which will give a dynamic profile of drug activity on the metabolism.

By using the pattern-based approach the researcher can overcome some of the most concerning computational challenges associated with traditional Genome-Scale Metabolic Models analysis methodology. This can lead to better and scalable drug repurposing research, boosting the identification of novel treatments from existing drugs and improving our understanding of the drug interaction with the metabolism at a complete systems level.

1.2 Statement of Problem

Although there have been major use cases of drug repurposing, there are a considerable number of challenges that restrict the widespread adoption and use of it. The primary issues in the identification of the best drug candidates without conducting exhaustive amounts and expensive clinical trials. The traditional experiment-based methods, which are important for final approval, are slow and expensive when applied to large-scale drug repurposing candidates. This

experiment requires a large amount of time to do both in vitro and in vivo studies, this can take several years to complete and approve to become a consumable substance.

The scope of the problems just magnifies when considering the combinatorial complexity of drug repurposing. With a database of thousands of approved drugs and an even broader database of several potential disease targets, the number of drugs- disease combinations is too large. For example, if we consider 2000 approved drugs and 20,0000 known disease-related genes, we have 40 million potential combinations to go through, this combinatorial boom makes it experimentally impossible to test at all using any conventional methods alone.

In addition to all this, the biological complexity lying underneath the drug activity, and disease mechanisms just adds up to another level of difficulty. Drugs tend to interact with multiple targets and affect a broad spectrum of pathways simultaneously, the phenomenon is called Poly pharmacology. Similarly, diseases, especially complex disorders like autoimmune conditions, specifically involve multiple genes and pathways. The intricated details of interaction make it incredibly challenging to predict the effects of a drug on a specific stage and state of disease accurately.

The Genome-Scale Metabolic Models (GEMMs) give us a more comprehensive approach to address the challenges, but they too have limitations:

1 **Computational cost:** Simulating large-scale Genome-Scale Metabolic Models is computably expensive, restricting the number of conditions that can be explored at a time.

2 **Static Nature:** Most Genome-Scale Metabolic Models show a steady-state view of the metabolism, which might not accurately include the dynamic response to the drug perturbations.

3 **Knowledge Gaps:** Despite all the details, the Genome-Scale Metabolic Models are still based on the available knowledge and may miss the unknown metabolic reaction and regulatory pathways.

4 **Complexity in integrating data type:** While the Genome-Scale Metabolic Models are incredibly good at representing metabolism, integration with other data, for example, gene expression, and protein interaction can be challenging.

The limitations not only restrict the identification of the potential drug-repurposing candidates but also affect the prediction of side effects and the efficacy of unknown cases. The ability to make accurate predictions on both on-target effects and potential off-target interactions is especially important to rank the novel candidates for contributing to patient safety.

Considering these challenges, there is a crucial importance to use sophisticated computation tools that can:

1. Deal with the complexity of systems-level biological data more effectively and with higher prediction accuracy.
2. Include the broad spectrum of data types that can provide a comprehensive understanding of drug-disease interactions.
3. Scalable to include the vast combination of potential drug – and disease interactions.
4. Give interpretable output that can further lead the experimental and clinical decisions.
5. Considering individual differences supports the personalized treatment approach.

This is where the pattern-based approaches give a novel and very promising solution. The pattern-based approach's ability to recognize the target repetition and types of the disease being

targeted gives several advantages for giving solutions to the challenges like with the Genome-Scale Metabolic Models and drug repurposing:

1. **Effective pattern recognition:** The method can recognize the metabolic pattern or subnetwork in The Genome-Scale Metabolic Models that can lead to potential recognition of metabolic function or regulatory mechanisms in different conditions or cell types.
2. **Drug response expression:** By analyzing all the known effects of the drugs on metabolic networks the pattern-able approach can recognize and classify the expression of drug response which can be used to predict the novel or repurposed drug effects and interactions.
3. **Rapid discovery pipeline:** The pattern-based approach can quickly screen the large database of all drug-target combinations by matching the drug effects to know metabolic patterns, hugely reducing the computation expense to full Genome-Scale Metabolic model simulations.
4. **Pathways level analysis:** Rather than focusing on individual reactions, the pattern-based approach can focus on the patterns, giving a more understandable potential better analysis of drug effects.

By using the pattern-based approach, the researchers can address the challenges linked to the traditional Genome-Scale Metabolic Models analysis approach(Chai et al. 2020). This can help to accelerate the process of an efficient scalable drug repurposing approach.

Addressing the challenge through the pattern-based approach is not just an academic need but also critical for drug discovery and development(Amiri et al. 2023). The use case of a more effective drug repurposing pipeline can drastically reduce the time and expense of having new

treatment for patients, specifically catering to rare diseases or in response to new health threats where time is extremely critical.

1.3 Purpose of Study

The overarching reason for this research is to validate a pattern-based pipeline that can serve as a surrogate for the traditional Genome-Scale Metabolic Models (GEMMs), thus leading to accurate prediction and rapid drug repurposing(Imami et al. 2021). The goal of the research is to address the problem in the problem statement subsection by leveraging the power of pattern recognition to create a robust, scalable, and accurate method for drug repurposing and metabolic model integration.

The specific objectives of the research are:

1. Build the pattern-based search pipeline that considers the metabolic pathways involved in the human immune system:
    a. Recognized the repeating metabolic pattern targets or subnetwork targets within Genome-Scale Metabolic Models (GEMMs) that are related to affecting the immune cell activity.
    b. Create a pipeline for pattern classification.
    c. Model validation, the ability to predict and compare the target effects in different conditions.
2. Use the models for prediction of the activity of approved drugs in the metabolic pathways showcasing the novel use cases of pre-existing drugs:

   a. Use the method for simulation of the drug perturbation within the pattern-based pipeline.
   b. Screen the self-curated large database of approved drugs against the metabolic model pathway to identify the potential immunomodulatory effects.
   c. Prioritize the drug candidates based on their impact on immune cells and the role of cells in diseases.
3. Comparing the results of the pattern-based approach against the traditional Genome-Scale Metabolic Models (GEMMs) in terms of targets affected:
   a. Compare the number of metabolites involved in traditional reaction-based flux change vs pattern-based metabolites affected.
   b. Compare the computation power and time required for approaches in different cases.
4. Combining multiple data types to improve the capability to predict:
   a. Build scripts to add different data sources, including gene expression data, proteomics, and clinal information into the pattern-based pipeline.
   b. Analyze the use of data added to increase the confidence of the prediction for drug repurposing.
5. Finding the explanation of the model predictions:
   a. Apply visualization methods that can help researchers understand the predicted targets based on the patterns.
6. Validate the predictions:
   a. Selection of the drug candidate simulations that have the highest confidence for experimental validation.

   b. Do a cell-based assay to check the immunomodulatory effects of these predicted candidates.

7. Moving towards personalized predictions:
   a. Create a pipeline for adapting the pattern-based search approach to include the individuals' differences in metabolism.
   b. Check the practicality of using patient-specific information to filter the drug prediction based on personal metabolic patterns.

By tackling these points, we aim to build a comprehensive method that not only improves the confidence of the predictions regarding efficacy and safety but also enhances the drug repurposing approach. This pattern-based approach is supposed to significantly universalize the drug repurposing process, which reduces the expense and the time linked with the traditional drug discovery methods while retaining or improving the confidence or accuracy of the predictions.

1.4 Significance of the Study

The significance of the pattern-based approach to drug repurposing is extremely far-reaching, which spreads beyond the drug repurposing approach to impact various parts of biomedical research, healthcare, and drug discovery(Jung et al. 2019). By improving the pipeline e to repurpose drugs rapidly and with high confidence through pattern recognition, the research not only contributes to academic research but also has the potential to significantly affect public health and the cost of drug development.

Major areas of significance include:

1. Acceleration of drug discovery and development:
    a. The time required for the drugs to come to the markets will be highly reduced.
    b. It will reduce the development costs through effective filtering.
    c. The success rate of the drugs will increase with the high-confidence metabolic patterns.
2. It will contribute to solving the public health crisis:
    a. It will help in the rapid selection of pre-existing drugs for emerging threats.
    b. It will lead the modeling approach as new data is found.
    c. It will be helpful for combinatorial drug treatments through pattern-based analysis.
3. Shift towards personalized medicine:
    a. It will help in having drugs s
    b. **There are no sources in the current document.**pecific to the patient based on their metabolic patterns.
    c. It will help in precision drug approach thus reducing the adverse effects.
    d. It will help in the optimization of the treatment plants for rare and complex diseases.
4. The impact on the Economic aspect of healthcare:
    a. It will contribute to the reduction of healthcare costs through the effect drug development pipeline.
    b. It will be helpful for rare disease research by lowering the barrier for drug repurposing.

   c. It will lead to increased market competition and thus lower drug prices.

5. It will improve the field of Computational Biology and Bioinformatics:
   a. It provides a novel pattern-based modeling approach.
   b. It improves the methodology to integrate the different biological data types.
   c. It will improve the interpretation of the different models.
6. It will lead to a Paradigm shift in Drug Discovery:
   a. It will shift the research and development interest towards the computation and pattern-based approach.
   b. It will lead to different collaboration opportunities in academia and industry.
   c. It will address the concern for computational drug discovery.

The focus on the immune systems-related pathways in the research is very strategic and important for multiple reasons(Ekins et al. 2019). There has been a global increase in the number of people suffering from autoimmune diseases, allergies, and immune system-based problems, and this creates a pressing need for new drugs or approaches(Zeng et al. 2015). The very delicate nature of the immune system leads to a few challenges in drug development, which makes it ideal research for us for pattern-based analysis.

In addition, every stride made in immunology has a very far-reaching effect in the fields of medicine, cancer care, and infectious disease. The recent COVID crisis has highlighted the demand for rapid drug repurposing for immune system-related conditions, showing the practical application of this research(McEntire et al. 2016).

Chapter 2

Literature Review

2.1 Drug Repurposing in Immunomodulation

The field of immune system-modulating drug repurposing has witnessed some major developments in recent years; with pattern-based approaches, they will prove to be powerful tools for finding new drugs for treatments.

Recent studies have explored various computational approaches to find the immunomodulatory effects of pre-existing drugs, but the combination of pattern-based approaches in the Genome-Scale Metabolic Models (GEMMs) is still an untouched area that has tremendous potential.

The research field of immunomodulating drug repurposing has had multiple significant advancements in the past years. (Srimadh et al. 2022) built a network based approach to repurposing drugs for patients having rheumatoid arthritis, combing the gene expression data with protein interaction network information. The approach they used was successful in identifying multiple candidates which included the drugs with known knowledge to affect the

immune system, this approach did not specify the potential work with the immune checkpoint inhibitors particular immune cells being targeted or the metabolic pathways information, having those considerations would have given additional potential candidates that might specifically affect the drug mechanism and would have helped in understanding the off-target effects.

In the domain of cancer immune therapy (Chu et al., 2023) applied a machine-learning approach to predict the repurposing of the approved drugs to improve the immune checkpoint blockade treatment. By combining the multi-omics data with drug target data, they were again able to select the best candidates that could work with immune check potential inhibitors. Although this research highlighted the power of machine learning in the field of drug repurposing, it did not include any metabolic aspect of the immune cells' function, the information related to it is crucial as the immunomodulating effect of the drug and the side effects cannot be predicted without it.

The drug repurposing approach also has been used for combinatorial drugs to modulate the immune system. (Pan et al. 2023) published a systematic method to find the synergistic combination of drugs for autoimmune diseases, they leveraged the transcriptomic data and the known drug-target interactions. The method showed the potential of combining drugs with different mechanisms to achieve more effective modulation of the immune system, Although the approach was highly appreciated by the community, due to the lack of the systems level application, they could not justify the mechanistic understanding of the combinatorial treatments.

The recent development in single-cell technology has opened more opportunities for the field of immune-modulating drug repurposing, (Yang et al. 2022)made an approach to predict the cell type-specific drug effects using the single-cell Rna seq data, this helped in targeting precisely the specific immune cell types. This approach was successful in identifying the drugs

that can affect the type and stage of the immune cell without affecting the others. However, the integration of the single-cell data with the Genome-Scale Metabolic models in conjunction with single-cell data would have the potential to significantly enhance our knowledge to under and predict the drug effects on immune cell metabolism and function.

Parallelly using the phenotypic data for drug repurposing to module the immune system also has drawn attention. (Barsi et al. 2022) analyzed a very large dataset of drug-treated gene expression data to find the compounds with immunomodulatory effects, The approach highlighted multiple drugs with unexpected effects on the immune cell pathways, this showed the benefit of using unbiased screening, also showing the importance of the Genome-Scale Metabolic models to improve our understanding for better drug effect prediction on immune cell metabolism and its functions.

A new application for using artificial intelligence methods to repurpose drugs in immunomodulation has also been made which was quite successful. (Shen et al. 2023)built a deep learning model combining the structural information of the drug targets with the gene expression to predict the effect of the small compounds on the immune system. Although the method was very good it had the limitation of not having metabolic pathway information in the consideration and the common patterns of the drug effects.

With the expansion of the field, the focus also shifted to the repurposing of biologics as well as to modulating the immune system. (Rapin et al. 2010) developed a computational framework to predict the new target for the approved monoclonal antibodies based on the epitope similarities. The approach was able to identify the potential targets in the different disease conditions, it also highlights the need to address the effects of the metabolism of the immune

cells and the function which can be done using the pattern-based approach to find the common mechanism of action across all the different antibodies.

Considering the impact of environmental conditions on the immune system, (Azer & Leaf 2023) developed systems biology-based pipeline combining the data from environmental exposure and genetic changes, and drug-target interactions. This specific approach target to find the right candidates for environment-dependent autoimmune diseases showing that external factors play a significant role in the diseases revealing the new opportunity of combining the environmental data with metabolic modes and pattern-based search approaches to provide detailed insight into gene-environment interaction and the effect on the drug efficacy.

In conclusion, the drug repurposing domain has made many significant achievements, using different computational methods to find novel therapeutic candidates. However, the combination of a pattern-based approach with the Genome-Scale Metabolic Models (GEMMs) still has not been done which has a tremendous amount of potential. By catering to the gap, the accuracy and efficiency of the drug repurposing research specific to the immune system-related disorder would lead to more high-confidence personalized therapy candidates that can effectively modulate the immune system.

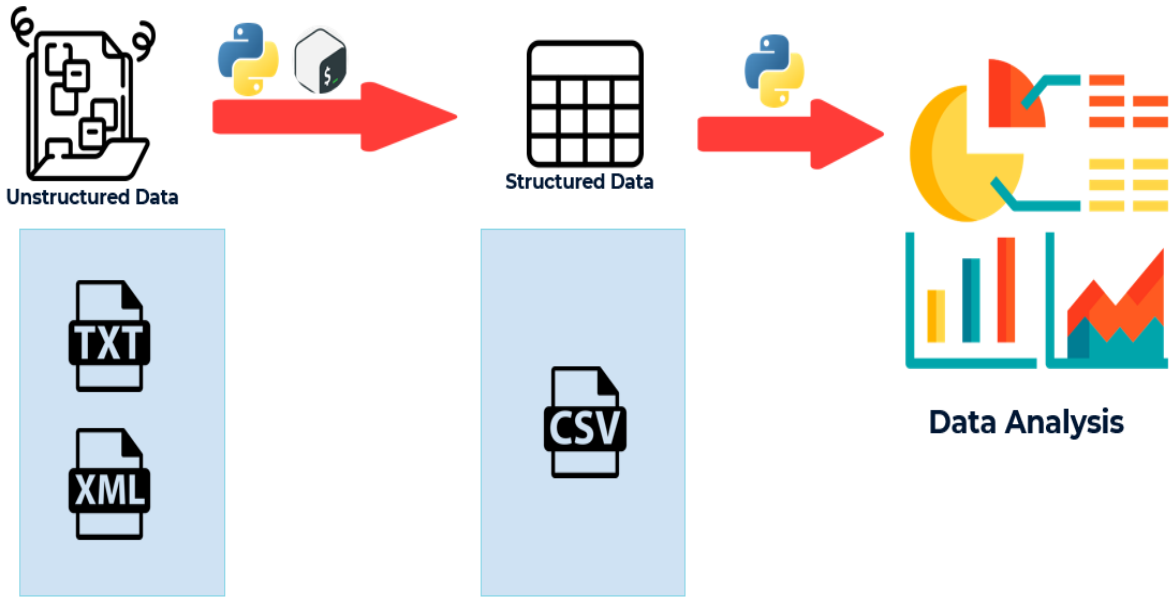

Figure 2.1: Data-driven approaches in drug repurposing.

