IDENTIFICATION OF IMMUNE-RELATED GENE SIGNATURES TO EVALUATE IMMUNOTHARAPEUTIC RESPONSE IN CANCER PATIENTS USING EXPLORATORY SUBGROUP DISCOVERY

A Dissertation

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

OLHA KHOLOD

Dr. Chi-Ren Shyu, Dissertation Supervisor

DECEMBER 2022

The undersigned, appointed by the Dean of the Graduate School, have examined the thesis entitled

IDENTIFICATION OF IMMUNE-RELATED GENE SIGNATURES TO EVALUATE IMMUNOTHARAPEUTIC RESPONSE IN CANCER PATIENTS USING EXPLORATORY SUBGROUP DISCOVERY

Presented by OLHA KHOLOD

A candidate for the degree of Doctor of Philosophy

And hereby certify that, in their opinion, it is worthy of acceptance.

_____

Dr. Chi-Ren Shyu

_____

Dr. Jonathan Mitchem

_____

Dr. Jussuf Kaifi

_____

Dr. Richard Hammer

DEDICATION

This dissertation is dedicated to all brave Ukrainian defenders from the Armed Forces of Ukraine who is currently fighting unjustified war against Russia. You are the heroes of our times.

ACKNOWLEDGMENTS

First, I would like to acknowledge my academic advisor Dr. Chi-Ren Shyu who gave me the opportunity to be trained in his laboratory and taught me skills required to become an independent researcher. Additionally, I would like to thank my former academic advisor Dr. Dmitriy Shin, with whom I spent the first years of my PhD and who immensely contributed to my computational training. Furthermore, I would like to acknowledge my committee members Dr. Jonathan Mitchem, Dr. Jussuf Kaifi and Dr. Richard Hammer for their guidance and encouragement throughout these years. I also would like to thank my laboratory mates Dr. Danlu Liu and William Basket for their precious help with exploratory subgroup discovery algorithm.

I want to thank all talented educators and scientists whom I encountered in my scientific career: my high school mentors – Dr. Kateryna Lystvan and Dr. Yuliya Luchakivska; my academic and research advisors from Taras Shevchenko National University – Dr. Iryna Kozeretska, Dr. Lyudmyla Shvachko, Dr. Dmytro Krasnienkov and Dr. Oleksandr Kolyada. You have always been a great inspiration to me, and I am very grateful for all time you invest in my personal and professional development.

I am so lucky to have a close circle of friends who were there for me in times of happiness and difficulties. Firstly, I want to thank Dr. William Meyers for numerous coffee meetings and his endless support for Ukrainian people, especially in such challenging times. Secondly, I want to thank Alyona Rozenblat, Viktoriia and Logan Muehlman and their daughter Mila for make me feel like home and for making me a better person. Thirdly, I want to thank Daryna Haponiuk and Mariia Bondarenko for

being my rock-solid support across the Atlantic Ocean. In addition, I want to thank my partner James McMillen for his patience, incredible kindness, and his love. Finally, I would like to express my very profound gratitude to my mother Halyna Vlasenko, my late father Volodymyr Kholod and to my amazing elder sister Mariia Kholod. I would not be where I am today without your endless help, love, and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

IDENTIFICATION OF IMMUNE-RELATED GENE SIGNATURES TO EVALUATE IMMUNOTHARAPEUTIC RESPONSE IN CANCER PATIENTS USING EXPLORATORY SUBGROUP DISCOVERY

Olha Kholod

Dr. Chi-Ren Shyu, Dissertation Supervisor

ABSTRACT

Phenotypic and genotypic heterogeneity are characteristic features of cancer patients. To tackle patients' heterogeneity, immune checkpoint inhibitors (ICIs) represent one of the most promising therapeutic approaches. However, approximately 50% of cancer patients that are eligible for treatment with ICIs will not respond well, which motivates the exploration of immunotherapy in combination with either targeted treatments or chemotherapy. Over the years, multiple patient stratification techniques have been developed to identify homogenous patient subgroups, although, matching patient subgroup to treatment option that can improve patients' health outcome remains a challenging task.

We extend our exploratory subgroup discovery algorithm to identify patient subpopulations that can potentially benefit from immuno-targeted combination therapies or chemoimmunotherapy in five cancer types: Head and Neck Squamous Carcinoma (HNSC), Lung Adenocarcinoma (LUAD), Lung Squamous Carcinoma (LUSC), Skin Cutaneous Melanoma (SKCM) and Triple-Negative Breast Cancer (TNBC). We employ various regression models to identify immune-related gene signatures and drug targets

that increase the likelihood of partial remission on combination therapies, either immuno-targeted regimen or chemoimmunotherapy. Moreover, our pipelines can pinpoint adverse drug effects associated with predicted drug combinations. In addition, we uncovered distinct immune cell populations (T-cells, B-cells, Myeloid, NK-cells) for TNBC patients that differentiate patients with partial remission from patients with progressive disease after chemoimmunotherapy. Finally, we incorporate our methodological developments on Mutational Forks Formalism that enable an assessment of patient-specific flow by leveraging information from multiple single-nucleotide alterations to adjust the transitional likelihoods that are solely based on the canonical view of a disease.

Our suit of methods can help to better select responders for combination therapies and improve health outcome for cancer patients with limited treatment options.

Chapter 1 Introduction

1.1 Motivations

In 2022, 1.9 million new cancer cases and 600 thousand cancer deaths are projected to occur in the United States[1]. Cancer immunotherapy is a promising innovative treatment for many forms of cancer. Although immunotherapy has been shown to be efficacious, patient response rates vary and, often only a small subset of the patients within a large cohort respond favorably to the treatment[2]. One potential explanation is that immunotherapies are not yet widely used as first-line treatments, they are typically given to patients whose immune systems are already compromised due to advanced disease and/or previous therapies[3]. Another reason might be that acquired genetic and epigenetic alterations activate immunosuppressive signaling pathways that contribute to immunotherapy resistance[4]. These issues promote the exploration of immunotherapy in combination with targeted treatments or chemotherapy.

To reduce the cost of prospective clinical trials, novel computational algorithms are needed to evaluate the efficacy of combination treatments with immunotherapy for cancer patients. It is also important to account for potential adverse reactions from these therapeutic combinations. The exploratory subgroup discovery algorithm[5] is capable of identifying multiple homogenous patients' subgroups based on both phenotypic and genotypic parameters in unsupervised manner. Public data repositories, such as cBioPortal[6], has a collection of multidimensional cancer genomics datasets and can be employed for data mining tasks. However, these datasets are usually lacking information on primary and secondary outcomes from clinical studies. By augmenting exploratory subgroup discovery with advanced regression analysis and with usage of high-quality clinical datasets, we can infer phenotypic and genotypic

predictors that would help to differentiate patients with partial remission from patients with progressive disease after combination treatments with immunotherapy. Ultimately, it will help to select potential responders and/or exclude potential non-responders for combination treatments with immunotherapy and it is expected that such therapies will eventually become the standard of care for cancer treatment.

1.2 Summary of Contributions

The contributions of this dissertation are summarized as follows:

- Creates probabilistic framework based on Mutational Forks Formalism to decipher patient-specific flow of signal transduction perturbed by multiple single-nucleotide alterations. In addition, this approach uses ontologically enriched pathway data that preserve the diversity of gene interaction terms. Our framework follows a specific, biologically explainable logic, and can thus be employed as hypothesis generation tool in precision medicine settings.
- Provides informatic framework based on exploratory mining algorithm to identify immuno-targeted combination therapies for cancer patients with EGFR wild-type gene. In addition, our pipeline can pinpoint adverse drug effects associated with predicted drug combinations. Our approach can help to better select responders for immuno-targeted combination therapies and improve health outcome for cancer patients with EGFR wild-type gene.
- Augments exploratory mining algorithm for predicting the efficacy of chemoimmunotherapy for cancer patients with limited treatment options, such as triple-negative breast cancer. In addition, our approach is capable of

distinguishing unique immune cell populations that differentiate patients with partial remission from patients with progressive disease after chemoimmunotherapy.