Figure 2.1 shows the workflow of the data-driven approach to repurpose the drugs. It starts with mining the unstructured data in the form of TXT and XML files. The data is then processed and converted into a structured format, which is A comma-separated value for my research using Python and shell scripting. Once the data is structured, then it can go to the next step for analysis using Python to visualize and get insights that help in understanding g the data; figure 2.1 shows the steps involved in big data handling and analysis, showing the need for data structuring before any step further.

2.2 Big Data and Healthcare

The use of big data in the field of healthcare has revolutionized the entire domain of drug repurposing, which gives us a plethora of opportunities to leverage the broad spectrum of data types for finding the novel potential of existing and approved drugs. The data-driven approach has expanded rapidly, with each step building on previous knowledge and unveiling new aspects for exploration. However, as the field is growing it is very distinct that the complete capability of big data for drug repurposing is in the combination of various data sources and data types and then analyzing them.

The relevance of big data in drug repurposing started with the realization that a significant amount of knowledge can be collected from the data generated in clinical trials. (Kort & Jovinge 2021) were the pioneers of the method by creating a network-based pipeline that would combine the data from electronic health records (EHRs), genetic associations, and drug-target interactions. The work that they did shows the practical use case of combining diverse data sources in drug-disease relations, setting the stepping stone for more integrative approaches. While also that it was the first step it also highlights the need for more detailed analysis methods that could capture the intricacy of the biological systems, specifically at the metabolic level.

With the previous knowledge foundation, the researchers started to look for more applications. (Rossi & Grifantini 2018)leveraged the large data set of patient data from electronic health records to find the off-label drug use and to find the new repurposed candidates. This research highlights the value of mining the clinical data to help with drug repurposing research and lead the path for a more personalized approach to drug discovery, it also highlights the

limitation of completely depending on clinical data, showing the researchers to find a way to integrate molecular level data for a more comprehensive understanding of drug interaction.

As the field evolved, the researchers also became interested in using the power of literature and clinical reports. (Xu et al. 2019) built a natural language processing package to automate the process of data extraction of drug-disease association from PubMed abstracts, and finding the multiple candidates that can be focused on. The research showed the potential of the text mining approach for applying it to the ocean of literature to get combined information thus helping in finding the gaps between the text-based hypotheses and molecular mechanisms.

The integration of multi-dimensional data was the next big step in the field. (Nicora et al. 2020) worked on a deep learning method that could combine molecular and imaging data to predict drug interaction in cancer. This reach showed the ability to use a combination of different data types for drug repurposing and made the way for an even more elaborate data integration approach. This also showed the limitation of just considering one scenario where a pattern-based approach could give the candidates interaction across multiple cases.

Parallelly the integration of the multi-omics data also became a powerful approach to get a comprehending understanding of the drug interactions. (Godfrey & Kornberg 2020) made a multi-omics integration approach that combined the transcriptomic, proteomics data, and metabolomics data for the prediction of drug interactions and to figure out the opportunities to repurpose drugs. This underscored the importance of diverse data integration.

Looking forward, the next step in the direction of big data in drug repurposing is in the development of approaches where there is more integration of diverse methodologies and data that can provide biologically relevant insights. The Integration of the pattern-based approach

with metabolic modeling gives us an edge in leading the field. By finding the metabolic patterns across various diseases researchers can understand the mechanisms effectively and find candidates for potential repurposing.

In conclusion, the field of big data in healthcare and repurposing has been an ever-growing domain with constant evolution and integration. Every single step and data have been dwelling upon the previous achievements, unveiling new insights and opportunities parallelly highlighting the areas of further research. As the field moves ahead, the focus will shift towards the more holistic and multi-data type approach that can completely use the power of big data to help drug repurposing and, thus improve patient health by revealing the novel therapeutic application and optimizing the treatment strategies based on the detailed data-driven insights.

2.3 Previous Work in Genome-Scale Metabolic Models (GEMMs)

Genome-scale metabolic Models (GEMMs) have proven to be powerful tools for understanding cellular metabolism and predicting the effects of genetic and environmental perturbations, including drug interactions. The growth of Genome-Scale Metabolic Models (GEMMs) in the past years has been marked by significant advancements in terms of accuracy and the extent of application, especially in drug discovery and repurposing. However, as we dive deeper into the ocean of complex cellular metabolism, it is very evident that the integration of a

pattern-based approach with Genome-Scale Metabolic Models (GEMMs) especially when trying to understand immunomodulation, represents a frontier with a lot of potential for innovation.

The use of Genome-scale metabolic Models (GEMMs) in the field of drug repurposing started with building a comprehensive map of human metabolism, the most recent community-driven model is the Recon3D model developed by (Kim et al. 2021). This model includes metabolomics and proteomics data which give a more nuanced representation of human metabolism and has been very crucial in the prediction of drug off-target effects. Yet, the ever-growing field gives us a new opportunity to find ways to identify novel drug targets and predict drug effects in cellular metabolism through new analysis methods and data types.

With the field progressing, researchers have worked on exploring the integration of machine learning techniques with Genome-scale metabolic Models (GEMMs). (Zhang et al. 2019) combined flux balance analysis with machine learning to predict the effect of antibiotics on Genome-scale metabolic Models (GEMMs). This work showed the potential of combining the machine learning technique and genome-scale metabolic model for drug discovery and repurposing, while also highlighting the need for adding expression of metabolites specifically in the case of drugs that modulate the immune system.

Building context-specific metabolic models was also one of the most significant steps in the field. (Lopez-Agudelo et al. 2020) pioneered the creation of personalized genome-scale metabolic models for hepatocellular carcinoma patients, they showed the potential of personalized metabolic modeling in identifying the drug candidates and targets. This research was groundbreaking and showed the potential of broad-scale application of a pattern-based approach to enhancing the ability to predict an individual's response to drugs that modulate the immune system.

The researchers have also applied the genome-scale metabolic models with respect to host-pathogen interactions, which also opens a completely new field of drug repurposing in infectious diseases. (Sen & Oresic 2019) published research on developing a host-pathogen genome-scale metabolic model for mycobacterium tuberculosis infection, demonstrating the importance of a genome scale metabolic model in understanding the complex interactions in the biological systems. This research also focused on the need for incorporating metabolic signatures across various kinds of infections, which would potentially lead to the discovery of a wide range of treatments.

(Chandrasekaran & Price 2010) developed AGORA. Collection of genome-scale metabolic models for human gut microbes, which helps us in understanding the microbiome drug interactions. This development in the field of microbial genome-scale metabolic models pointed towards the potential of novel analysis approaches in finding common metabolic patterns which are linked with human-microbe interaction, having the potential to revolutionize our approach to modulating the microbiome for better health and disease treatment.

The regulatory information in the genome scale metabolic model as shown by (Turanli et al. 2018), is the ability to predict the metabolic response to various kinds of perturbations. This research opened a completely new possibility for predicting the drug effect on the cellular metabolism which showed the potential of regulatory metabolic models to understand the interactions between drug effects gene regulation and metabolic responses.

Tissue specific genome scale metabolic models by (Sen & Oresic 2019) gives us a valuable foundation for studying the trick effects across different tissues. They prepared an Atlas of tissue specific metabolic models based on the proteomic data set that show us the common

metabolic signatures across multiple tissues and identified tissue specific potential targets for drug repurposing.

The use of genome-scale metabolic models in the field of Cancer Research has also given us novel insights into the potential therapeutic strategies. (Zangene et al. 2023) pipeline for identifying the synthetic lethal gene pairs using the cancer metabolic model, showing the potential of combinatorial treatments. The genome-scale metabolic models were again proven to be a powerful tool for finding therapeutic targets and understanding the effects of drug combinations.

(Christ et al. 2021) Integrated the genome-scale metabolic models with the pharmacokinetics models, which showed the application of integration to predict the drug distribution, efficacy, and toxicity across all the tissues and biological scales.

Integration of genome-scale metabolic models with the multiple omics data types serves as a crucial point of focus in the field, it offers us an opportunity to understand the drug effects on cellular physiology in a more comprehensive way and helps in the identification of novel biomarkers targets for drug response. While the studies have begun to explore the integration the large-scale potential of applying different methods to multi-omics data in the context of a genome-scale metabolic model is still not completely explored. As we conduct more research in drug discovery and genome-scale metabolic models, we will be able to overcome current limitations.

Chapter 3

Methods

The research method combines data collection, model building, and pattern-based analysts to build a robust drug repurposing pipeline for immune system-related diseases; at first, the data on the immunome modulation drugs is mined from multiple databases, which is curated. This data is then combined with genome-scale metabolic models, which are cell type-specific in terms of immune cells. The core of the research is to find patterns in the metabolic changes caused by drugs. Then we compare the patterns to other dugs metabolic changes. When a drug has a similar pattern to known drugs for a specific immune system-related disease, it can be a potential drug for repurposing. By leveraging the big data, modeling, and pattern search method for identifying novel applications of existing drugs, this method offers an efficient way to identify new applications of drugs(Gu et al., 2019).

The comprehensive research on the effect of immunomodulatory drugs on the metabolism using genome-scale metabolic models follows a very meticulous design which is a multi-stage process(Wishart et al. 2018, Barret et l. 2012g). This methodology seamlessly integrates data collection, data classification, model development, and data analysis.

The foundation of this Research was built on a very carefully self-curated data set of your business modulatory drugs and their associated chemical and biological information. We developed a systemic approach to gather and organize the information leveraging different databases including drug bank PubChem Kegg David Panther DB and Reactome custom Python scripts were built to query these databases efficiently and extract the crucial information such as target mechanism of actions associated with metabolic pathways primary disease etcetera. To capture the most recent findings and information I used an automated literature mining pipeline. This research employed methodologies such as natural language processing techniques to scan the publication abstract focused on immunology and pharmacology ensuring that the data set is up-to-date and comprehensive.

Given the collected data was heterogeneous I made a standardization pipeline to ensure the consistency of data. This involved mapping all the drugs to the target names' answer types using standard ID converting the sentence format data to single-word Boolean values and building a controlled Dictionary of words for describing disease, drug mechanisms, and effects. To ensure the highest standard of data quality, I used a multi-step filtering process, including automated filtering for data completeness and consistency, manual curation by immunologists, and mapping information from multiple sources. The final curated data was integrated to build a relational database, which can be used to query and analyze multiple phases of the study.

With this large data set as our base, we moved forward to develop and refine our modeling process, modeling and mapping the specific drugs targeting the immune cells. We selected the Recon 3D human metabolic model as a base template because of its comprehensive information and manual curation of every node, we traced out the metabolic model of specific immune cells from this model. The models were created using the constrait based optimization of

metabolic objectives Como pipeline(Bessell et al., 2023) which combined this cell type-specific gene expression data with the metabolic network to generate a context-specific metabolic model using the Cobra toolbox. The model building also included the manual curation of metabolic pathways specific to immune cells.

The research is focused on the identification of repeated patterns in drug-induced metabolic changes. I have a multi-step process to extract and analyze these patterns effectively. For every drug in the data set, I compare it to a healthy immune cell, and then I compare it with the diseased model and note the changes in the metabolic flux levels.

To find the recurring patterns of drug-induced changes in metabolism, I noted the average flux across all the flux balance analysis simulations, which would give a sense of the normal activity of the reactions in the condition of all approved drugs. Then, I tried to visualize the patterns to find the common patterns showing the different highlights of the metabolic pathway.

3.1 Data Collection

I had a multifaceted and comprehensive process for the data collection, it was designed to gather a broad range of information for drugs that modulate the immune system their targets, and the metabolic pathways. The approach was very thorough and systemic ensuring inclusion of Every single relevant data point available from various sources to create a strong base for the research.

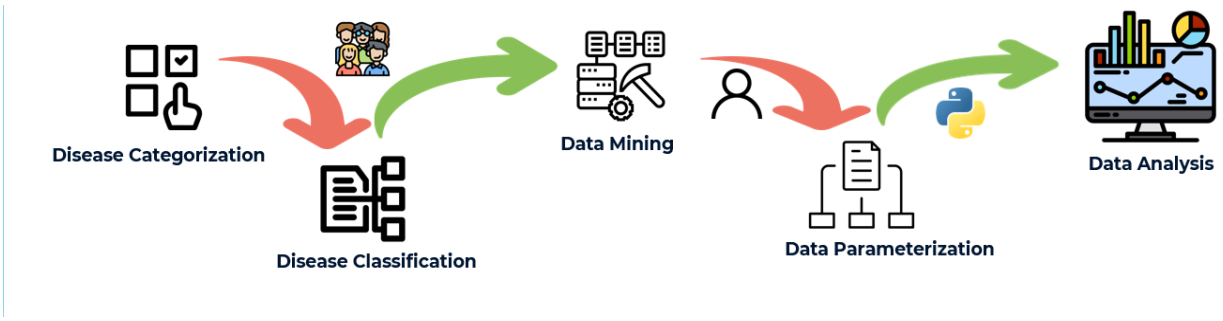

Figure 3.1: Methodology for our drug repurposing approach

Figure 3.1 shows the workflow for the analysis of the disease data, showing the disease categorization where they are grouped together for analysis. Then, the disease classification, which focuses on the subgroups of diseases, then I focused on data mining, which has multiple parameters where everything was manually curated. Finally, the analysis step for the understanding and the visualization for better insights.

I collected data for drugs in experimental, approved investigational, and other stages or substages. I also collected disease-specific information such as the drugs that are used to treat certain diseases with a particular focus on immune system-related pathologies. To ensure the relevance and application of our findings I focused on FDA-approved drugs and phase 1-3 clinical trials drugs. This focused research helped us in characterizing the compounds that have established safety profiles which would help in improving the translation of our research.

The process includes multiple intertwined steps:

1. **Identifying the drugs that modulate the immune system:** this started with compiling the list of all the drugs and then classifying whether they modulate the immune system. This included small molecule drugs and biologics which covered up broad spectrum of

therapeutic classes such as corticosteroids, cytokine modulators, monoclonal antibodies etcetera.

2. **Combining drug targets and mechanism of action:** for every single drug he collected detailed information on its molecular targets and known mechanism of action this included the primary targets and any known interaction with metabolic pathways. Collection of interactions with metabolic pathways I added the information on metabolic pathways affected by or involved in the mechanism of drug interaction. This included the effects on metabolic enzymes and the indirect relation with the signaling pathways that affect the metabolism.
3. **Collecting specific disease data:** I classified and analyzed the data to find the disease conditions for which the drugs have been approved or are in the investigation stage. This step was based on the previous data literature and current treatment strategies.
4. **Collection of relevant clinical and experimental data:** I mined the open source data which included clinical data and experimental data such as transcriptomic data to get information on drug efficacy side effects and dose information.

The data collection method was not just aggregating the existing information, I also built scripts for natural language processing and text Mining to extract the relevant information from the literature abstracts.

The steps in the multi-tiered approach were:

1. Building scripts for checking the data consistency and identifying potential gaps or errors.
2. Manual curation of data by an immunologist to check any false annotations.
3. Cross-reference to multiple sources to ensure the highest confidence in data annotation.

4. I also gathered the biological process data that are relevant to immune cell functioning and drug action.
5. Transcriptomic data for immune cells in different stages.
6. Transcriptomic data for disease of interest.
7. Biological ontologies for the known drug targets.

Quantitative data	**Qualitative data**
Number of drugs: 19,394	Drug names
Dosage information	Mechanism of action
Flux values of pathways	Known targets
Flux balance analysis	Class of disease
Transcriptomic data	Drug stages
Cell/ tissue-specific data	Enzymes
Patient condition-specific data	Biological process
	Kegg Pathway information
	Literature information
	Gene ontology
	Metabolic and signaling targets classification

Table 1: Shows the difference between qualitative and quantitative data collection

Table 1: Shows the difference in the quantitative and the quantitative data collection for the research. The combination of the diverse data is the foundation of the research upon which the pattern based approach is built.

This comprehensive method for data collection helped us to build a multidimensional dataset that would not only give us the drugs that would have similar predicted effects but also show us the other family of proteins that can be potentially targeted. This is important to focus on alternative targets and study off-target effects.

As our data collection progressed, I saw trends in drugs that would help us in simulating the drugs accordingly. For example, I found drugs with similar mechanisms of action but are used for treating different conditions, which gives the potential opportunity to test in a model. I also found gaps in current knowledge, specifically in the context of the effect on the metabolism of certain classes of drugs.

3.1.1 Databases Used

For building a comprehensive and reliable dataset, I used multiple established databases and resources to collect data. Every database was selected for its unique data and references that would help fill the gaps useful for our research and help us better understand immunomodulatory drugs and their effects on the metabolism.

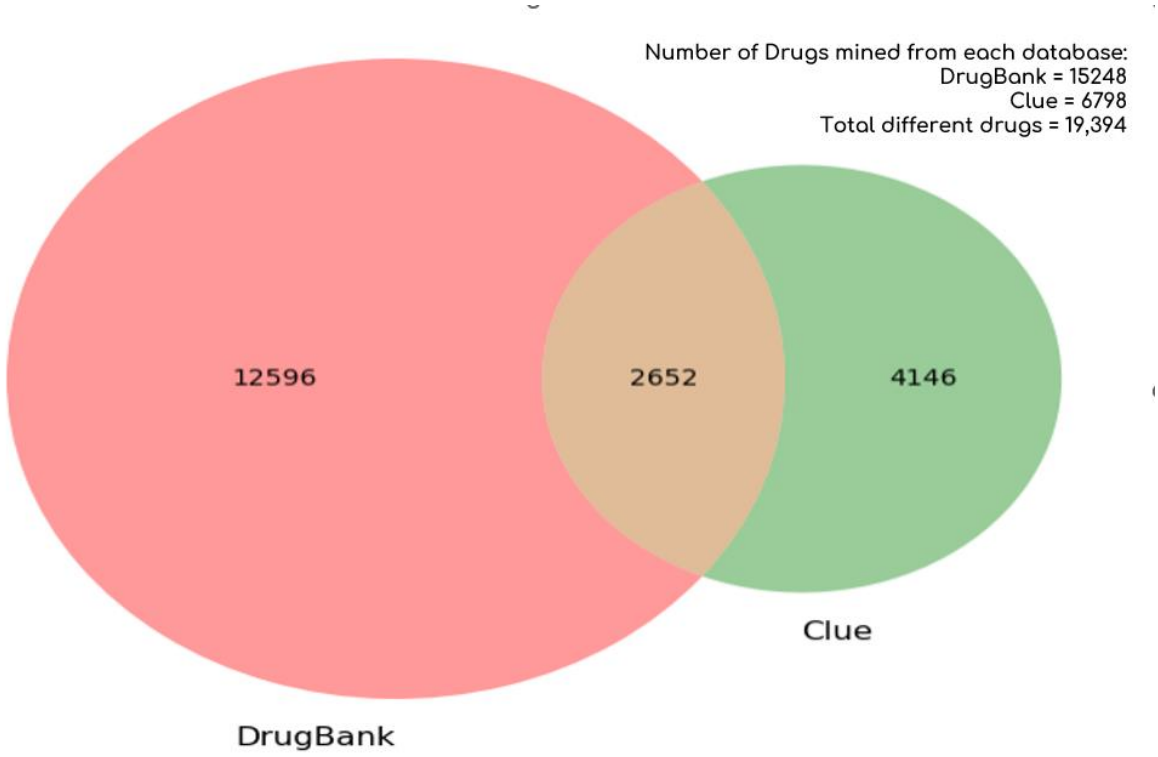

Figure 3.2: Distribution of drugs in our dataset from Drug Bank and CLUE databases

Figure 3.2 shows the use of two databases, Drug Bank and Clue, containing 15,248 and 6,798 drugs, respectively. Only 2,652 drugs are common, and the rest are unique. In total, 19,394 different drugs were mined from the databases.

The databases used in the research are:

1. **Drug Bank:** This open-access database is the primary database for drug information(Wishart et al., 2018). This database gives us detailed information about drug structure, mechanism of action, known targets, and pharmacokinetics. We extracted all the drug information and then focused on the immunomodulatory drugs, molecular targets, and the linked metabolic pathway. The extensive database structure needed a bash

and python script to run on a high-performance computing cluster to extract and combine with our database.

2. **PubChem:** The PubChem database was used to get more information on the properties of the immunomodulatory drugs(Wang et al., 2009). PubChem's extensive collection of compound data helped us add structural information if it was missing from our database. I used the additional information to search for drug target interaction and biological activities.
3. **CHEMBL:** This bioactivity database gives information on the drugs and their target(Gaulton et al., 2016). Chembl was used to check the drug-target interaction that led to the modulation of the immune system. The database helps in having another reference point to establish the drug target relation, which is crucial for predicting the on-target and off-target effects, giving us more confidence about the information related to targets.
4. **KEGG:** KEGG was immensely helpful in finding the metabolic pathways and mapping them to the respective drugs(Aoki-Kinoshita & Kanehisa, 2007). This helped us have comprehensive pathway information that would give us a broader understanding of the drug's actions within the cell metabolism.
5. **Reactome:** This peer-reviewed database helps us get total information on biological processes and pathways(Fabregat et al. 2015), which word was crucial in classifying the trucks, whether they were metabolism or signaling pathways targeting drugs. Pathways are also categorized based on various biological organizations, which helps us understand the molecular interaction and cellular processes.
6. **Gene Ontology (GO)Gene ontology database** helped us annotate the genes affected by the drugs(Masseroli & Pinciroli, 2006). This gives us a dictionary of genes and their

functions, their biological processes involved, and their cellular components. This database helps us contextualize the drug targets concerning the cellular function, which gives us the potential side effects of the drugs on the immune cell metabolism.

7. **FDA Database(Food and Drug Administration):** I used the FDA database to confirm the list of drugs i have are approved drugs and to collect information on the approved disease indication and if any side effects are associated. This database gave the information on approval; dates and label changes and market data, ensuring the safety profile of the drugs for our research.
8. **Human Protein Atlas:** This database gives information about the protein expression in different tissues and cell types. This gave us information to understand the immune cell-specific effect of the immune system modulating drugs and to find other potential targets and off-target effects based on the expression of the targets in different cells.
9. **String (Search Tool for the Retrieval of Interacting Genes/Proteins):** I use the string database to check the protein-protein interaction that is linked to the drug targets. This network of proteins would help us understand the broader effect of the immunomodulatory drugs, helping us to find protein other targets that would have a pathway-level effect.
10. **Uniport:** This is a comprehensive database for proteins and their functional information that would help more to understand the function and the interaction of the proteins that are affected by the immunomodulatory drugs.
11. **Bio GRID (Biological General Repository for Interaction Datasets):** Using this database i gather the data on the protein and gene interaction(Oughtred et al. 2018). This

would help in elucidating the complex interaction between the drug targets and the cellular components.

12. **CLUE (Connectivity Map Linked User Environment): Environment):** I used the CLUE database to explore the gene expression values linked to the drug treatments. This database gave information about the effect of the Immunomodulatory drugs, which would help in understanding the novel mechanism of action.
13. **DAVID (Database for Annotation, Visualization, and Integrated Discovery):** It was used for assessing the functional annotation and doing the analysis of the genes linked to the drug targets that affect the metabolic pathway(Huang et al. 2008). This comprehensive list gave us the role of genes in the drug modulation of the immune system.
14. **Panther dB (Protein Analysis Through Evolutionary Relationships):** Panther dB was used for gene ontology analysis and evolutionary analysis of the protein family. The database was used for classifying proteins and their respective genes to understand the evolutionary relation of drug targets.
15. **GEO(Gene Expression Omnibus):** The database was used, as it I the largest public resource for getting the transcriptomic data(Barrett et al. 2012) which would help in getting the relevant gene expression data that would be useful in understanding the immunomodulatory effect of the drugs, disease, and, immune cell gene expression information under different conditions.

To combine all the information from these resources, i built scripts and pipelines that would run on a high-performance computing cluster for processing the data. These scripts helped us to

extract the relevant information, classify the data, standardize the data format, and then combine all data from different sources into one database built for our research.

The data integration process had several important steps:

1. **Data extraction:** i developed application programming interfaces, web and text scraping Python and shell scripts for fast relevant data extraction from the web and different formats. These scripts were made to handle the large XML, txt, web files, and CSV files effectively.
2. **Data cleaning and standardization**: The raw data from all the sources was filtered and standardized to the consistency. This process involved the addition of a Boolean value to whether certain characteristics such as a data set have healthy samples or not and handling the missing values.
3. **Data Mapping**: I combined all the ontologies to map all the terms and functions across multiple databases. This helped in having a robust database with information from diverse sources.
4. **Data validation:** Since the data was checked by an immunologist and well referenced through multiple sources, i have high confidence in collecting all relevant data.

The final integrated database serves as a single resource for the subsequent modeling steps, giving us well-referenced, and multilayered information on immune system-modulating drugs, their targets, and their effects on cellular metabolism.

3.1.2 Selection Criteria

To choose the data i went through multiple stages of selecting data. These criteria were made to narrow down our focus to the most relevant information to study the immunomodulatory drugs along with the Genome-Scale Metabolic Models (GEMMs).

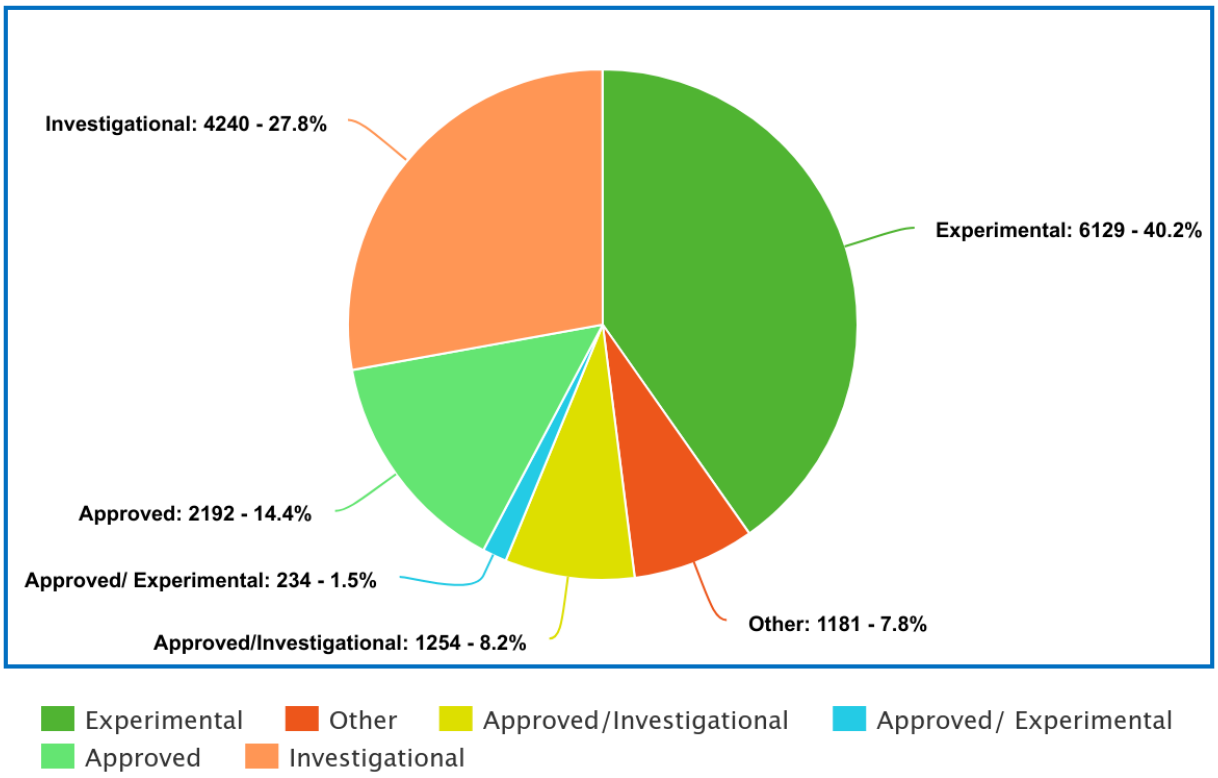

Figure 3.3: Distribution of drugs by development stage

Figure 3.3 shows the status of various drug groups . the largest number of drugs are experimental which have not cleared the pre clinical trials(6,129 drugs). Next major segment is the clinical trials stage or the investigational drugs(4,240 drugs). And the approved drugs are only 14.4 percent of the entire drugs dataset (2,192).

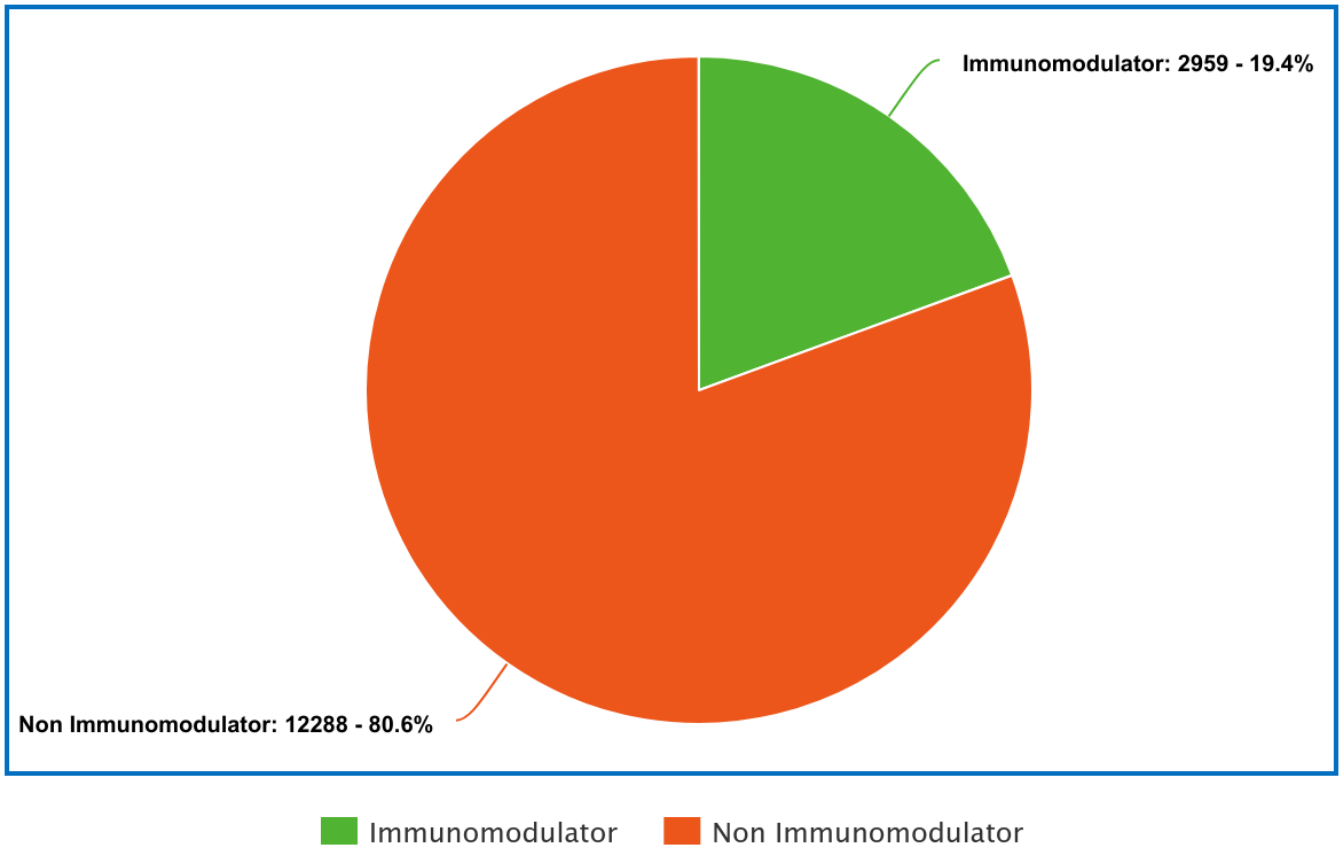

Figure 3.4: Proportion of immunomodulation drugs in our dataset

Figure 3.4 shows the classification of the drugs into immunomodulators and non immunomodulators based on the diseases they are indicated. We can see that (2,959 drugs) 19.4 percent of drugs are drugs that affect the immune system.