- Disseminates informatic tools and data to scientific community. The extension of exploratory mining algorithm will be available as Jupiter notebooks under iDAS lab GitHub repository.

1.3 Thesis Organization

This dissertation is organized as follows. Chapter 2 introduces our Mutational Forks Formalism to identify patient-specific flow of signal transduction perturbed by multiple single-nucleotide alterations. Chapter 3 explains the extension of exploratory subgroup discovery algorithm for identifying immuno-targeted combination therapies for cancer patients with EGFR wild-type gene. Chapter 4 describes the augmentation of explanatory subgroup discovery algorithm for uncovering immune-related gene signatures to predict the effectiveness of chemoimmunotherapy in triple-negative breast cancer patients. Finally, we summarize this dissertation and discuss future works at the end.

Chapter 2 Identifying Patient-Specific Flow of Signal Transduction Perturbed by Multiple Single-Nucleotide Alterations

2.1 Introduction

Signaling pathways play a crucial role in the transformation of external signals into gene regulation mechanisms. In this process, external information is transferred through a series of biochemical reactions, which can be referred to as 'signal flow'. Dysregulation of signal flow by a certain mutation may trigger the development of pathological processes, including cancer[7]. It is especially evident in individual patient cases, where the inferred flow can illustrate the deviation from a canonical form of the disease and therefore explain drug resistance or pave the way to new therapeutic regimens.

Various computational methods have been developed for pathway analysis in recent years[8, 9]. However, only a small number of papers have focused on the development of models that do not primarily rely on kinetic parameters. To overcome the difficulty in accurately estimating kinetic parameters, pathway topology methods employ various characteristics of a pathway network, such as node degree, betweenness, clustering coefficient, and shortest path length to predict functionally important nodes in the network. For example, Zhao & Liu analyzed topological parameters of complex disease genes and revealed the importance of their location in a biomolecular network[10]. Erten et al. developed an algorithm called Vavien for harnessing the topological similarity of proteins in a network of interactions to prioritize candidate disease-associated genes[11]. Ning et al. attempted to examine the placement of essential proteins and the network topology by determining the correlation of protein essentiality and reverse nearest

neighbor topology (RNN)[12]. However, such methods are limited only to a small network of proteins and do not infer the gene-to-gene direction of signal propagation in biological systems.

To address these limitations, Lee & Cho developed a novel algorithm that can estimate signal flow using only topological information to predict the direction of activity change in various signaling networks[13]. They extended their work to the Signal Flow Control (SFC) method, which is capable of identifying control targets without the information of kinetic parameter values[14]. The limitation of these probabilistic topological methods is that the flow of signal transduction is equally distributed among connected nodes. Therefore, the weights of the edges are roughly approximated from topology and heavily depend on the number of connecting nodes. In addition, only binary interactions have been encoded to the model, such as activation and inactivation.

A separate group of methods deals with inference of perturbation patterns in signaling systems. For example, Santolini & Barabasi proposed a series of network topology-based models named "DYNamics-Agnostic Network MOdels (DYNAMO)", which can retrieve the relative magnitude of biological perturbation patterns when lacking knowledge of the kinetic parameters[15]. The DYNAMO series models achieve an average of 65% accuracy in predicting perturbation patterns of both distance-based and biochemical models, which shows promising results. Li & Gao enhanced the distance model in DYNAMO and leveraged graph convolutional networks (GCN) to improve perturbation pattern prediction[16]. However, the limitation of pathway perturbation methods, such as DYNAMO and GCN, is that they do not focus on inferring the gene-to-gene direction of signal propagation, instead pinpointing localities where pathway perturbation occurred.

In this work, we present an improvement of the Mutational Forks formalism which can handle a combination of single-nucleotide alterations for inferring patient-specific flow. By definition, a mutational fork is a subgraph where signal transduction can flow through alternative paths. We have assigned prior probabilities to the flow of signal transduction through particular entities based on literature findings, rather than approximating it solely from topology. In addition, we employed ontologically enriched pathway data that preserve the diversity of gene interaction terms, such as phosphorylation-activation, ubiquitination, binding, degradation, etc. We have also conducted two case studies to show that the proposed algorithm follows a specific, biologically explainable logic, and can thus be employed as hypothesis generation tool in precision medicine settings.

2.2 Materials and Methods

2.2.1 Definitions

This section provides a brief overview of definitions in the Mutational Fork formalism (Figure 1). Here, we use bold font for all pathway entities and italic font for functions and probability distributions. For more details, please refer to our previous publication[17].

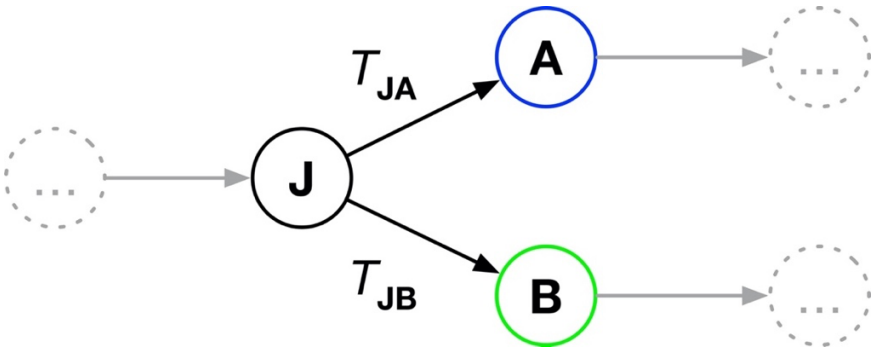

Figure 1: An example of a dual mutational fork. **J** stands for junction gene, **A** and **B** denote shoulder genes, T_{JA} signifies transition of signal flow from the node **J** to the node **A**, T_{JB}

represents transition of signal transduction flow from the node **J** to the node **B**, black arrows depict the direction of signal flow in the fork. Dashed-line circles represent incoming and outcoming nodes in the fork, grey arrows denote the direction of signal flow from incoming and outcoming nodes in the fork.

- $\mathbf{J_{AB}}$ – *mutational fork*, a place where signal transduction can flow through alternative paths
- **J** – *junction* gene, where signal transduction flow splits
- **A**, **B** – *shoulder* genes, the first genes in alternative paths
- **JA**, **JB** – legs of the mutational fork $\mathbf{J_{AB}}$
- $T_{\mathbf{JA}}(\mathbf{J_{AB}})$ – transition of signal transduction flow to leg **JA** of the mutational fork $\mathbf{J_{AB}}$
- $T_{\mathbf{JB}}(\mathbf{J_{AB}})$ – transition of signal transduction flow to leg **JB** of the mutational fork $\mathbf{J_{AB}}$
- $P(\mathrm{T}_{\mathbf{JA}}(\mathbf{J_{AB}}))$ – probability of $T_{\mathbf{JA}}(\mathbf{J_{AB}})$
- $P(\mathrm{T}_{\mathbf{JB}}(\mathbf{J_{AB}}))$ – probability of $T_{\mathbf{JB}}(\mathbf{J_{AB}})$
- $M(\mathbf{X})$ – mutation of gene **X**
- $C(\mathrm{M}(\mathbf{X}))$ – mutation condition of gene **X**
- $A(\mathbf{X})$ – status of gene **X**
- $C(A(\mathbf{X}))$ – activation condition of gene **X**
- $P(M(\mathbf{A}) \mid T_{\mathbf{JA}}(\mathbf{J_{AB}}))$ – probability of $M(\mathbf{A})$ given $T_{\mathbf{JA}}(\mathbf{J_{AB}})$
- $P(M(\mathbf{A}) \mid T_{\mathbf{JB}}(\mathbf{J_{AB}}))$ – probability of $M(\mathbf{A})$ given $T_{\mathbf{JB}}(\mathbf{J_{AB}})$
- $P(M(\mathbf{B}) \mid T_{\mathbf{JB}}(\mathbf{J_{AB}}))$ – probability of $M(\mathbf{B})$ given $T_{\mathbf{JB}}(\mathbf{J_{AB}})$
- $P(\mathrm{M}(\mathbf{B}) \mid T_{\mathbf{JA}}(\mathbf{J_{AB}}))$ – probability of $M(\mathbf{B})$ given $T_{\mathbf{JA}}(\mathbf{J_{AB}})$
- $P(T_{\mathbf{JA}}(\mathbf{J_{AB}}) \mid M(\mathbf{A})=1)$ – probability of $T_{\mathbf{JA}}(\mathbf{J_{AB}})$ given $M(\mathbf{A}) = 1$