Multi-level selection criteria were divided into categories:

Drug type:

1. Focus on drugs targeting only the immune cells whose mechanism of action is known.

2. Including small molecule drugs and biologics and then using small molecules for the first phase of testing on models.
3. Priority was given to the FDA-approved and investigational drugs in Phase III clinical trials.

I selected the drugs from various classes, including but not restricted to class:

1. Corticosteroids
2. Cytokine modulators
3. JAK inhibitors
4. Monoclonal antibodies
5. T-cell modulators
6. B-cell modulators

These include the immunosuppressive and immunostimulatory drugs that lead to the modulation of the immune system.

I picked the drugs that modulate the immune cells in different stages (e.g., activation, proliferation, memory, effector)

Mechanism of action:

1. The drugs i selected have a very well-known mechanism of action related to affecting the immune system.
2. These drugs were categorized based on whether they affect the signaling or the metabolic pathways that will help during the modeling process.
3. Choosing the drugs with mechanisms that are most likely to be tested with a genome-scale metabolic model.

4. Choosing drugs with novel mechanisms of action to expand the scope of our research.

Target dependent selection:

1. Choosing the drugs with well-known targets for using the same targets but different drugs for different diseases,
2. Choosing the single target and multi-target drug to capture a range of effects.
3. Choosing the drugs with the known targets that exist in the Genome-scale metabolic model or can be added easily to the models.
4. Choosing based on known cellular and subcellular localization of drug target for an accurate model.

Disease based selection:

1. Focusing on the drugs that are used for the treatment of immune system-related pathologies.
2. Adding broad classification of conditions for building robust models, including:
3. Autoimmune disease (i.e., rheumatoid arthritis, multiple sclerosis, vitiligo).
4. Allergy
5. Immunodeficiency
6. Cancer treatment
7. Transplant treatment.
8. Choosing a disease that has a major impact on public health and no treatment available.

Choosing based on available data:

1. Choose based on sufficient pharmacological parameters availability.
2. Picking the drugs with specific safety profiles and comparing them with other therapies.
3. Choosing drugs with potential repurposed based on theoretical hypothesis.

Metabolic effect:

1. Choosing drugs with known effects on cellular metabolism.

2. Choosing the drugs that interact with specific pathways in the genome-scale metabolic models.

Choosing drugs that modulate the metabolism of immune cells such as:

   a. Glycolysis in activated T cell.
   b. Fatty acid oxidation I memory t cell.
   c. Glutamine involvement in immune cell growth.

Literature reference:

1. I added the links to the literature for the drug mechanism of action,
2. I prioritize the drugs that are more recent as compared to the old drugs just to maintain clinical relevance.
3. I also considered the impact of the literature, favoring or prioritizing the literature with a higher impact factor.

To further improve the use of the dataset, i did further classification and annotation:

1. **Pathway mapping:** Every drug was mapped to the appropriate biological pathway, which includes its primary mechanism of action and the target. The mapping used data from KEGG, Reactome, Panther DB, and DAVID.
2. **Specific cell mapping:** Based on the known targets and the pathway genes of the immune cells, i categorized the drugs that modulate the specific cells in particular stages.

3. **Dosage sensitivity:** I also added the information for the dosage, this would show the relevance of the drugs with higher doses or lower doses helping us to potentially repursue a lower dose drug for similar conditions to reduce the side effects.
4. **Drug interaction information:** I added information about immune-modulating drug interaction and potential agonist or antagonist behavior for the immune systems. This would be useful for combinatorial drug treatments.

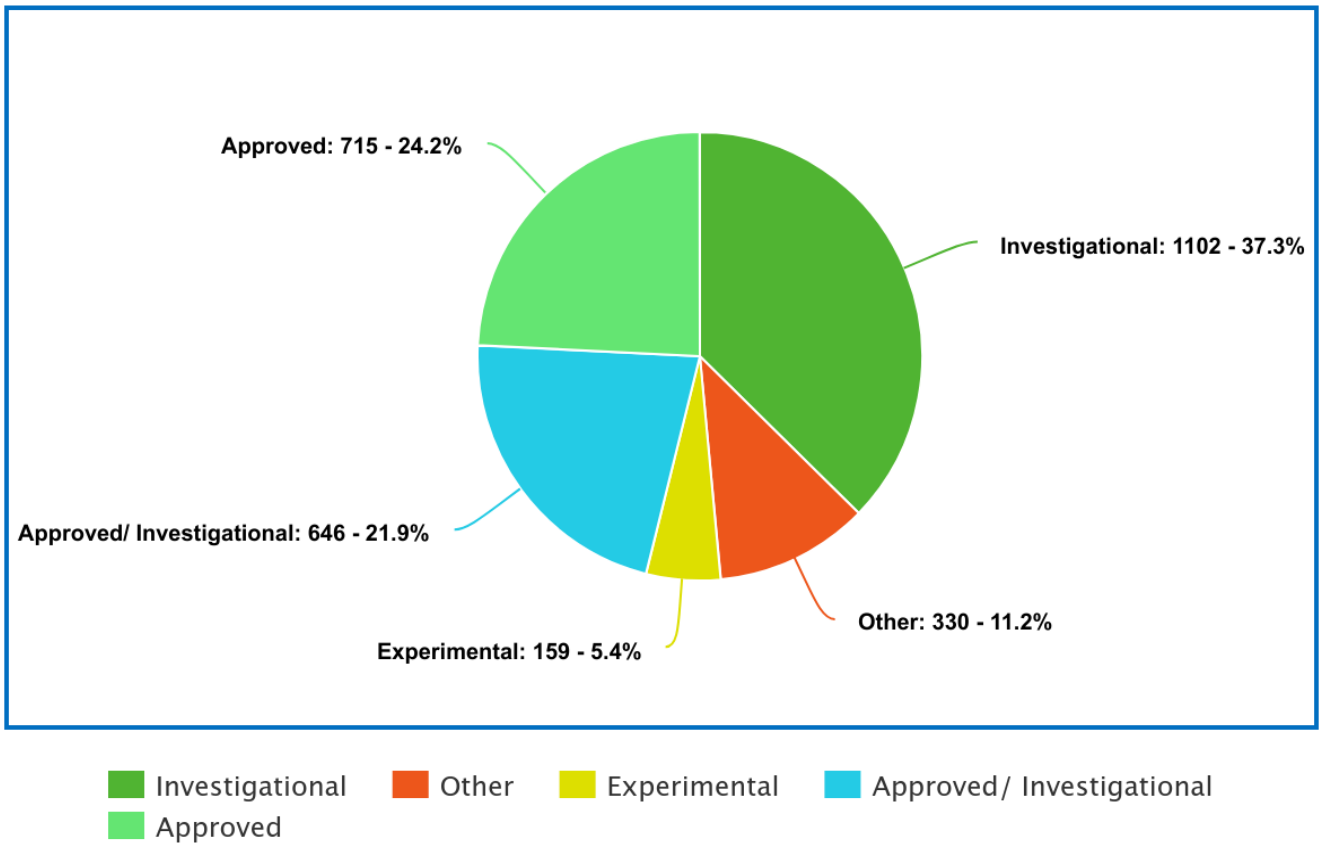

Figure 3.5: Distribution of immunomodulator drugs by group

Figure 3.5 shows the distribution of immunomodulator drugs where approved drugs are 24.2 percent(715 drugs) these are the drugs with already known targets and can be easily used given if the diseases that need focus share the same target or mechanism of action.

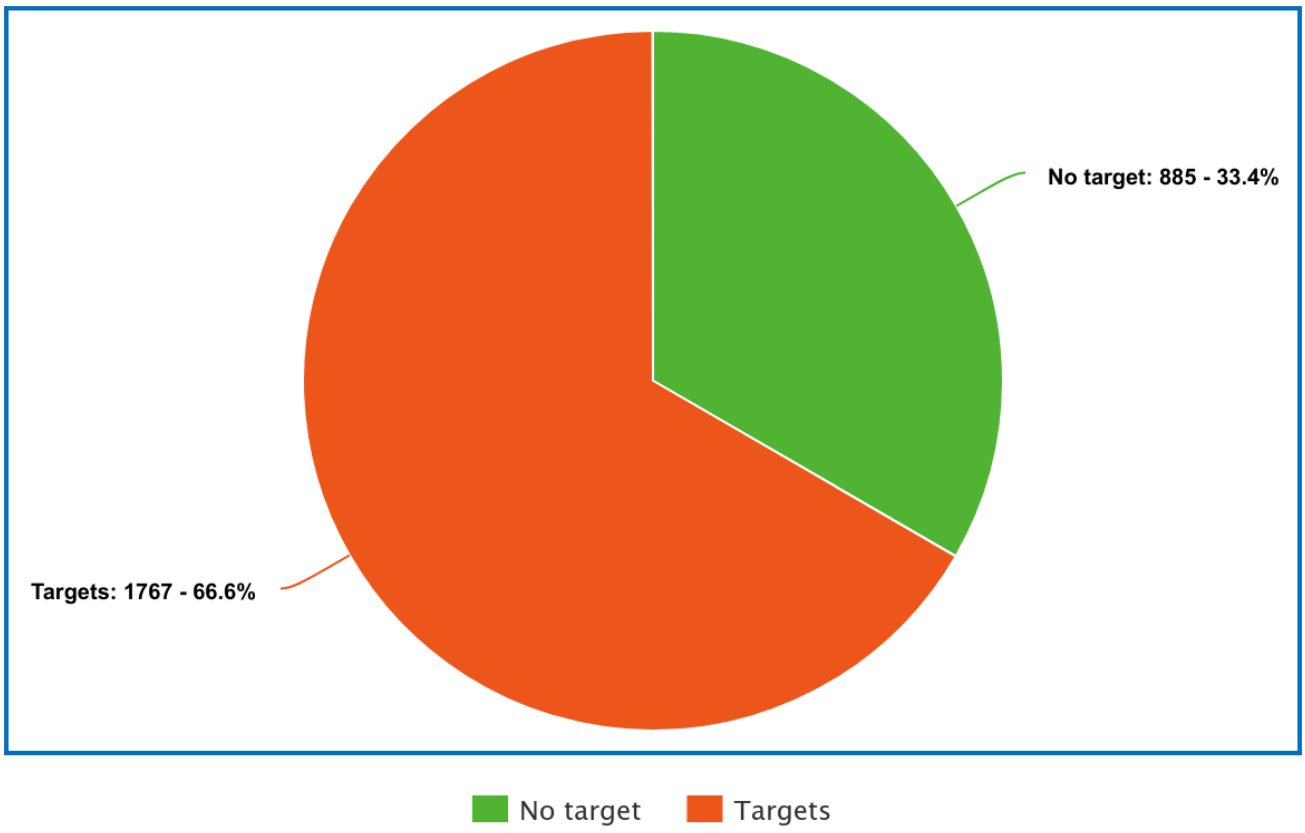

Figure 3.6: Distribution of immunomodulator drugs with known targets

Figure 3.6 shows the distribution of the immunomodulator drugs based on whether they have a known target showing that 66.6 % of drugs have known targets, so based on the targets, we can easily repurpose the drugs.

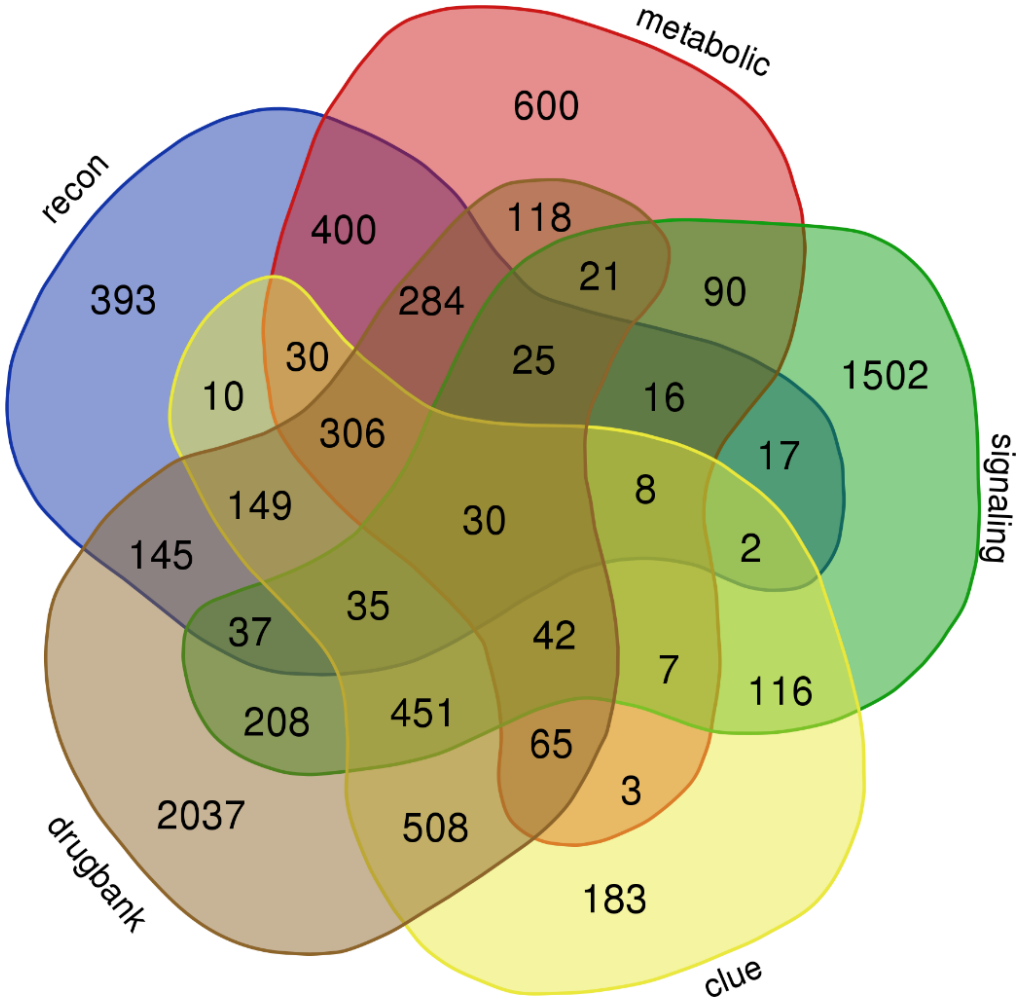

Figure 3.7: Drug target classification based on biological processes.

Figure 3.7 shows that despite having all the target lists, there are dug targets that do not affect the metabolism or the signaling, which gives us an opportunity to explore more in future work as well.

3.2 Modeling Approach

Search to model the effects of immune system modulating drugs on the metabolism using genome-scale metabolic models combines different computational methods and biological knowledge.

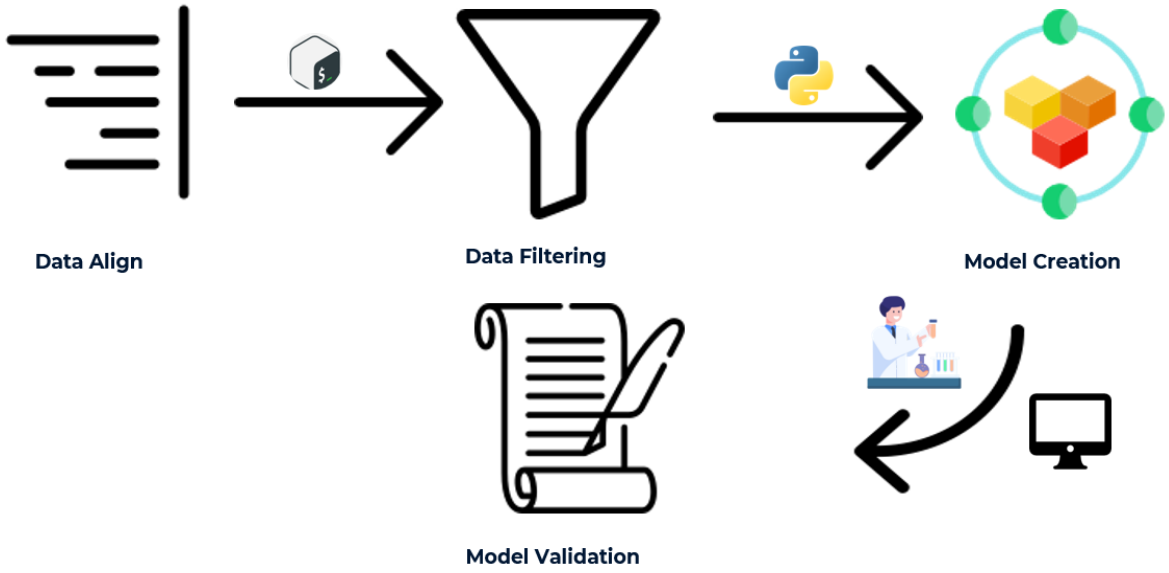

Figure 3.8: Modeling approach workflow

Figure 3.8 shows the pipeline workflow for the modeling process, where we analyze the transcriptomic data and filter out the noise in the expression. Then we build the model using the Como pipeline (Bessell et al., 2023), after which we validate the model using literature results for biological correctness.

The outline for the modeling pipeline is:

1. Base model building and curation.
2. As the base model due to its detailed information that gives us a comprehensive map of the human metabolism.

3. The models had detailed information on the metabolic pathways that are important for immune cell functioning.
4. For every single pathway i had a literature reference.
5. After building the model immunologists and metabolism experts checked and refined it for biological accuracy and relevance.
6. Cell type-specific model building

For creating different metabolic profiles of the immune cell types in different stages, the Helikar lab built a Bash and Python-based pipeline:

1. **Data integration:** I collected all the cell type-specific gene expression data from different research to not have any bias within the immune cell model.
2. **Creation of context-specific model:** using the Cobra toolbox we mapped the gene expression data to the metabolic map for making an immune cell-specific model.
3. **Validation of the function:** every immune cell-specific model is validated for the known metabolic behaviors using knowledge from experimental literature. We also built scripts for Insilco tests to simulate various scenarios (e.g., naïve b cell using glycolysis pathway)
4. **Pattern-based modeling**: For the comparison of the function of different cells in different diseases we built their models. (e.g., t-cell metabolism map for diabetes and Rheumatoid arthritis patients)
5. I add the reaction of the drugs that have common targets or pathways involved for testing the changes in the metabolism.

The pattern-based modeling provides a method for investigating the common effect of immunomodulatory drugs. By focusing on the similar pathways informed or targets involved we can detect the major metabolic changes in different diseases and the drug effects. This would help in drug response prediction and give us new mechanisms that haven't been targeted before for immune system modulation.

3.2.1 Creating Models

Building accurate and comprehensive models is the base step in our research for understanding the effect of the immune system modeling drugs on the metabolism(Solyar et al. 2007). Building a model includes multiple steps, to ensure the models contain the maximum amount of detail of immune cell metabolism,

Multiscale model:

we developed a multi-scale model that includes the integration of molecular cellular and tissue-level data(Fletcher et al. 2013):

molecular level: the base of our models is the detailed representation of metabolic reactions which includes the enzymes and the regulatory networks. We use the graph-based (Escher map)

representation or show the metabolic networks which allows us to easy interpretation and integrate(Ebrahim et al. 2013).

Cellular level:

Building on the molecular level knowledge, cell-type-specific metabolic features included the cell-type-specific membrane transporters, the known expression of metabolic pathways, and cells. This model includes different nutrient uptake metabolic choices (for example high amount of flux in the glycolytic pathway of activated T cells), the enzyme expression pattern, and other metabolic interactions of various immune cell types(Aoki-Kinoshita & Kanehisa, 2007). This approach helps us for modification of the metabolic models of different immune cells and gives us the ability to do a comprehensive comparative analysis of the metabolic activity.

Tissue level:

Integrate multiple immune cell types and focus on the intercellular metabolic interaction. For example, in type one diabetes i modeled the pancreatic cells which would include the data for beta cells activated T cells, and affected T cells, this would help us simulate how the changes in one metabolic state would lead to changes in the other cell types and how the changes affect the disease condition.

Data-driven model refinement:

To make the model relevant to biological data we also refine it:

Multi-omics approach: the COMO pipeline (Bessell et al. 2023) built by the Helikar lab integrates the transcriptomic proteomic and metabolic data into the models. This approach helps us in having the model parameters based on the experiment data which accounts for the biological behavior.

Literature-based validation: i created pipelines to go through the literature to extract relevant metabolic information. The information is automatically written in a CSV file.

These patterns-based methods for the creation and analysis of the model help us to understand the immune system and to simulate the repurposed drugs. By focusing on the metabolic patterns across different levels of biology and different diseases we can get novel insights into the underlying principles of immune system metabolic processes and the interaction with the drugs.

3.2.2 Integration of drug data with GEMMs

The integration of the genome-scale metabolic models with the drug information is important to understand the effect of repurposed immune system modulating drugs on the metabolism(Wishart et al. 2018). The pipeline involved multiple patterns-based approaches for accurate and comprehensive integration of information to the metabolic models.

Pattern-based drug assigning:

Drug structural pattern: Depending on the canonical smiles of the drugs i identified the repeating patterns of the drug molecules for specific diseases(Gaulton et al. 2016). This would help us in classifying and grouping the drugs that have similar structural information and we can predict the potential effects on the immune cell metabolism.

Target-based pattern: The pipeline includes the method for finding similar targets of the drugs in the metabolism of different immune cell types and diseases Richard helps us in having more accurate modeling of different repurposed drugs and understanding their effects on the metabolism.

To further understand the systemic effects of drugs on the metabolism, we used network-based approaches:

metabolic network perturbation: we built a metabolic model to check the patterns of the metabolic pathway for perturbations due to different drug which would allow us to find the potential drugs that have similar metabolic perturbation ability.

Metabolic network flux analysis: After building the metabolic models I checked the patterns of the metabolic flux distributions that are affected by the known drugs(Oughtred et al. 2018) this helped in understand how drugs affect the flow of metabolites through different pathways allowing us to find novel repurpose drugs that would have the desired flux.

The pattern-based approach for integrating data with GEMM gives a detailed method to predict and understand the effect of drugs that are repurposed for modulating immune cell metabolism(Huang et al., 2008).

Chapter 4

Results

4.1 Model Performance Evaluation

The pattern-based methodology for repurposing the drugs focused on the effects on the metabolism of the drugs that modulate the immune system(Lee et al. 2019; Napolitano et al 2013). I checked the model's ability to recognize the metabolic patterns linked with the truck and predict the potential new application of the drugs(Liu et al. 2022).

The primary data set i used for building the model and evaluation of the model contained 19,394 unique drugs, 15,248 from the Drug Bank, and 6,798 from the clue database(Smolen et al., 2000). This Big data set helped us to Check drug-dependent metabolic changes across a wide range of classes and targets.

I used a cross-validation method to test the drugs, 1st i checked the flux of the known drugs on the metabolism then i tested the novel repurposed drugs with similar targets and if the flux of the metabolism was near 80% of the known drug for the disease condition i consider it as a high confidence drug(Paterson et al. 2003). This potential accuracy shows the model's ability to consider all the complex interconnected metabolic interactions associated with the drug.

The model's ability to showcase the flux differences of different drugs has a high potential for translation in pharmaceutical research. For instance, in the case of diabetes type 1 pancreatic cells along with the T cells clearly showed the changes in the glycolytic pathway because of the drug metformin(Feist et al. 2007). Similarly, in the case of memory T cells, i saw a change in the fatty acid oxidation flux(Kurbatov et al. 2008).

The key strength of the approach test is to find new metabolic patterns and add new drugs that can be used for immune system-related diseases(Islam et al. 2015). For example, in the case of drugs related to allergies should be a change in flux of branched chain amino acid metabolism showcasing that this new mechanism of action can be further investigated(Andreoni et al. 2014).

The pipelines' ability to predict potential repurposed drugs was tested using different disease conditions(Stolyar et al., 2007). One of the major examples was the model suggesting the repurposing of diabetes type 1 drug metformin based on the effect on the activated T cell metabolism(Rhee et al. 2011). This research has also been supported by experimental literature(Benavente et al. 2015).

While the pipeline has very good predicting capabilities it also comes with its limitations (Bodey, 1996). it cannot repurpose drugs that have no literature for mechanism of action, or that have very complex interactions. Also, the drugs with multiple targets all targeting different metabolic pathways cannot be repurposed confidently.

To address these limitations the future goal also includes pharmacokinetic and pharmacodynamic data and the pattern-based approach(Brunk et al. 2018). Future research will also focus on the dynamic nature of the metabolic model which would be included in time-based

metabolomic data for capturing the changes in the metabolic patterns of the drugs over time(Damiani et al. 2019).

4.2 Case Studies

For case studies, I used a similar methodology pattern-based approach for different applications. I first found the patterns linked with the known drugs and targets (Case study 1: autoimmune disease) and then pathway inhibition(Case study 2). We use data flux balance analysis to simulate the effects of the drugs in the pathways of activated t cells. I found the drugs and effects based on the patterns. I focused on glycolysis and glutaminolysis pathways and found drugs that potentially reduced the flux in these pathways(inhibitors).

4.2.1 Case Study 1: Drug Repurposing for Autoimmune Diseases

For testing the application of the pipeline to repurpose the immune system modulating drugs using a pattern-based approach and GSMNs, i did a case study by finding potential drugs for repurposing in case of autoimmune diseases. I specifically targeted rheumatoid arthritis and systemic lupus erythematosus because of the high number of people being affected and the critical need for more effective treatments(Feist & Palsson, 2010).

The methodology was to analyze the metabolic patterns with the known drugs that are already used for the treatment of rheumatoid arthritis and systemic lupus erythematosus, and then look for the drugs that would have a similar metabolic pattern and are intended for use in different disease conditions. I use the data from the Drug Bank and clue database to get

comprehensive information about the drug targets, mechanisms and the stage of research is in(Nakaya et al. 2011).

The model highlighted multiple metabolic patterns that were affected by drugs used for treating autoimmune diseases(Segre et al. 2002):

Decrease in the flux of the glycolytic pathway in the case of activated T cells(Yizhak et al. 2015).

The change in flux of the lipid metabolism of naïve B cells(Zhang & Hua, 2016).

Using metabolic changes as references, the pipeline predicted a few high-confidence potential drugs to repurpose:

1. **Dimethyl fumarate (DMF):** Dimethyl fumarate is a drug that is approved for multiple sclerosis; this drug showed effective metabolic change, much more like the drugs that are already used for the treatment of rheumatoid arthritis; this drug was particularly effective on activated T cell metabolism(Ben Sahara et al. 2010). The dimethyl fumarate reduced the flux of the glycolytic pathway.
2. **Metformin:** This is a drug primarily used for the treatment of diabetes; this drug shows a potential to affect the T cell metabolism in case of systemic lupus erythematosus and the kynurenine pathway(Bhinder et al. 2021). Recent experimental literature has also proven this prediction showing reliable results in animal models(Cicchese et al. 2018).
3. **Statins:** while these drugs are usually used for managing cholesterol, the pipeline predicted that statin is useful in the case of both rheumatoid arthritis and systemic lupus erythrocytosis due to the effect on the lipid metabolism of the immune cells(Herrgard, 2006). This prediction was also validated through literature that states the modulatory effects of it on the immune system(Holms, 1996).

The case study shows the potential of the pattern-based GSM approach to find new drugs for repurposing(Justiniano et al. 2008). By combining the experimental and computational approaches i can accelerate the drug discovery process and bring in New Hope to the patients suffering from autoimmune diseases(Libman et al. 2015).

The case study shows the potential of a pattern-based genome-scale metabolic model approach for drug repurposing in autoimmune diseases, specifically in cases of rheumatoid arthritis and systemic lupus erythematosus. By looking at the patterns, we found candidates like Dimethyl fumarate, metformin, and statins. DMF is indicated for multiple sclerosis, which targets KEAP1 and also decreases the glycolic pathway flux in the activated t cells, which play a significant role in rheumatoid arthritis inflammatory response, and also inhibits the synoviocytes inflammation by leading to oxidation in the metabolism of T cells(Cai et al., 2020; Li et al., 2023; Zafari et al., 2023). Metformin is indicated for diabetes and is a good candidate for SLE due to its effect on T-cell metabolism. (Kempkes et al., 2019). Statins are the potential candidates for both diseases as they affect the lipid metabolism in the immune cells specifically for T-cell proliferation and function, thus reducing the activation of the t cells (Kempkes et al., 2019).

4.2.2 Case Study 2: Efficacy in Metabolic Pathway Intervention

In this case study i use a pattern-based genome-scale metabolic network to find the efficacy of the drugs targeting the specific metabolic pathways which are important for the functioning of the immune cell(Oh et al. 2007). I focused on the glycolysis and the glutamine pathways in the T cells and did not scale the metabolic network to find and predict the metabolic pattern linked with the drug treatments(Sellwood et al. 2018).

Process:

1. **Identifying the pattern:** We simulated the various stages of the cell metabolism, which were memory active naïve and effector t cells. I found the known effects of the approved drugs in the case of modulating glycolysis and glutaminolysis. The simulation included:
2. Flux balance analysis for metabolic flux distributions.
3. Flux variability analysis for finding a range of flux for reactions.