- $P(T_{\mathbf{JA}}(\mathbf{J_{AB}}) \mid M(\mathbf{B})=1)$ – probability of $T_{\mathbf{JA}}(\mathbf{J_{AB}})$ given $M(\mathbf{B})=1$
- $P(T_{\mathbf{JB}}(\mathbf{J_{AB}}) \mid M(\mathbf{B})=1)$ – probability of $T_{\mathbf{JB}}(\mathbf{J_{AB}})$ given $M(\mathbf{B})=1$
- $P(T_{\mathbf{JB}}(\mathbf{J_{AB}}) \mid M(\mathbf{A})=1)$ – probability of $T_{\mathbf{JB}}(\mathbf{J_{AB}})$ given $M(\mathbf{A})=1$

We define mutation and activation statuses of a particular gene **X**, given a corresponding mutational event as shown in Equation 1 and Equation 2. The gene is mutated/activated if the corresponding conditions $C(M(\mathbf{X}))$ and $C(\boldsymbol{A}(\mathbf{X}))$ are met. In this work, the activation condition is set when the expression value of **X** exceeds the average gene expression values for **X** among all patients after a mutational event occurred. The mutation condition $C(M(\mathbf{X}))$ is enabled when a single-nucleotide polymorphism (SNP) in gene **X** occurred.

$$M(\mathbf{X}) = \begin{cases} 1, \textit{ if } C\big(M(\boldsymbol{X})\big) \textit{ is met}; \\ 0, \textit{ otherwise;} \end{cases}, \text{ value 1 means gene } \mathbf{X} \text{ is mutated;} \quad (1)$$

$$A(\mathbf{X}) = \begin{cases} 1, \textit{ if } C\big(A(\boldsymbol{X})\big) \textit{ is met}; \\ 0, \textit{ otherwise}; \end{cases}, \text{ value 1 means gene } \mathbf{X} \text{ is active;} \quad (2)$$

2.2.2 Mutational Forks Formalism for Combinatorial Perturbations

Here, we introduce the notion of *transitional confidence* (TC) to characterize transitions from the junction gene to any of the shoulder genes, based on *simultaneous mutations in several genes* in a fork. Consider an example of combinatorial perturbation, where the shoulder genes **A** and **B** are mutated, while the junction gene **J** is not mutated:

$$TC(T_{\mathbf{JA}}(\mathbf{J_{AB}}) \mid M(\mathbf{J}) = 0, M(\mathbf{A}) = 1, M(\mathbf{B}) = 1) = \quad (3)$$

$$\frac{P(M(\mathbf{J}) = 0, M(\mathbf{A}) = 1, M(\mathbf{B}) = 1 \mid T_{\mathbf{JA}}(\mathbf{J_{AB}})) * P(\mathrm{T}_{\mathbf{JA}}(\mathbf{J_{AB}}))}{P(M(\mathbf{J}) = 0, M(\mathbf{A}) = 1, M(\mathbf{B}) = 1)}$$

- $P(M(\mathbf{J}) = 0, M(\mathbf{A}) = 1, M(\mathbf{B}) = 1) \mid T_{\mathbf{JA}}(\mathbf{J_{AB}}))$ – number of patient cases, where shoulder genes **A** and **B** are mutated and activated, while junction gene **J** is not mutated, divided by all cases in the cohort
- $P(M(\mathbf{J}) = 0, M(\mathbf{A}) = 1, M(\mathbf{B}) = 1)$ – number of patient cases, where shoulder genes **A** and **B** are mutated, and junction gene **J** is not mutated regardless of the activation status of gene **A** and **B**, divided by all cases in the cohort
- $P(T_{\mathbf{JA}}(\mathbf{J_{AB}}))$ – frequency of topological linear motif from junction gene **J** to shoulder gene **A**

Following the logic in Equation 3, we can infer a perturbation pattern for any combination of mutations in the mutational fork. Please note that unlike previously described pathway analysis tools[14-18], our approach uses literature findings about pathway interactions as a prior to predict the flow of signal transduction. The TC concept is fundamental in this paper, because it not only enables the identification of gene-to-gene perturbation patterns but is also essential for determining patient-specific flow through a particular mutational fork. Since the previous version of the method addressed only single mutational events in the inference of transitional confidences[19], we improved the current version by accounting for all possible combinations of mutations in forks and, therefore, captured a larger scope of pathway deregulation events for lung adenocarcinoma patients.

2.2.3 Data Acquisition and Preprocessing

The pathway data have been acquired from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database[20]. These pathway maps have been processed using a two-step curation approach.

The first step of curation consists of removing inconsistencies and errors (e.g. missing or erroneous gene-to-gene interactions) that were present in original KGML files. During this process, all gene entities and their interactions were verified using cited literature to ensure accurate data provenance.

The second step of curation consists of data formalization and data transformation procedures. The data formalization procedure affects gene-to-gene interactions and their directionality. To formalize gene-to-gene interactions, we have replaced simplistic interaction types in KEGG pathway maps by a set of more specific biological relationships (e.g. *binds_associates*, *phosphorylates_activates*, *phosphorylates_inhibits*, etc.). Overall, we have modeled 27 gene-to-gene interactions as in our previous work[21]. These interaction terms have been encoded as predicates in RDF triples. To formalize the directionality of interactions, signal transduction is considered to flow from the effectors to regulated gene entities, which correspondingly determine the *Start* and the *End* gene entities in a path. The data transformation procedure translates curated KGML files into traversable RDF *subject-predicate-object* triples.

Finally, we have integrated five different pathway maps, including the MAPK signaling pathway (hsa04010), Calcium signaling pathway (hsa04020), JAK-STAT signaling pathway (hsa04630), NSCLC signaling pathway (hsa05223), and PI3K-AKT signaling pathway (hsa04151) into the RDF knowledge network as per our previous work[22]. To address

computational complexity, we have introduced the notion of *Protein Family* to represent proteomic isoforms and enable traversal through singular protein family instances (Table 1). A protein family can take an "un-collapsed" form when a detailed analysis of a specific isoform is needed. All gene entities were linked to terms in standard gene and protein databases and ontologies such as GO[23] and Uniprot[24] through the Human Protein Reference Database (HPRD)[25].

Protein family	**Members**
RAS	HRAS, KRAS, NRAS
RAF	ARAF, BRAF, RAF1
PI3K	PI3KCA, PI3KCB, PI3KCD, PI3KR1, PI3KR2, PI3KR3
MAP2K	MAP2K1, MAP2K2
$MAPK_1$	MAPK1, MAPK3
AKT	AKT1, AKT2, AKT3
$MAPK_2$	MAPK8, MAPK9, MAPK10
CRK	CRK, CRKL
MAPK8IP	MAPK8IP1, MAPK8IP2

Table 1: A list of protein families.

The lung adenocarcinoma patients' profiles have been extracted from TCGA Pan-Cancer Atlas (TCGA-LUAD) at the cBioPortal platform[19]. Whole-exome sequencing (WES) has been performed on tumor and germ-line DNA with a mean coverage 97.6× and 95.8×, respectively. The mean somatic mutation rate across the lung adenocarcinoma cohort was 8.87 mutations per megabase (Mb) of DNA. The single-nucleotide mutation rate was 6.86 per Mb. Seventy-five percent of somatic mutations identified by WES were present in the RNA transcriptome when

the locus in question was expressed (minimum 5×). The data consists of 566 patients, each represented by 114 attributes.