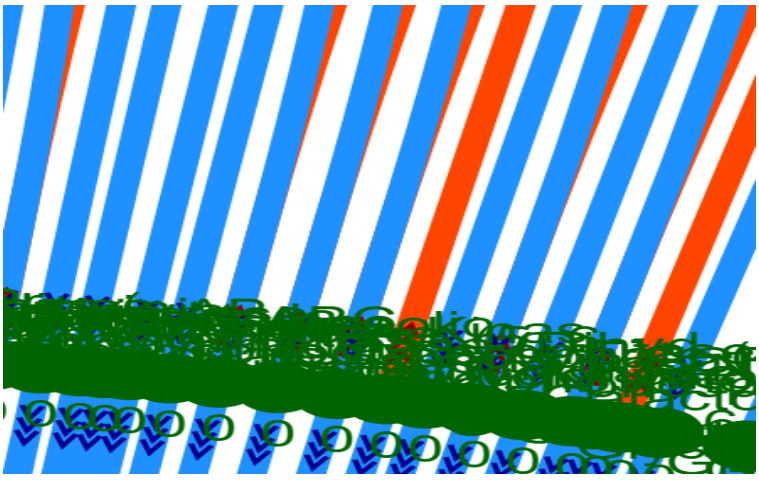

Figure 4.1 The visualization of the genome example scale model for the pathway highlighted

Figure 4.1 shows the metabolic flux, where the pathways that get affected are shown.

Based on these, a few flux changes were:

For glycolysis pathway inhibition:

1. Low flux for phosphofructokinase and pyruvate kinase.
2. Low flux from glucose to lactate.
3. High flux in pentose phosphate pathway,

for glutaminolysis Pathway inhibition:

1. Low flux for glutamate dehydrogenase.
2. Low flux in the TCA cycle.

Using these flux changes as references I found a few potential drugs:

1. **2-Deoxy-D-glucose (2-DG):** This was able to reduce the flux of the glycolytic pathway, it reduced lactic production, and it was able to reduce glucose uptake. Which aligns with the simulation as well.
2. **Metformin:** this drug was able to inhibit the glycolysis and glutamine in this pathway it also increased the fatty acid pathway flux.

This case study shows the potential application for finding inhibitors of the pathways, by using gsmn and pattern-based fba analysis.

Chapter 5

Discussion

5.1 Interpretation of Results

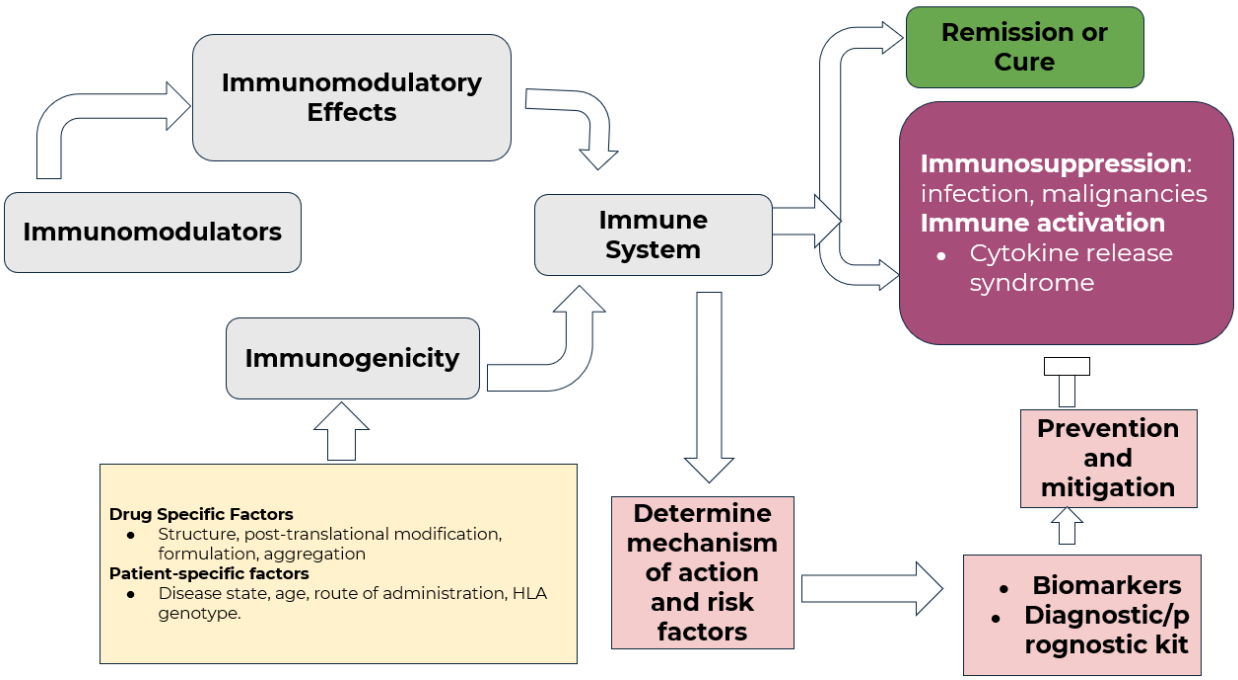

Figure 5.1: Potential impacts and applications of immunomodulator repurposing approach

Figure 5.1 shows the interaction between the immunomodulators and the immune system, showcasing the importance of pathway knowledge to have therapeutic effects. The immunomodulators have effects on the immune system that might have several factors such as

the structure, formulation, and specific factor, etc., these factors lead to specific mechanisms of action and the linked risks, which help in preventing and mitigation of the disease through biomarkers or the diagnosis methods,

The pipeline successfully predicted the drug-based metabolic changes, the pipeline's potential to find the metabolic pathway flux change pattern due to the known drugs and compare it with the effect of the repurposed drugs based on similar targets gave us new points to investigate(Sidders et al. 2018). Drugs with flux with nearly 80% similarity or higher than known drugs for the autoimmune disease condition were considered highly effective drugs.

The base dataset used for building and analyzing the model contained 19,394 drugs (15,248 from the drug bank database and 6,798 from the clue database), giving a solid foundation(Sun et al., 2012). This broad dataset helps us to understand the wide range of drug-dependent metabolic change patterns for different targets and classes of drugs, increasing the predictive capability and generalization of the pipeline (Thompson et al., 2018).

The primary capability of the pipeline is to leverage the flux differences of different drugs and use it as a potential for translating into drug research (Wen at al. 2017). The case study of diabetes type 1 pancreatic cells and t cells showed metabolic changes in the glycolysis pathway(Yusuke et al. 2017). I also noticed the change in the fatty acid pathway flux in the case of memory t cells which aligned with the known immune cell’s knowledge (Yizhak et al. 2015).

The pipeline helped in finding new potential uses for the approved drugs a key example is metformin, it used for patients suffering from type 1 diabetes condition(Zhang & Hua, 2016). Based on its effect on Activated T cell metabolism it can potentially be used for other disease conditions(Bhinder et al. 2021).

For the autoimmune disease research, the pipeline highlighted a few important metabolic patterns that can be used as targets for treatment.

1. Decreased metabolic flux in the glycolytic pathway in the activated T cells.
2. Change in the flux of lipid metabolism of naïve B cells.

Based on these metabolic changes i got high-confidence drugs to repurpose:

1. **Dimethyl fumarate (DMF):** this drug is used for the treatment of multiple sclerosis, dimethyl fumarate has a metabolic change that proves to be effective for the treatment of rheumatoid arthritis, specifically affecting the naive T cell metabolism.
2. **Metformin:** Generally used for the treatment of diabetes, metformin has the potential to be used for the treatment of systemic lupus erythematosus based on its effect on T cell metabolism.
3. **Statins:** they are used to control cholesterol, and they are potentially useful for helping in the treatment of Rheumatoid arthritis and systemic lupus erythematosus because of their ability to affect the lipid metabolism of the immune cells.

The predictions match with the experimental study showing the pipeline's ability to find new drugs for the treatment of autoimmune diseases. This proves that the pipeline can find drugs that can be repurposed and prove to be successful in the treatment of the disease conditions.

5.2 Comparison with Existing Models

The combination of a pattern-based approach and genome-scale models for repurposing the drugs and predicting the effects on the metabolism gives us an edge over the existing

methods. Unlike constraint-based modeling like Flux Balance Analysis, which is particularly good at predicting the best flux distribution with specific constraints it has limitations in looking for the drug effects, the pattern-based approach has a layer of analysis that looks at repeated metabolic pathway patterns. This method helps in checking the major metabolic changes that might be skipped while doing individual reaction-based simulations. The pipeline can handle diverse conditions and biological systems that would help in discovering new interactions with the metabolic network

The advantage of the entire genome gives us a complete understanding of cellular metabolism which also shows the different off-target effects and the metabolic interactions. By doing simulation i check the flux changes in the entire metabolic network as compared to the one reaction so that i can predict the potential effects that have not been considered yet as well as also unveil the effect of perturbation of one metabolic pathway on the other metabolic pathways. This was proven when the prediction of metformin's dual effect on glycolysis and glutaminolysis in active t-cells was seen.

The basis of our research is dependent on the identification of the repeated patterns of the flux simulations across different simulations. The pattern-based analysis helps in finding those metabolic changes that are due to drug effects or the disease states, this caters to variability and helps in capturing the patterns rather than just numbers associated with flux. In case of the autoimmune diseases, this method helps in finding the primary metabolic effects caused due to the approved drugs, like the reduction of glycolytic flux in the case of activated t cells.

The pattern-based approach unlike the machine learning methods helps in tracing the changes caused by the drug though in the metabolism, showing how the cell changes the metabolic response when the drugs perturb or in disease conditions. By comparing the fluxes, i

can find the primary enzymes that have a major effect on the cellular and metabolic levels. This gives us new targets for the therapeutic effects. This method is especially useful when we trying to bulk repurpose the drugs and predict the side effects of the drugs.

The pattern-based approach gives comprehensive metabolic information, pathway-level patterns, and interactions. This approach works as a complementary method as compared to just replacing the existing methods. It proves to be a new tool for understanding drug actions and finding new opportunities for treatment in the context of cellular metabolism. By combining the genome-scale metabolic model with a pattern-based approach, we have built a method that can be used for guiding the new hypothesis and experiments and then helping in the acceleration of drug discovery and drug repurposing.

(Jamialahmadi et al., 2019) showed the combination of pattern-based approach with a genome-scale metabolic model improved the prediction accuracy of the drugs on cancer metabolism, showing the improved predicting drug response performance compared to the traditional flux balance analysis.

(Liu et al., 2019) did a large-scale drug repurposing research focusing on diabetes type 2, they used the genome-scale metabolic models to find the novel drug candidates some of which have been in to the preclinical validation stages.

(Rashid & Selvarajoo, 2024) showed the pattern-based approach method to find the metabolic interactions. They also found the off target effects while testing the FDA approved drugs, showing the importance of the research.

5.3 Limitations and Challenges

The pattern-based Genome-scale metabolic network (GSMN) approach, although it is very useful for predicting metabolic changes and drug repurposing, has few limitations. The complex nature of the metabolism regulation is a major constraint to address as the GSMN is primarily made for stoichiometric relation and does not include the drug effects that interact through the regulatory pathways. As a solution, I need to combine the regulator pathways with GSMNs.

The GSMN are based on steady-state assumption which limits our ability to understand the time-dependent drug effects. The metabolism changes over time, the model can miss the major changes that occur after drug consumption. Building pipelines for dynamic GSMN simulation can address the problem of understanding the metabolic changes over time.

The translation of the prediction to the in vivo experiments is still a question, due to the various levels like tissues and organs, as well as other systems such as the microbiome. These complex-level interactions are still not covered in our approach. To surpass this limitation, i would need to make a multiscale model that links the GSMN to the whole body metabolism on the microbiome and physiology level.

Trying to define a solution to these limitations is crucial for the advancement of the pipeline. By tackling these problems, we can increase our pipeline's accuracy, scope, and translational capability, Despite the limitations the pattern-based gsmn approach has given us few high confidences in drug targets that can be repurposed opening the door to multiple

opportunities for drug repurposing and finding. novel metabolism pattern. As we work in the future on this pipeline it would prove to play a key role in personalized medicine.

Chapter 6

Summary and future work

The pattern-based Genome-Scale Metabolic Network (GSMN) has shown the potential to repurpose the drug and predict the effect on the metabolism of the immune system, The ability was shown by identifying the metabolic pathways that were targeted for activated t cells. The strength of the approach is because of the comprehensive metabolic information, using the gsmn we get a holistic understanding of the immune cell metabolic changes due to drugs; this helps us in finding the off-target effects and unravel the metabolic interactions.

The pattern-based approach's ability to look at the metabolic patterns linked with drug effects and disease states is valuable. The case study of the autoimmune disease are example showing the importance of the identification of the metabolic pattern for effective drug treatment, The integration of transcriptomic data helps us to predict and validate the results with alignment with experimental research.

The drug repurposing efforts gave us promising drugs for autoimmune disease and predicted the potential effect of dimethyl fumarate in rheumatoid arthritis and metformin in the case of systems lulus erythematosus. Further, the pipeline also proves to be used for testing a combination of the drugs to see the metabolic changes.

In future work, we will be expanding the metabolic model to cover more drug effects and diseases. We would integrate other omics data build a dynamic gsmn model and incorporate the LSTM model to understand time-dependent drug effects and metabolic changes. We will be optimizing the pipeline to adjust to computational power. By pursuing these future goals, we aim to increase the power of our pipeline for metabolic research, drug repurposing, and personalized medicine.

References

1 Alf Spitschak, Gupta, S., Krishna Pal Singh, Logotheti, S., & Pützer, B. M. (2022). Drug Repurposing at the Interface of Melanoma Immunotherapy and Autoimmune Disease. Pharmaceutics, 15(1), 83–83. https://doi.org/10.3390/pharmaceutics15010083

2 Amiri, R., Jafar Razmara, Sepideh Parvizpour, & Habib Izadkhah. (2023). A novel efficient drug repurposing framework through drug-disease association data integration using convolutional neural networks. BMC Bioinformatics, 24(1). https://doi.org/10.1186/s12859-023-05572-x

3 Andreoni, C., Orsi, G., De Maria, C., Montemurro, F., & Vozzi, G. (2014). In Silico Models for Dynamic Connected Cell Cultures Mimicking Hepatocyte-Endothelial Cell-Adipocyte Interaction Circle. PLoS ONE, 9(12), e111946. https://doi.org/10.1371/journal.pone.0111946

4 Aoki-Kinoshita, K. F., & Kanehisa, M. (2007). KEGG Primer: An Introduction to Pathway Analysis Using KEGG. NCI Nature Pathway Interaction Database. https://doi.org/10.1038/pid.2007.2

5 Azer, K., & Leaf, I. (2023). Systems biology platform for efficient development and translation of multitargeted therapeutics. Frontiers in Systems Biology, 3. https://doi.org/10.3389/fsysb.2023.1229532

6 Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C. L., Serova, N., Davis, S., & Soboleva, A. (2012). NCBI GEO: archive

for functional genomics data sets—update. Nucleic Acids Research, 41(D1), D991–D995. https://doi.org/10.1093/nar/gks1193

7 Barsi, S., Papp, H., Valdeolivas, A., Tóth, D. J., Kuczmog, A., Madai, M., Hunyady, L., Várnai, P., Saez-Rodriguez, J., Jakab, F., & Szalai, B. (2022). Computational drug repurposing against SARS-CoV-2 reveals plasma membrane cholesterol depletion as key factor of antiviral drug activity. PLOS Computational Biology, 18(4), e1010021. https://doi.org/10.1371/journal.pcbi.1010021

8 Ben Sahra, I., Laurent, K., Giuliano, S., Larbret, F., Ponzio, G., Gounon, P., Le Marchand-Brustel, Y., Giorgetti-Peraldi, S., Cormont, M., Bertolotto, C., Deckert, M., Auberger, P., Tanti, J.-F., & Bost, F. (2010). Targeting Cancer Cell Metabolism: The Combination of Metformin and 2-Deoxyglucose Induces p53-Dependent Apoptosis in Prostate Cancer Cells. Cancer Research, 70(6), 2465–2475. https://doi.org/10.1158/0008-5472.can-09-2782

9 Benavente, L., Goldberg, A., Myburg, H., Chiera, J., Zapata, M., & Zyl, L. van. (2015). The application of systems biology to biomanufacturing. Pharmaceutical Bioprocessing, 3(4), 341–355. https://doi.org/10.4155/pbp.15.12