The skin cutaneous melanoma patients' profiles have been extracted from TCGA Pan-Cancer Atlas (TCGA-SKCM) at the cBioPortal platform[19]. The melanoma cohort is focused on metastatic cases (11.6% regional skin cutaneous or subcutaneous metastatic tissue, 56.4% regional metastatic lymph node, 25.1% distant or unspecified metastatic tissue) because melanoma is most often discovered after it has metastasized[26]. WES was performed on paired tumor and germline normal genomic DNA, including primary and metastatic melanomas with a mean exon coverage of 87×, adequate for detecting a single-nucleotide variant at an allelic fraction of 0.3 with a power of 80%. The data consists of 448 patients, each represented by 114 attributes.

The following attributes were inputted to the mutational fork:

- number of cases, where gene X (or Y or Z) is mutated. This attribute requires information about specific alteration (e.g. Pro369Gln);
- number of cases, where gene X (or Y or Z) is mutated and activated. This attribute requires information about specific alteration (e.g. Pro369Gln) and the corresponding FPKM value;
- number of cases, where gene X (or Y or Z) is not mutated;
- topological frequency for transition from X to Y (or X to Z). This attribute requires information derived from gene-to-gene interaction pairs, for example, how frequent a specific gene pair appears in the set of extracted linear paths.

Whole exome sequencing is not required, and a combination of several candidate gene sequencings is acceptable. We have considered both non-synonymous and synonymous mutations for the downstream analysis because they have significant impact on translation or co-translational protein folding and play an important role in many diseases including cancer[27]. In addition, we have averaged the expression values across cancer patients for each gene to measure the activation effect from the mutational event.

2.2.4 Mutational Forks Extraction from RDF Knowledge Network

To reflect the direction of signal propagation, surface receptors have been encoded as *start genes*, while nuclear transcriptional factors have been encoded as *end gene*s[28]. For this study, we chose the EGFR gene as a start gene, which corresponds to the cellular receptor and the CCND1 gene as the end gene, which corresponds to the nuclear target. The objective behind selection of these genes was that activation of the epidermal growth factor receptor (EGFR)-tyrosine kinases is a key reason for lung adenocarcinoma progression[29]. On the other hand, deregulation of CCND1 has been implicated in the pathogenesis of lung adenocarcinoma and is associated with poor prognosis[30]. However, other genes relevant to lung adenocarcinoma pathogenesis can be selected as well, including multiple-input, multiple-output options. The k-shortest path graph algorithm has been applied to extract a set of *linear paths* from start to end genes. Finally, a custom script has identified frequently occurring linear paths and combined them into mutational forks in a form of *dual* and *triple* forks.

2.2.5 Identification of Patient-Specific Signaling Routes

We have performed three blocks of TC calculations for dual forks and four blocks of TC calculations for triple forks:

- Block 1: one mutational event occurred, either in the junction gene or one of the shoulder genes;
- Block 2: two mutational events occurred simultaneously, either in the junction gene and one of the shoulder genes or in two shoulder genes;
- Block 3: three mutational events occurred conjointly, either in the junction gene and two of the shoulder genes (in the case of dual and triple forks) or in three shoulder genes (in the case of triple forks only). A special case of this scenario has occurred when all genes are mutated in the triple fork.

We then calculated the TC for each block using Equation 3.

2.3 Results

After setting the EGFR gene as a start gene and the CCND1 gene as an end gene, we identified 178 unique linear paths between these gene entities. The length of the linear paths varied between eight and eleven gene entities. By combining frequently occurring linear paths, we have extracted eleven mutational forks from the RDF knowledge network using a custom script. Out of these eleven, six mutational forks were selected as having distinct frequencies for each transition (Figure 2). Gene entities corresponding to protein families are highlighted in bold.

2.3.1 Case Study 1: Lung Adenocarcinoma

Table 2 presents an example of TC calculations for the KRAS mutational fork in lung adenocarcinoma patients. Detailed TC calculations for the rest of the mutational forks have been provided in the Supplementary Materials.

Transition	Description	TC
	Block 1	
KRAS-BRAF	$TC(T_{\mathbf{JA}}(\mathbf{J}_{\mathbf{AB}}) \mid M(\mathbf{J}) = 1, M(\mathbf{A}) = 0, M(\mathbf{B}) = 0)$	0.0416
	$TC(T_{\mathbf{JA}}(\mathbf{J}_{\mathbf{AB}}) \mid M(\mathbf{J}) = 0, M(\mathbf{A}) = 1, M(\mathbf{B}) = 0)$	0.0448
	$TC(T_{\mathbf{JA}}(\mathbf{J}_{\mathbf{AB}}) \mid M(\mathbf{J}) = 0, M(\mathbf{A}) = 0, M(\mathbf{B}) = 1)$	0.0442
KRAS-PIK3CA	$TC(T_{\mathbf{JB}}(\mathbf{J}_{\mathbf{AB}}) \mid M(\mathbf{J}) = 1, M(\mathbf{A}) = 0, M(\mathbf{B}) = 0)$	0.8807
	$TC(T_{\mathbf{JB}}(\mathbf{J}_{\mathbf{AB}}) \mid M(\mathbf{J}) = 0, M(\mathbf{A}) = 1, M(\mathbf{B}) = 0)$	**0.9539**
	$TC(T_{\mathbf{JB}}(\mathbf{J}_{\mathbf{AB}}) \mid M(\mathbf{J}) = 0, M(\mathbf{A}) = 0, M(\mathbf{B}) = 1)$	0.942
	Block 2	
KRAS-BRAF	$TC(T_{\mathbf{JA}}(\mathbf{J}_{\mathbf{AB}}) \mid M(\mathbf{J}) = 1, M(\mathbf{A}) = 1, M(\mathbf{B}) = 0)$	0.0389
	$TC(T_{\mathbf{JA}}(\mathbf{J}_{\mathbf{AB}}) \mid M(\mathbf{J}) = 1, M(\mathbf{A}) = 0, M(\mathbf{B}) = 1)$	0.0383
	$TC(T_{\mathbf{JA}}(\mathbf{J}_{\mathbf{AB}}) \mid M(\mathbf{J}) = 0, M(\mathbf{A}) = 1, M(\mathbf{B}) = 1)$	0.0435
KRAS-PIK3CA	$TC(T_{\mathbf{JB}}(\mathbf{J}_{\mathbf{AB}}) \mid M(\mathbf{J}) = 1, M(\mathbf{A}) = 1, M(\mathbf{B}) = 0)$	0.8294
	$TC(T_{\mathbf{JB}}(\mathbf{J}_{\mathbf{AB}}) \mid M(\mathbf{J}) = 1, M(\mathbf{A}) = 0, M(\mathbf{B}) = 1)$	0.8217
	$TC(T_{\mathbf{JB}}(\mathbf{J}_{\mathbf{AB}}) \mid M(\mathbf{J}) = 0, M(\mathbf{A}) = 1, M(\mathbf{B}) = 1)$	**0.9260**
	Block 3	
KRAS-BRAF	$TC(T_{\mathbf{JA}}(\mathbf{J}_{\mathbf{AB}}) \mid M(\mathbf{J}) = 1, M(\mathbf{A}) = 1, M(\mathbf{B}) = 1)$	0.0231
KRAS-PIK3CA	$TC(T_{\mathbf{JB}}(\mathbf{J}_{\mathbf{AB}}) \mid M(\mathbf{J}) = 1, M(\mathbf{A}) = 1, M(\mathbf{B}) = 1)$	0.4913

Table 2: Transitional confidence calculations for KRAS mutational fork in lung adenocarcinoma patients.

KRAS mutational fork

Interestingly, the mutational status of the BRAF gene increase the possibility of transitioning by the KRAS-PIK3CA path (see Block 1 in Table 2). While BRAF and PIK3CA mutations occur simultaneously, the possibility of transitioning by the KRAS-PIK3CA path is higher than in the other two cases (see Block 2 in Table 2). When junction gene KRAS is not mutated, a transition by the KRAS-PIK3CA path is more likely. However, when junction gene KRAS is mutated, the differences between various TC values are more pronounced.

Using functional classification of network-attacking mutations (NAMs), the KRAS mutational fork represents an example of dysregulation of network dynamics, specifically the "node inactivation" scenario. Inactivating mutations on the essential residues of BRAF kinase led to the interruption of information flow through this node. Thus, the phosphorylation of subsequent BRAF substrate – MAP2K1 protein is disabled, which in turn allows PIK3CA protein kinase through generation of secondary messenger PIP3 to recruit AKT1 and PDPK1 proteins to the cellular membrane, thus activating the PI3K-AKT axis[30, 31].

AKT1 mutational fork

Notably, mutation in the MAP2K4 gene alone increases the chance of transitioning by the AKT1-MAP2K4 route. While MAP2K4 and RAF1 mutations occur conjointly, the possibility of transitioning by the AKT1-MAP2K4 path is higher than in other cases. Similarly, conjoint mutations in the AKT1, RAF1 and MAP2K4 genes increase the chance of transitioning by the AKT1-MAP2K4 route, comparing to other combinations of mutations. Overall, the mutational status of junction gene AKT1 does not profoundly impact transition by the AKT1-MAP2K4 route.

Following functional classification of NAMs, the AKT1 mutational fork represents an example of dysregulation of network dynamics, specifically the “node activation” scenario. It has been shown that AKT1 regulates the storage of glucose in the form of glycogen by phosphorylating GSK3B at the 'Ser-9' residue, resulting in inhibition of its kinase activity[32]. In addition, AKT1 phosphorylates RAF1 at the 'Ser-259' residue and negatively regulates its activity as well[33]. However, mutations in the MAP2K4 gene allow avoiding of negative regulation from AKT1 phosphorylation, thus enabling direct activation of downstream targets, such as MAPK8, MAPK9 and MAPK10[34].

MAP2K4 mutational fork

Interestingly, mutations in junction gene MAP2K4 can alone increase chances of transitioning by the MAP2K4-MAPK8 path. Conjoint mutations in MAP2K4 and CRK genes increase the possibility of transitioning by the MAP2K4-MAPK8 path. Furthermore, conjoint mutations in the MAP2K4, MAPK8 and CRK genes increase chances of transitioning by the MAP2K4-MAPK8 path as well. Overall, the mutational status of junction gene MAP2K4 can profoundly impact transition by the MAP2K4-MAPK8 path.

Based on the functional classification of NAMs, the MAP2K4 mutational fork represents an example of network rewiring, specifically the “downstream rewiring” scenario. Evidently, MAP2K4 mutations shift the signaling network structure to the MAP2K4-MAPK8 route by “rewiring” downstream interactions with MAPK8IP3, MAPK8 and CRK proteins. Such network behavior can be explained by drifts in the peptide specificity upon mutation of the determinants of specificity (DoS) in the MAP2K4 kinase. Therefore, activated MAPK8 phosphorylates a

number of transcription factors, primarily components of AP-1 such as JUN, JDP2 and ATF2, and thus regulates AP-1 transcriptional activity[35].

MAPK8IP1 mutational fork

Notably, a mutation in the junction gene MAPK8IP1 alone increases chances of transitioning by the MAPK8IP1-MAP3K11 path. While MAPK8IP1 and MAP2K7 mutations occur conjointly, a transition by the MAPK8IP1-MAP3K11 path is more probable. A similar TC value has been obtained when the MAPK8IP1 and MAP3K11 genes were mutated simultaneously (TC = 0.403 in both cases). Furthermore, conjoint mutations in the MAPK8IP1, MAP3K11 and MAP2K7 genes increase a chance of transitioning by the MAPK8IP1-MAP3K11 path as well. Overall, the mutational status of junction gene MAPK8IP1 profoundly impact the transition of the MAPK8IP1-MAP3K11 path. Unfortunately, since MAPK8IP1 is not a phosphoprotein, we were unable to derive functional implications of its single-nucleotide mutations for corresponding junction genes.

MAP3K11 mutational fork

Interestingly, mutation in the shoulder gene MAP2K7 alone increases the chance of transitioning by the MAP3K11-MAPK8IP3 path. A similar TC value has been obtained, while the junction gene MAP3K11 mutated alone (TC = 0.5080). Moreover, when both genes are mutated, the chance of transitioning by the MAP3K11-MAPK8IP3 path is higher than the remaining cases. Overall, the mutational status of the junction gene MAP3K11 profoundly impacts the transition by the MAP3K11-MAPK8IP3 path.

Using the functional classification of NAMs, the MAP3K11 mutational fork represents an example of dysregulation of network dynamics, specifically the “node inactivation” scenario.

Inactivating mutations on the essential residues of the MAP2K7 kinase led to an interruption of information flow through this node. Thus, the direct activation of the MAP2K7 downstream substrates, such as MAPK8, MAPK9 and MAPK10, is disabled, which in turn allows the MAPK8IP3 scaffold protein to selectively mediate JNK signaling by aggregating specific components of the MAPK cascade to form a functional JNK signaling module[35].

MAPK8 mutational fork

Notably, mutation in junction gene MAPK8 alone increases chances of transitioning by the MAPK8-MAPK8IP1 path. Furthermore, while junction gene MAPK8 conjointly mutated with shoulder gene MAPK8IP1, the possibility of the MAPK8-MAPK8IP1 path is the highest. Overall, the mutational status of junction gene MAPK8 impacts the transition by the MAPK8-MAPK8IP1 route.

From the functional classification of NAMs, the MAPK8 mutational fork represents an example of network rewiring, specifically the "downstream rewiring" scenario. Evidently, MAPK8 mutations shift the signaling network structure to the MAPK8-MAPK8IP1 route by "rewiring" downstream interactions with MAPK8IP1 and MAPK8IP3 proteins. It has been shown that cytoplasmic MAPK8IP1 causes inhibition of JNK-regulated activity by retaining JNK in the cytoplasm and inhibiting JNK phosphorylation of c-Jun[36].