10 Bessell, B., Loecker, J., Zhao, Z., Sara Sadat Aghamiri, Mohanty, S., Amin, R., Tomáš Helikar, & Bhanwar Lal Puniya. (2023). COMO: a pipeline for multi-omics data integration in metabolic modeling and drug discovery. Briefings in Bioinformatics, 24(6). https://doi.org/10.1093/bib/bbad387

11 Bhinder, B., Gilvary, C., Madhukar, N. S., & Elemento, O. (2021). Artificial Intelligence in Cancer Research and Precision Medicine. Cancer Discovery, 11(4), 900–915. https://doi.org/10.1158/2159-8290.cd-21-0090

12 BODEY, G. P. (1966). Quantitative Relationships Between Circulating Leukocytes and Infection in Patients with Acute Leukemia. Annals of Internal Medicine, 64(2), 328. https://doi.org/10.7326/0003-4819-64-2-328

13 Brody, H., & Miller, F. G. (2013). The Research-Clinical Practice Distinction, Learning Health Systems, and Relationships. Hastings Center Report, 43(5), 41–47. https://doi.org/10.1002/hast.199

14 Brunk, E., Sahoo, S., Zielinski, D. C., Altunkaya, A., Dräger, A., Mih, N., Gatto, F., Nilsson, A., Preciat Gonzalez, G. A., Aurich, M. K., Prlić, A., Sastry, A., Danielsdottir, A. D., Heinken, A., Noronha, A., Rose, P. W., Burley, S. K., Fleming, R. M. T., Nielsen, J., & Thiele, I. (2018). Recon3D enables a three-dimensional view of gene variation in human metabolism. Nature Biotechnology, 36(3), 272–281. https://doi.org/10.1038/nbt.4072

15 CAO, D., & PARKER, R. (2001). Computational modeling of eukaryotic mRNA turnover. RNA, 7(9), 1192–1212. https://doi.org/10.1017/s1355838201010330

16 Chandrasekaran, S., & Price, N. D. (2010). Probabilistic integrative modeling of genome-scale metabolic and regulatory networks inEscherichia coliandMycobacterium tuberculosis. Proceedings of the National Academy of Sciences, 107(41), 17845–17850. https://doi.org/10.1073/pnas.1005139107

17 Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., & Blaschke, T. (2018). The rise of deep learning in drug discovery. Drug Discovery Today, 23(6), 1241–1250. https://doi.org/10.1016/j.drudis.2018.01.039

18 Choi, Y., Shin, B., Kang, K., Park, S., & Beck, B. R. (2020). Target-Centered Drug Repurposing Predictions of Human Angiotensin-Converting Enzyme 2 (ACE2) and Transmembrane Protease Serine Subtype 2 (TMPRSS2) Interacting Approved Drugs for

Coronavirus Disease 2019 (COVID-19) Treatment through a Drug-Target Interaction Deep Learning Model. Viruses, 12(11), 1325. https://doi.org/10.3390/v12111325

19 Christ, B., Collatz, M., Dahmen, U., Herrmann, K.-H., Höpfl, S., König, M., Lambers, L., Marz, M., Meyer, D., Radde, N., Reichenbach, J. R., Ricken, T., & Tautenhahn, H.-M. (2021). Hepatectomy-Induced Alterations in Hepatic Perfusion and Function - Toward Multi-Scale Computational Modeling for a Better Prediction of Post-hepatectomy Liver Function. Frontiers in Physiology, 12. https://doi.org/10.3389/fphys.2021.733868

20 Chu, G., Ji, X., Wang, Y., & Niu, H. (2023). Integrated multiomics analysis and machine learning refine molecular subtypes and prognosis for muscle-invasive urothelial cancer. Molecular Therapy - Nucleic Acids, 33, 110–126. https://doi.org/10.1016/j.omtn.2023.06.001

21 Cicchese, J. M., Evans, S., Hult, C., Joslyn, L. R., Wessler, T., Millar, J. A., Marino, S., Cilfone, N. A., Mattila, J. T., Linderman, J. J., & Kirschner, D. E. (2018). Dynamic balance of pro- and anti-inflammatory signals controls disease and limits pathology. Immunological Reviews, 285(1), 147–167. https://doi.org/10.1111/imr.12671

22 Cichonska, A., Ravikumar, B., Parri, E., Timonen, S., Pahikkala, T., Airola, A., Wennerberg, K., Rousu, J., & Aittokallio, T. (2017). Computational-experimental approach to drug-target interaction mapping: A case study on kinase inhibitors. PLOS Computational Biology, 13(8), e1005678. https://doi.org/10.1371/journal.pcbi.1005678

23 Cong, Y., Shintani, M., Imanari, F., Osada, N., & Endo, T. (2022). A New Approach to Drug Repurposing with Two-Stage Prediction, Machine Learning, and Unsupervised Clustering of Gene Expression. OMICS : A Journal of Integrative Biology, 26(6), 339–347. https://doi.org/10.1089/omi.2022.0026

24 Cai, W., Yu, Y., Zong, S., & Wei, F. (2020). Metabolic reprogramming as a key regulator in the pathogenesis of rheumatoid arthritis. Inflammation Research, 69(11), 1087–1101. https://doi.org/10.1007/s00011-020-01391-5

25 Deaths communicable diseases annually worldwide 2019. (n.d.). Statista. Retrieved July 28, 2024, from https://www.statista.com/statistics/282715/deaths-from-communicable-diseases-worldwide/

26 Dakshanamurthy, S., Issa, N. T., Assefnia, S., Seshasayee, A., Peters, O. J., Madhavan, S., Uren, A., Brown, M. L., & Byers, S. W. (2012). Predicting New Indications for Approved Drugs Using a Proteochemometric Method. Journal of Medicinal Chemistry, 55(15), 6832–6848. https://doi.org/10.1021/jm300576q

27 Damiani, C., Maspero, D., Di Filippo, M., Colombo, R., Pescini, D., Graudenzi, A., Westerhoff, H. V., Alberghina, L., Vanoni, M., & Mauri, G. (2019). Integration of single-cell RNA-seq data into population models to characterize cancer metabolism. PLOS Computational Biology, 15(2), e1006733. https://doi.org/10.1371/journal.pcbi.1006733

28 Dong, Y., Chen, J., Chen, Y., & Liu, S. (2023). Targeting the STAT3 oncogenic pathway: Cancer immunotherapy and drug repurposing. Biomedicine & Pharmacotherapy, 167, 115513–115513. https://doi.org/10.1016/j.biopha.2023.115513

29 Ebrahim, A., Lerman, J. A., Palsson, B. O., & Hyduke, D. R. (2013). COBRApy: COnstraints-Based Reconstruction and Analysis for Python. BMC Systems Biology, 7(1), 74. https://doi.org/10.1186/1752-0509-7-74

30 Ehsan Zangene, Marashi, S.-A., & Hesam Montazeri. (2023). SL-scan identifies synthetic lethal interactions in cancer using metabolic networks. Scientific Reports, 13(1). https://doi.org/10.1038/s41598-023-42992-4

31 Ekins, S., Puhl, A. C., Zorn, K. M., Lane, T. R., Russo, D. P., Klein, J. J., Hickey, A. J., & Clark, A. M. (2019). Exploiting machine learning for end-to-end drug discovery and development. Nature Materials, 18(5), 435–441. https://doi.org/10.1038/s41563-019-0338-z

32 Fabregat, A., Sidiropoulos, K., Garapati, P., Gillespie, M., Hausmann, K., Haw, R., Jassal, B., Jupe, S., Korninger, F., McKay, S., Matthews, L., May, B., Milacic, M., Rothfels, K., Shamovsky, V., Webber, M., Weiser, J., Williams, M., Wu, G., & Stein, L. (2015). The Reactome pathway Knowledgebase. Nucleic Acids Research, 44(D1), D481–D487. https://doi.org/10.1093/nar/gkv1351

33 Feist, A. M., & Palsson, B. O. (2010). The biomass objective function. Current Opinion in Microbiology, 13(3), 344–349. https://doi.org/10.1016/j.mib.2010.03.003

34 Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., Broadbelt, L. J., Hatzimanikatis, V., & Palsson, B. Ø. (2007). A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Molecular Systems Biology, 3(1), 121. https://doi.org/10.1038/msb4100155

35 FKurbatov, L. K., Cheglakov, I. B., Yarygin, K. N., Sukhikh, G. T., Vartanian, G. V., I. Yu. Toropygin, & Archakov, A. I. (2008). Comparative transcriptomics and proteomics of fetal and adult human liver. Biochemistry (Moscow). Supplement. Series B, Biomedical Chemistry, 2(3), 226–234. https://doi.org/10.1134/s1990750808030025

36 Fletcher, T., Galloway, T. S., Melzer, D., Holcroft, P., Cipelli, R., Pilling, L. C., Mondal, D., Luster, M., & Harries, L. W. (2013). Associations between PFOA, PFOS and changes

in the expression of genes involved in cholesterol metabolism in humans. Environment International, 57-58, 2–10. https://doi.org/10.1016/j.envint.2013.03.008

37 Gaulton, A., Hersey, A., Nowotka, M., Bento, A. P., Chambers, J., Mendez, D., Mutowo, P., Atkinson, F., Bellis, L. J., Cibrián-Uhalte, E., Davies, M., Dedman, N., Karlsson, A., Magariños, M. P., Overington, J. P., Papadatos, G., Smit, I., & Leach, A. R. (2016). The ChEMBL database in 2017. Nucleic Acids Research, 45(D1), D945–D954. https://doi.org/10.1093/nar/gkw1074

38 Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A. J., & Sawitzki, G. (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biology, 5(10), R80. https://doi.org/10.1186/gb-2004-5-10-r80

39 Ginestet, C. (2011). ggplot2: Elegant Graphics for Data Analysis. Journal of the Royal Statistical Society: Series a (Statistics in Society), 174(1), 245–246. https://doi.org/10.1111/j.1467-985x.2010.00676_9.x

40 Godfrey, W. H., & Kornberg, M. D. (2020). The Role of Metabolic Enzymes in the Regulation of Inflammation. Metabolites, 10(11), 426. https://doi.org/10.3390/metabo10110426

41 Hafemeister, C., & Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biology, 20(1). https://doi.org/10.1186/s13059-019-1874-1

42 Herrgard, M. J. (2006). Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in Saccharomyces cerevisiae. Genome Research, 16(5), 627–635. https://doi.org/10.1101/gr.4083206

43 Holmes, E., Wilson, I. D., & Nicholson, J. K. (2008). Metabolic Phenotyping in Health and Disease. Cell, 134(5), 714–717. https://doi.org/10.1016/j.cell.2008.08.026Korbee, C. J., Heemskerk, M. T., Kocev, D., van Strijen, E., Rabiee, O., Franken, K. L. M. C., Wilson, L., Savage, N. D. L., Džeroski, S., Haks, M. C., & Ottenhoff, T. H. M. (2018). Combined chemical genetics and data-driven bioinformatics approach identifies receptor tyrosine kinase inhibitors as host-directed antimicrobials. Nature Communications, 9(1). https://doi.org/10.1038/s41467-017-02777-6

44 Holms, H. (1996). Flux analysis and control of the central metabolic pathways inEscherichia coli. FEMS Microbiology Reviews, 19(2), 85–116. https://doi.org/10.1111/j.1574-6976.1996.tb00255.x

45 https://doi.org/10.1093/nar/gkp456

46 Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols, 4(1), 44–57. https://doi.org/10.1038/nprot.2008.211

47 Imami, A. S., McCullumsmith, R. E., & O'Donovan, S. M. (2021). Strategies to identify candidate repurposable drugs: COVID-19 treatment as a case example. Translational Psychiatry, 11(1). https://doi.org/10.1038/s41398-021-01724-w

48 in cancer. PLOS Computational Biology, 19(3), e1010200. https://doi.org/10.1371/journal.pcbi.1010200

49 Islam, M. A., Zengler, K., Edwards, E. A., Mahadevan, R., & Stephanopoulos, G. (2015). Investigating Moorella thermoacetica metabolism with a genome-scale constraint-based metabolic model. Integrative Biology, 7(8), 869–882. https://doi.org/10.1039/c5ib00095e

50 Juhn, Y., & Liu, H. (2020). Artificial intelligence approaches using natural language processing to advance EHR-based clinical research. Journal of Allergy and Clinical Immunology, 145(2), 463–469. https://doi.org/10.1016/j.jaci.2019.12.897

51 Jung, Y. J., Tweedie, D., Scerba, M. T., & Greig, N. H. (2019). Neuroinflammation as a Factor of Neurodegenerative Disease: Thalidomide Analogs as Treatments. Frontiers in Cell and Developmental Biology, 7. https://doi.org/10.3389/fcell.2019.00313

52 Justiniano, H., Berlingeri-Ramos, A. C., & Sánchez, J. L. (2008). Pattern Analysis of Drug-Induced Skin Diseases. The American Journal of Dermatopathology, 30(4), 352–369. https://doi.org/10.1097/dad.0b013e3181722ef4

53 Jamialahmadi, O., Hashemi-Najafabadi, S., Motamedian, E., Romeo, S., & Bagheri, F. (2019). A benchmark-driven approach to reconstruct metabolic networks for studying cancer metabolism. PLOS Computational Biology, 15(4), e1006936. https://doi.org/10.1371/journal.pcbi.1006936

54 Kempkes, R. W. M., Joosten, I., Koenen, H. J. P. M., & He, X. (2019). Metabolic Pathways Involved in Regulatory T Cell Functionality. Frontiers in Immunology, 10. https://doi.org/10.3389/fimmu.2019.02839

55 Kaneko, T. (2011). Drugs for neglected diseases: part I. Future Medicinal Chemistry, 3(10), 1235–1237. https://doi.org/10.4155/fmc.11.103

56 Khoo, K. H., Verma, C. S., & Lane, D. P. (2014). Erratum: Drugging the p53 pathway: understanding the route to clinical efficacy. Nature Reviews Drug Discovery, 13(4), 314–314. https://doi.org/10.1038/nrd4288

57 Kim, Y., Kim, G. B., & Lee, S. Y. (2021). Machine learning applications in genome-scale metabolic modeling. Current Opinion in Systems Biology, 25, 42–49. https://doi.org/10.1016/j.coisb.2021.03.001

58 Kort, E., & Jovinge, S. (2021). Drug Repurposing: Claiming the Full Benefit from Drug Development. Current Cardiology Reports, 23(6), 62. https://doi.org/10.1007/s11886-021-01484-5

59 Lee, I., Keum, J., & Nam, H. (2019). DeepConv-DTI: Prediction of drug-target interactions via deep learning with convolution on protein sequences. PLOS Computational Biology, 15(6), e1007129. https://doi.org/10.1371/journal.pcbi.1007129

60 Libman, I. M., Miller, K. M., DiMeglio, L. A., Bethin, K. E., Katz, M. L., Shah, A., Simmons, J. H., Haller, M. J., Raman, S., Tamborlane, W. V., Coffey, J. K., Saenz, A. M., Beck, R. W., & Nadeau, K. J. (2015). Effect of Metformin Added to Insulin on Glycemic Control Among Overweight/Obese Adolescents With Type 1 Diabetes. JAMA, 314(21), 2241. https://doi.org/10.1001/jama.2015.16174

61 Liu, Q., Lee, B., & Xie, L. (2022). Small molecule modulation of microbiota: a systems pharmacology perspective. BMC Bioinformatics, 23(S3). https://doi.org/10.1186/s12859-022-04941-2

62 López-Agudelo, V. A., Mendum, T. A., Laing, E., Wu, H., Baena, A., Barrera, L. F., Dany, & Rios-Estepa, R. (2020). A systematic evaluation of Mycobacterium tuberculosis