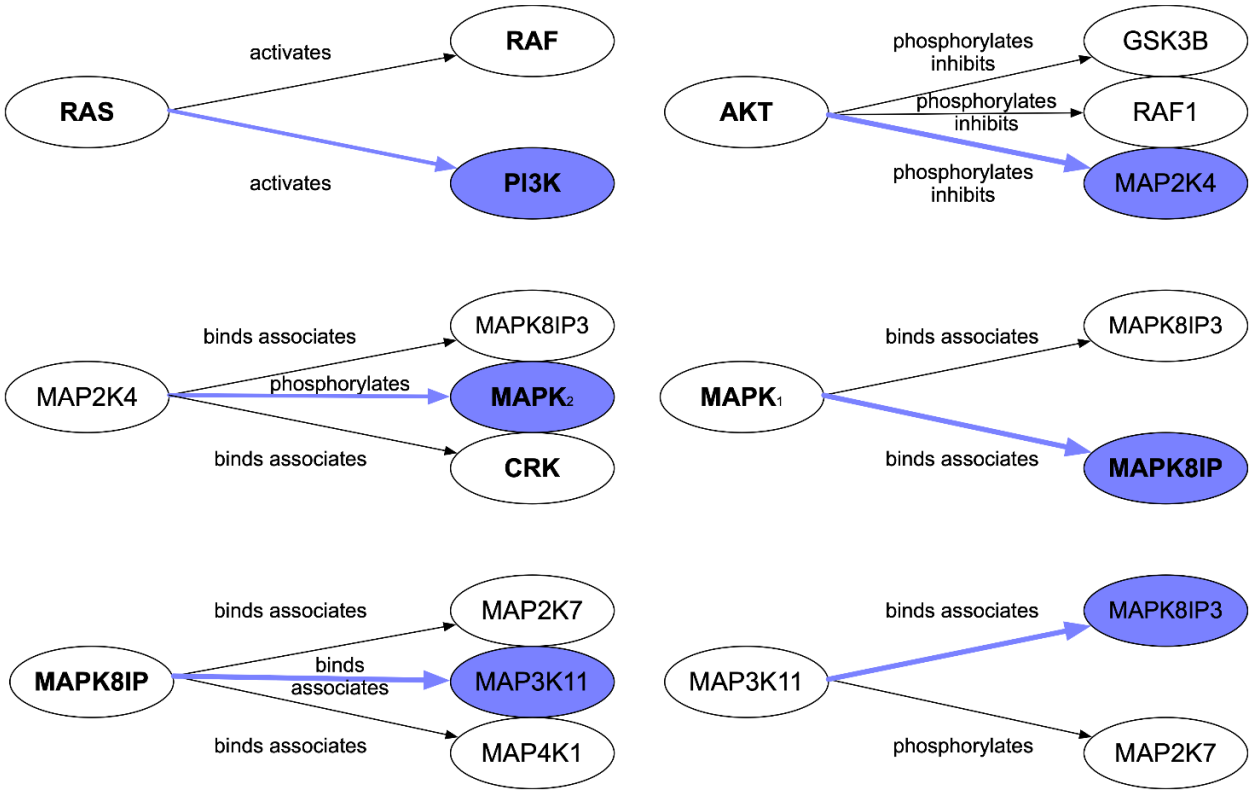

Figure 2: Extracted mutational forks for lung adenocarcinoma patients. The high likelihood route highlighted in purple, along with the corresponding shoulder gene. Protein families are highlighted in bold (see Table 1). MAPK1 refers to either MAPK1 or MAPK3; MAPK2 refers to MAPK8, MAPK9, or MAPK10.

2.3.2 Case Study 2: Skin Cutaneous Melanoma

KRAS mutational fork

Interestingly, the mutational status of the junction gene KRAS increase the possibility of transitioning by the KRAS-PIK3CA path. While KRAS and PIK3CA mutations occur simultaneously, the possibility of transitioning by the KRAS-PIK3CA path is higher than in the other two cases. These results differ from the KRAS mutational fork in lung adenocarcinoma, where mutations in the shoulder gene BRAF plays more profound role in transitioning by the KRAS-PIK3CA axis.

AKT1 mutational fork

Notably, mutation in the shoulder gene MAP2K4 alone increases the chance of transitioning by the AKT1-MAP2K4 route. While MAP2K4 and GSK3B mutations occur conjointly, the possibility of transitioning by the AKT1-MAP2K4 path is higher than in other cases. Similarly, conjoint mutations in the AKT1, GSK3B and MAP2K4 genes increase the chance of transitioning by the AKT1-MAP2K4 route, comparing to other combinations of mutations. Unlike the AKT1 mutational fork in lung adenocarcinoma, where combination of MAP2K4 and RAF1 mutations increase the possibility of transitioning by the AKT1-MAP2K4 route, mutations in the GSK3B gene are more likely to define the flow of signal towards AKT1-MAP2K4 axis.

MAP2K4 mutational fork

Interestingly, mutations in the junction gene MAP2K4 can alone increase chances of transitioning by the MAP2K4-MAPK8 path. Conjoint mutations in MAP2K4 and CRK genes, as well as in another pair of genes – MAPK8 and CRK – increase the possibility of transitioning by

the MAP2K4-MAPK8 path. Furthermore, conjoint mutations in MAP2K4, MAPK8 and CRK genes increase chances of transitioning by the MAP2K4-MAPK8 path as well. Overall, the mutational status of junction gene MAP2K4 can profoundly impact transition by the MAP2K4-MAPK8 path. In comparison to the MAP2K4 mutational fork in lung adenocarcinoma, CRK gene is widely implicated into transitioning by MAP2K4-MAPK8 axis.

MAPK8IP1 mutational fork

Notably, a mutation in the shoulder gene MAP2K7 alone increases chances of transitioning by the MAPK8IP1-MAP3K11 path. While MAPK8IP1 and MAP2K7 mutations occur conjointly, a transition by the MAPK8IP1-MAP3K11 path is more probable. Furthermore, conjoint mutations in the MAPK8IP1, MAP2K7 and MAP4K1 genes increase a chance of transitioning by the MAPK8IP1-MAP3K11 path as well. These results differ from the MAPK8IP1 mutational fork in lung adenocarcinoma, where junction gene MAPK8IP1 plays more profound role in transitioning by the MAPK8IP1-MAP3K11 axis.

MAP3K11 mutational fork

Interestingly, mutation in the shoulder gene MAP2K7 alone increases the chance of transitioning by the MAP3K11- MAP2K7 path. Moreover, when MAP3K11 and MAPK8IP3 genes are mutated concomitantly, the chance of transitioning by the MAP3K11- MAP2K7 path is higher than the remaining cases. Unlike the MAP3K11 mutational fork in lung adenocarcinoma, where signal flow through the MAP3K11-MAPK8IP3 route, the predominant transition for the MAP3K11 mutational fork using melanoma dataset is MAP3K11-MAP2K7 axis.

MAPK8 mutational fork

Notably, mutation in junction gene MAPK8 alone increases chances of transitioning by the MAPK8-MAPK8IP1 path. Furthermore, while junction gene MAPK8 conjointly mutated with shoulder gene MAPK8IP1, the possibility of the MAPK8-MAPK8IP1 path is the highest. Overall, the mutational status of junction gene MAPK8 impacts the transition by the MAPK8-MAPK8IP1 route. These findings are consistent with the MAPK8 mutational fork in lung adenocarcinoma.

2.4 Discussion

In this paper, we have presented the Mutational Forks formalism for inferring patient-specific flow of signal transduction based on multiple single-nucleotide alterations. The fundamental idea behind the Mutational Forks formalism is an application of a Bayesian "flip" principle to infer the most probable transitions of signaling flow given the mutational profile of the case. In this setup, the marginal *prior probability* of a transition is computed based on its frequency as the transition appears in literature-derived knowledge bases such as KEGG. This quantity, i.e. the *initial belief* of the transition, reflects a chance of the transition from a *canonical* view of the disease, without accounting for any patient-specific data. The *normalized conditional probability* of a mutation given the transition is then computed and represents the *support* the mutation provides to the transition. In this scheme, the Bayesian "flip" enables the assessment of a chance of the transition based on mutations in a specific patient cohort. In other words, the initial probability of the transition is adjusted by the mutational profile of the cohort. The resulting quantity, i.e. the *posterior probability of the transition given mutation*s, may not strictly follow the mathematics of the probabilistic systems. Therefore, we call it the *Transitional*

Confidence (TC) and compute it for a set of network junctions, which we call *Mutational Forks* (MF), in context of this study.

The proposed method provides a more comprehensive view of pathway deregulation in cancer compared to existing pathway tools[13-18]. For example, our approach uses literature findings about pathway interactions as a prior to predict the flow of signal transduction, rather than approximating it from the topology of the interactome. In addition, the method expends semantics of gene-to-gene interaction terms to 27 interactions, including *binds_associates*, *phosphorylates_activates*, *phosphorylates_inhibits*, etc. Finally, our method helps to pinpoint mutated genes that are directly involved in transitioning by a specific pathway route. For example, mutations in BRAF gene increases the possibility of signal flow through the KRAS-PIK3CA axis in lung adenocarcinoma, while mutations in KRAS gene have more profound impact in transitioning by the same route in melanoma cohort.

The power of the pathway-based approach is that it may provide clues about the possible mechanisms underlying the differences in observed cancer patients' survival. Mutational forks may be useful for suggesting therapeutic targets or to select the most appropriate patients for clinical trials. For example, the KRAS G12C mutation is a well-known marker of particular forms of lung cancer that are treatable by the drug AMG 510 (Amgen)[37]. However, some patients with the KRAS G12C mutation have tumors that are resistant to treatment. By considering combinations of mutational events incorporated in the inference of transitional confidence, our approach can identify a cohort of patients with the KRAS G12C mutation that is functionally inactive at the pathway level due to a cumulative silencing effect from other mutations. Thus, the treatment-resistant nature of these patients' can be explained by their

pathway mutational status and when this pattern is observed, different therapy can be chosen. This approach provides us with the opportunity to avoid toxicity and wasting patient time with predictably ineffective therapy, thus directing clinicians to potentially more efficacious treatment upfront, which is the primary goal of precision health and patient-centered medicine.

We chose lung adenocarcinoma and skin cutaneous melanoma as case studies for several reasons. Both malignancies have high mutational burden[38, 39] and it is still unclear which path will be activated given a combination of such mutational events if, for example, a cancer patient has both BRAF and PIK3CA mutations. By knowing the exact signaling path affected by the cumulative impact of multiple mutations, clinicians would be able to tackle treatment resistance more effectively. Although our experiments had been conducted on genomics profiles derived from lung adenocarcinoma and melanoma patients, our method can be relevant to another types of cancer as well.

There are some limitations to our method. Due to the algorithmic design, our approach employs a limited number of curated signaling pathways relevant to cancer theranostics. Furthermore, our method accounts for the functional impact of single-nucleotide mutations, considering the sporadic nature of investigated malignancies. Due to the small scope of our study, we were unable to utilize additional database resources, such as SynMICdb[40], however, we will employ this database in our future work . Since our intention was to show a holistic picture of pathway deregulation for cancer patients given multiple mutational events, we did not differentiate among various phenotypical populations of cancer patients. Finally, the clonal evolution of solid tumors is beyond the scope of this article.

From a data standpoint, Leek et al. argued that differences in sample collection and processing across multiple centers could lead to batch effects, or variation in the data due to technical factors that mask relevant biological variation. For example, while considering TCGA data, indel calls were more sensitive to technical artifacts than single-nucleotide variant (SNV) calls[41]. Therefore, to avoid false positive associations we considered only single-nucleotide mutations in this paper.

Chapter 3 Identification of Immuno-Targeted Combination Therapies Using Explanatory Subgroup Discovery for Cancer Patients with EGFR Wild-Type Gene

3.1 Introduction

Immunotherapy represents one of the most promising therapeutic approaches in cancer[42]. Immune checkpoint inhibitors (ICIs) can lead to long-term remission and improved survival in patients with locally advanced/metastatic cancer[43, 44]. However, almost 50% of cancer patients that are eligible for treatment with ICIs will not respond[45, 46]. This is particularly true in patients without targetable mutations[47]. Conventional cancer treatments, such as radiation therapy[48], cytotoxic chemotherapy[49, 50], and targeted therapy[51, 52], have immunomodulatory effects along with direct tumor cell-killing activities. Their clinical utility in combination with ICIs potentially create synergetic effects with improved and durable clinical response[53, 54]. Therefore, there is a growing interest in searching for predictive biomarkers of therapeutic response and identifying therapeutic targets that could extend the benefits of ICIs.

Recent studies show that various factors contribute to the response to ICIs[55]. Tumor mutational burden (TMB) has emerged as one of the most crucial factors to determine the efficacy of ICIs[56, 57]. TMB quantitatively assesses the number of mutations per megabase (muts/Mb). High TMB values are indication of better survival in cancer patients receiving immunotherapy[58]. However, the effectiveness of immunotherapy has been demonstrated in some patients with low TMB, while unfavorable outcomes have been observed in a significant number of patients with a high TMB[59, 60]. It is also unclear how patients that harbor specific genetic alterations would respond to immunotherapy. For example, anti-PD-1 monotherapy was unable to improve survival outcomes for non-small cell lung cancer (NSCLC) patients with EGFR

mutations, even in patients with high PD-L1 expression[61]. Meanwhile, the ATLANTIC trial dispute this observation and highlights the benefits of ICIs for EGFR mutated tumors[62]. Therefore, the question of how to effectively utilize thousands of cell-intrinsic and extrinsic characteristics to identify patient subpopulations that will benefit from combinatorial treatments with ICIs remains unanswered.

Over the past decades, Subgroup Discovery (SD) methods have been used to find homogenous subpopulations of patients that share common genetic profiles and may respond similarly to therapeutic regimens[63, 64]. Existing SD approaches can be classified into two major categories: statistical methods and data mining methods. Statistical methods include regression analysis[65], clustering techniques[66] and latent class analysis (LCA)[67]. For example, LCA is a finite mixture model that aims to uncover unobserved groups within population. However, this technique does not automatically determine the number of latent classes and produce solutions that heavily rely upon expert knowledge, which significantly limits the capability to discover novel subgroups in large heterogenous cancer datasets. Data mining methods for SD are comprised of two major categories based on the search strategy for potential candidates – heuristic approaches and exhaustive approaches[68]. For example, CN2-SD[69] is a heuristic algorithm that searches for the statistically "most interesting" subgroups that are as large as possible and have the most unusual distributional characteristics with respect to the property of interest. However, CN2-SD suffers from the standard scaling problem that appear in the evaluation of large datasets, including heterogenous cancer datasets.

In this work, we extend our Subgroup Discovery algorithm[70] to predict immuno-targeted combination therapies for EGFR WT patient subpopulations. We focused on EGFR WT

subgroups because EGFR-TKIs is widely used for first-line treatment of patients with EGFR sensitizing mutations, leading to longer progression-free survival (PFS)[71, 72]. However, beyond first line, especially for EGFR WT tumors, the role of EGFR-TKIs is elusive. We employed a proportional odds model to identify significant drug targets and corresponding compounds that increase the likelihood of stable disease versus progressive disease in cancer patients that are EGFR WT. This approach will help to better select responders for immuno-targeted combination therapies and improve health outcomes for cancer patients with no targetable mutations.

3.2 Materials and Methods

Our informatic framework consists of two modules: Subgroup Discovery and Immuno-Targeted Combination Therapies Discovery. The Subgroup Discovery module identifies homogenous patient subgroups based on both phenotypic and genotypic parameters and explain the differences between these subgroups using gene expression patterns. The Immuno-Targeted Combination Therapies Discovery module predicts potential drug targets and corresponding compounds for uses in combination therapies with immunotherapy for cancer patients with no targetable mutations. The details of modules are described below.

3.2.1 Data Mapping

The TCGA dataset from PanCancer Atlas[73] consists of 1952 cancer patients. We have focused on four different cancer types: (1) Head and Neck Squamous Carcinoma (HNSC), n = 515; (2) Lung Adenocarcinoma (LUAD), n = 510; Lung Squamous Carcinoma (LUSC), n = 484 and Skin Cutaneous Melanoma (SKCM), n = 443. We chose these cancer types, because in several clinical trials involving patients with advanced lung cancer, skin cutaneous melanoma

and other solid tumors they demonstrated significant response rates to nivolumab and pembrolizumab, both monoclonal IgG4 antibodies against PD-1[74].

The input data consists of phenotypic and genotypic variables for a disease population. The phenotypic variables divide disease population into subgroups, while genotypic variables describe the main characteristics (patterns) of these subgroups. Eleven phenotypic variables were chosen for this study, including demographic data (e.g., Age, Gender), clinical-pathologic data (e.g., Tumor Type, Tumor Mutational Burden (TMB)), treatment history (e.g., Administration of Radiotherapy), and behavioral data (e.g., Smoking Status). As a part of the human-in-the-loop process, a physician panel specialized in the care of cancer patients selected the phenotypic variables to be included in the analysis. Many of these phenotypic variables were categorical. For example, the Smoking Status can be categorized as follows: (1) Never Smoker, (2) Current Smoker and (3) Former Smoker. We ensure all continuous variables in the dataset were categorized based on clinical literature or experience. To deal with missing data, we have excluded patients with too many missing values in both phenotypic and genotypic variables and used 'NA' as a new category to represent missing values. However, 'NA' variables are never used to form subgroups.