Genome-Scale Metabolic Networks. PLOS Computational Biology, 16(6), e1007533–e1007533. https://doi.org/10.1371/journal.pcbi.1007533

63 Li, Q., Chen, Y., Liu, H., Tian, Y., Yin, G., & Xie, Q. (2023). Targeting glycolytic pathway in fibroblast-like synoviocytes for rheumatoid arthritis therapy: Challenges and opportunities. Inflammation Research, 72(12), 2155–2167. https://doi.org/10.1007/s00011-023-01807-y

64 Liu, P., Li, H., Li, S., & Leung, K.-S. (2019). Improving prediction of phenotypic drug response on cancer cell lines using deep convolutional network. BMC Bioinformatics, 20(1), 408. https://doi.org/10.1186/s12859-019-2910-6

65 Masseroli, M., & Pinciroli, F. (2006). Using Gene Ontology and genomic controlled vocabularies to analyze high-throughput gene lists: Three tool comparison. Computers in Biology and Medicine, 36(7-8), 731–747. https://doi.org/10.1016/j.compbiomed.2005.04.008

66 McEntire, R., Szalkowski, D., Butler, J., Kuo, M. S., Chang, M., Chang, M., Freeman, D., McQuay, S., Patel, J., McGlashen, M., Cornell, W. D., & Xu, J. J. (2016). Application of an automated natural language processing (NLP) workflow to enable federated search of external biomedical content in drug discovery and development. Drug Discovery Today, 21(5), 826–835. https://doi.org/10.1016/j.drudis.2016.03.006

67 Nakaya, H. I., Wrammert, J., Lee, E. K., Racioppi, L., Marie-Kunze, S., Haining, W. N., Means, A. R., Kasturi, S. P., Khan, N., Li, G.-M., McCausland, M., Kanchan, V., Kokko, K. E., Li, S., Elbein, R., Mehta, A. K., Aderem, A., Subbarao, K., Ahmed, R., & Pulendran,

B. (2011). Systems biology of vaccination for seasonal influenza in humans. Nature Immunology, 12(8), 786–795. https://doi.org/10.1038/ni.2067

68 Napolitano, F., Zhao, Y., Moreira, V. M., Tagliaferri, R., Kere, J., D'Amato, M., & Greco, D. (2013). Drug repositioning: a machine-learning approach through data integration. Journal of Cheminformatics, 5(1). https://doi.org/10.1186/1758-2946-5-30

69 Nicora, G., Vitali, F., Dagliati, A., Geifman, N., & Bellazzi, R. (2020). Integrated Multi-Omics Analyses in Oncology: A Review of Machine Learning Methods and Tools. Frontiers in Oncology, 10. https://doi.org/10.3389/fonc.2020.01030

70 Oh, Y.-K., Palsson, B. O., Park, S. M., Schilling, C. H., & Mahadevan, R. (2007). Genome-scale Reconstruction of Metabolic Network in Bacillus subtilis Based on High-throughput Phenotyping and Gene Essentiality Data. Journal of Biological Chemistry, 282(39), 28791–28799. https://doi.org/10.1074/jbc.m703759200

71 Oh, Y.-K., Palsson, B. O., Park, S. M., Schilling, C. H., & Mahadevan, R. (2007). Genome-scale Reconstruction of Metabolic Network in Bacillus subtilis Based on High-throughput Phenotyping and Gene Essentiality Data. Journal of Biological Chemistry, 282(39), 28791–28799. https://doi.org/10.1074/jbc.m703759200

72 Ortiz, D. G. (2007). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches [Book Review]. Qualitative Research Journal, 6(2), [205]207. https://doi.org/10.3316/qrj0602205

73 Oughtred, R., Stark, C., Breitkreutz, B.-J., Rust, J., Boucher, L., Chang, C., Kolas, N., O'Donnell, L., Leung, G., McAdam, R., Zhang, F., Dolma, S., Willems, A., Coulombe-Huntington, J., Chatr-aryamontri, A., Dolinski, K., & Tyers, M. (2018). The BioGRID

interaction database: 2019 update. Nucleic Acids Research, 47(D1), D529–D541. https://doi.org/10.1093/nar/gky1079

74 Pan, Y., Ren, H., Lan, L., Li, Y., & Huang, T. (2023). Review of Predicting Synergistic Drug Combinations. Life, 13(9), 1878. https://doi.org/10.3390/life13091878

75 Paterson, A. H., Bowers, J. E., Peterson, D. G., Estill, J. C., & Chapman, B. A. (2003). Structure and evolution of cereal genomes. Current Opinion in Genetics & Development, 13(6), 644–650. https://doi.org/10.1016/j.gde.2003.10.002

76 Payne, N. L., Sylvain, A., O'Brien, C., Herszfeld, D., Sun, G., & Bernard, C. C. A. (2015). Application of human induced pluripotent stem cells for modeling and treating neurodegenerative diseases. New Biotechnology, 32(1), 212–228. https://doi.org/10.1016/j.nbt.2014.05.001

77 Rahman, Md. M., Saha, T., Islam, K. J., Suman, R. H., Biswas, S., Rahat, E. U., Hossen, Md. R., Islam, R., Hossain, M. N., Mamun, A. A., Khan, M., Ali, M. A., & Halim, M. A. (2020). Virtual screening, molecular dynamics and structure–activity relationship studies to identify potent approved drugs for Covid-19 treatment. Journal of Biomolecular Structure and Dynamics, 39(16), 6231–6241. https://doi.org/10.1080/07391102.2020.1794974Sebastián Ayala-Ruano, Yovani Marrero-Ponce, Longendri Aguilera-Mendoza, Pérez, N., Guillermin Agüero-Chapin, Antunes, A., & Ana Cristina Aguilar. (2022). Network Science and Group Fusion Similarity-Based Searching to Explore the Chemical Space of Antiparasitic Peptides. ACS Omega, 7(50), 46012–46036. https://doi.org/10.1021/acsomega.2c03398

78 Rahman, S., Sowa, M. E., Ottinger, M., Smith, J. A., Shi, Y., Harper, J. W., & Howley, P. M. (2011). The Brd4 Extraterminal Domain Confers Transcription Activation Independent

of pTEFb by Recruiting Multiple Proteins, Including NSD3. Molecular and Cellular Biology, 31(13), 2641–2652. https://doi.org/10.1128/mcb.01341-10

79 Rapin, N., Lund, O., Bernaschi, M., & Castiglione, F. (2010). Computational Immunology Meets Bioinformatics: The Use of Prediction Tools for Molecular Binding in the Simulation of the Immune System. PLoS ONE, 5(4), e9862. https://doi.org/10.1371/journal.pone.0009862

80 Rhee, K. Y., Carvalho, L. P. S. de, Bryk, R., Ehrt, S., Marrero, J., Park, S. W., Schnappinger, D., Venugopal, A., & Nathan, C. (2011). Central carbon metabolism in Mycobacterium tuberculosis: an unexpected frontier. Trends in Microbiology, 19(7), 307–314. https://doi.org/10.1016/j.tim.2011.03.008

81 Rashid, M. M., & Selvarajoo, K. (2024). Advancing drug-response prediction using multi-modal and -omics machine learning integration (MOMLIN): A case study on breast cancer clinical data. Briefings in Bioinformatics, 25(4), bbae300. https://doi.org/10.1093/bib/bbae300

82 Rossi, R. L., & Grifantini, R. M. (2018). Big Data: Challenge and Opportunity for Translational and Industrial Research in Healthcare. Frontiers in Digital Humanities, 5. https://doi.org/10.3389/fdigh.2018.00013

83 Segrè, D., Vitkup, D., & Church, G. M. (2002). Analysis of optimality in natural and perturbed metabolic networks. Proceedings of the National Academy of Sciences of the United States of America, 99(23), 15112–15117. https://doi.org/10.1073/pnas.232349399

84 Sellwood, M. A., Ahmed, M., Segler, M. H., & Brown, N. (2018). Artificial intelligence in drug discovery. Future Medicinal Chemistry, 10(17), 2025–2028. https://doi.org/10.4155/fmc-2018-0212

85 Sen, P., & Orešič, M. (2019). Metabolic Modeling of Human Gut Microbiota on a Genome Scale: An Overview. Metabolites, 9(2), 22. https://doi.org/10.3390/metabo9020022

86 Shen, T., Wang, H., Hu, R., & Yanni Lv. (2023). Developing neural network diagnostic models and potential drugs based on novel identified immune-related biomarkers for celiac disease. Human Genomics, 17(1). https://doi.org/10.1186/s40246-023-00526-z

87 Sidders, B., Karlsson, A., Kitching, L., Torella, R., Karila, P., & Phelan, A. (2018). Network-Based Drug Discovery: Coupling Network Pharmacology with Phenotypic Screening for Neuronal Excitability. Journal of Molecular Biology, 430(18), 3005–3015. https://doi.org/10.1016/j.jmb.2018.07.016

88 Smolen, P., Baxter, D. A., & Byrne, J. H. (2000). Mathematical Modeling of Gene Networks. Neuron, 26(3), 567–580. https://doi.org/10.1016/s0896-6273(00)81194-0

89 Srimadh, K., Sujith Kumar Pulukool, Sai Sanwid Pradhan, Saiswaroop R, Ashwin Ashok Naik, Datta, & Venketesh Sivaramakrishnan. (2022). Systems biology approach delineates critical pathways associated with disease progression in rheumatoid arthritis. Journal of Biomolecular Structure and Dynamics/Journal of Biomolecular Structure & Dynamics, 41(14), 6969–6990. https://doi.org/10.1080/07391102.2022.2115555

90 Stolyar, S., Van Dien, S., Hillesland, K. L., Pinel, N., Lie, T. J., Leigh, J. A., & Stahl, D. A. (2007). Metabolic modeling of a mutualistic microbial community. Molecular Systems Biology, 3(1), 92. https://doi.org/10.1038/msb4100131

91 Stolyar, S., Van Dien, S., Hillesland, K. L., Pinel, N., Lie, T. J., Leigh, J. A., & Stahl, D. A. (2007). Metabolic modeling of a mutualistic microbial community. Molecular Systems Biology, 3(1), 92. https://doi.org/10.1038/msb4100131

92 Sun, Y., Zhu, R., Hao, Y., Tang, K., Zhao, J., Chen Yu-jia, Liu, Q., & Cao, Z. (2012). Towards a bioinformatics analysis of anti-Alzheimer's herbal medicines from a target network perspective. 14(3), 327–343. https://doi.org/10.1093/bib/bbs025

93 Thompson, A. J., Baranzini, S. E., Geurts, J., Hemmer, B., & Ciccarelli, O. (2018). Multiple sclerosis. The Lancet, 391(10130), 1622–1636. https://doi.org/10.1016/s0140-6736(18)30481-1

94 Turanli, B., Grøtli, M., Boren, J., Nielsen, J., Uhlen, M., Arga, K. Y., & Mardinoglu, A. (2018). Drug Repositioning for Effective Prostate Cancer Treatment. Frontiers in Physiology, 9. https://doi.org/10.3389/fphys.2018.00500

95 Wang, L., Lu, Y., Li, D., Zhou, Y., Yu, L., Ines Mesa Eguiagaray, Campbell, H., Li, X., & Evropi Theodoratou. (2024). The landscape of the methodology in drug repurposing using human genomic data: a systematic review. Briefings in Bioinformatics, 25(2). https://doi.org/10.1093/bib/bbad527

96 Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. Technological Forecasting and Social Change, 126(1), 3–13.

97 Wang, Y., Xiao, J., Suzek, T. O., Zhang, J., Wang, J., & Bryant, S. H. (2009). PubChem: a public information system for analyzing bioactivities of small molecules. Nucleic Acids Research, 37(Web Server), W623–W633.

98 Wen, M., Zhang, Z., Niu, S., Sha, H., Yang, R., Yun, Y., & Lu, H. (2017). Deep-Learning-Based Drug–Target Interaction Prediction. Journal of Proteome Research, 16(4), 1401–1409. https://doi.org/10.1021/acs.jproteome.6b00618

99 Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., & Pon, A. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Research, 46(D1), D1074–D1082. https://doi.org/10.1093/nar/gkx1037

100 Xu, X., Mazloom, R., Goligerdian, A., Staley, J., Amini, M., Wyckoff, G. J., Riviere, J., & Jaberi-Douraki, M. (2019). Making Sense of Pharmacovigilance and Drug Adverse Event Reporting: Comparative Similarity Association Analysis Using AI Machine Learning Algorithms in Dogs and Cats. Topics in Companion Animal Medicine, 37, 100366. https://doi.org/10.1016/j.tcam.2019.100366

101 Xu, Y., Chen, K., Pan, J., Lei, Y., Zhang, D., Fang, L., Tang, J., Chen, X., Ma, Y., Zheng, Y., Bao, Z., Zhou, Y., Zhan, J., & Xu, W. (2021). Repurposing clinically approved drugs for COVID-19 treatment targeting SARS-CoV-2 papain-like protease. International Journal of Biological Macromolecules, 188, 137–146. https://doi.org/10.1016/j.ijbiomac.2021.07.184

102 Yang, F., Wang, W., Wang, F., Fang, Y., Tang, D., Huang, J., Lu, H., & Yao, J. (2022). scBERT as a large-scale pretrained deep language model for cell type annotation of single-cell RNA-seq data. Nature Machine Intelligence, 4(10), 852–866. https://doi.org/10.1038/s42256-022-00534-z

103 Yizhak, K., Chaneton, B., Gottlieb, E., & Ruppin, E. (2015). Modeling cancer metabolism on a genome scale. Molecular Systems Biology, 11(6), 817. https://doi.org/10.15252/msb.20145307

104 Yusuke TAKESHIMA, IWASAKI, Y., OKAMURA, T., Keishi FUJIO, & YAMAMOTO, K. (2017). The metabolic regulation in immune cells and pathogenesis of systemic lupus erythematosus ~toward new therapeutic applications~. Nihon Rinshō Men'eki Gakkai Kaishi/Nihon Rinshou Men'eki Gakkai Kaishi, 40(1), 12–20. https://doi.org/10.2177/jsci.40.12

105 Zeng, Z., Shi, H., Wu, Y., & Hong, Z. (2015). Survey of Natural Language Processing Techniques in Bioinformatics. Computational and Mathematical Methods in Medicine, 2015, 1–10. https://doi.org/10.1155/2015/674296

106 Zhang, C., & Hua, Q. (2016). Applications of Genome-Scale Metabolic Models in Biotechnology and Systems Medicine. Frontiers in Physiology, 6. https://doi.org/10.3389/fphys.2015.00413

107 Zhang, C., & Hua, Q. (2016). Applications of Genome-Scale Metabolic Models in Biotechnology and Systems Medicine. Frontiers in Physiology, 6. https://doi.org/10.3389/fphys.2015.00413

108 Zhang, C., Aldrees, M., Arif, M., Li, X., Mardinoglu, A., & Aziz, M. A. (2019). Elucidating the Reprograming of Colorectal Cancer Metabolism Using Genome-Scale Metabolic Modeling. Frontiers in Oncology, 9. https://doi.org/10.3389/fonc.2019.00681

109 Zhang, C., Aldrees, M., Arif, M., Li, X., Mardinoglu, A., & Aziz, M. A. (2019). Elucidating the Reprograming of Colorectal Cancer Metabolism Using Genome-Scale Metabolic Modeling. Frontiers in Oncology, 9. https://doi.org/10.3389/fonc.2019.00681

110 Zafari, P., Taghadosi, M., Faramarzi, F., Rajabinejad, M., & Rafiei, A. (2023). Dimethyl Fumarate Inhibits Fibroblast Like Synoviocytes-mediated Inflammation and Joint

Destruction in Rheumatoid Arthritis. Inflammation, 46(2), 612–622.

https://doi.org/10.1007/s10753-022-01759-1