The genotypic data in this study is z-score transformed values for 730 immune-related genes (including PD-1 and PD-L1 genes)[75] and 40 housekeeping genes between normal and tumor tissues. The categorization of genotypic variables is based z-score transformed values, where downregulated genes correspond to the range (-Inf, -2), not significantly expressed genes correspond to the range (-2, 2) and upregulated genes correspond to the range (2, Inf). Therefore, each gene represents a genotypic variable.

3.2.2 Subgroup Discovery

The main goal of the Subgroup Discovery module is to identify homogeneous patient subpopulations. A set of common phenotypic and genotypic parameters specifies each unique subgroup. For example, Females with TMB defined as High (more than 10 mutations per Mb) with SKCM could be considered as a subgroup. The unique subgroup is reevaluated each time a phenotypic variable is included. The statistical significance of each subgroup is defined by genotypic patterns that distinguish this subgroup from the rest of the outer population.

The Subgroup Discovery module includes three levels: *Path Expansion*, *Floating Subgroup Selection*, and *Inclusion*/*Exclusion* criteria. This method differs from decision tree algorithm in a way that allows the same patient to be a member in multiple unique subgroups. For example, LUSC patients could be members of (LUSC, Former Smoker) subgroup and (LUSC, Stage IIIA) subgroup. The key objective of subgroup stratification process is to determine a large number of existing subgroups based on phenotypic parameters, where most patients in that subgroup share unique genotypic patterns. This method is not greedy, because the algorithm is tracking the top potential subgroups based on local optimal selection[76].

By iterating over various paths in the search space, the *Floating Subgroup Selection* traverses' multiple paths. A phenotypic variable with the highest contrast score against the outer population (e.g. *Tumor Type = LUAD* on Figure 3) forms a base subgroup via *Path Expansion* process. On the first inclusion step, it adds a new variable, e.g. *EGFR = WT*, to the base subgroup (*Tumor Type = LUAD* AND *EGFR = WT*). To eliminate less significant inclusion steps, the exclusion function is used after each inclusion step. For example, when at the third inclusion step the subgroup is (*Tumor Type = LUAD* AND *EGFR = WT* AND *Smoker = Former* AND *Gender =*

Female), the exclusion function will eliminate the less significant move (*Smoker* = *Former*) from the existing subgroup if the newly created subgroup (*Tumor Type* = *LUAD* AND *EGFR* = *WT* AND *Gender* = *Female*) has higher contrast score. When the algorithm reaches the subgroup with the highest contrast score, the exploratory search will be terminated.

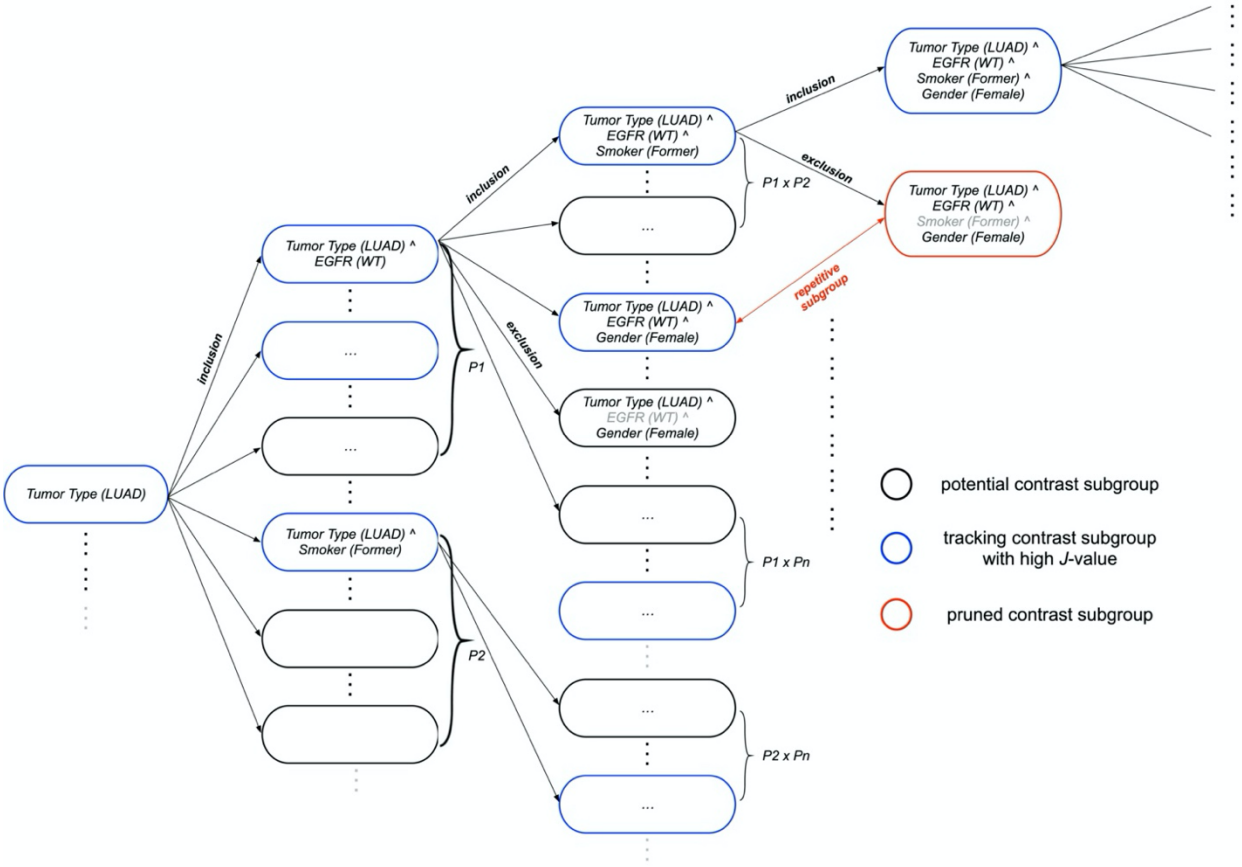

Figure 3: The Subgroup Discovery module. The floating and path expansion processes initiate multiple distinct starting points on various computational nodes. Based on the contrast score, the node would be added or eliminated for each point. Each point portrays a potential subgroup. By applying contrast pattern mining, each candidate subgroup is scored against the outer population.

The algorithm identifies the genotypic patterns that are frequently occur within each candidate subgroup but rare in the rest of the population. To evaluate the frequency for a specific

genotypic pattern in the homogenous subgroup the support metric[77] (Equation 4) is utilized. The growth rate metric[78] (Equation 5) is employed to evaluate the contrast of the pattern in the selected subgroup:

$$Support(p, D) = \frac{|\langle D, p \rangle|}{|D|} \tag{4}$$

where p is a pattern, $|<D, p>|$ is the number of patients with specific genotypic pattern, and $|D|$ is the total number of patients in the collection.

$$Growth(cp, S_{G1}, S_{G2}) = \frac{\text{Max}|s_1, s_2|}{\text{Min}|s_1, s_2|} \tag{5}$$

where cp is the contrast pattern, S_{G1} is the focus subgroup, S_{G2} is the outer population, s_1 is the support of cp in the focus subgroup, s_2 is the support of cp in the outer population.

We have employed a J-value[79] (Equation 6) to calculate the contrast for all patterns in each subgroup.

$$J\text{-value} = \frac{T * J_{org} + M * \bar{J}_{avr}}{T + M} \tag{6}$$

where T is a parameter related to the population size preference that depends on the type of the disease under study and, whether it is a rare disease or not, based on the concept of the Bayesian average[80].

There are six major parameters for Subgroup Discovery module: *min_support_proportion*, *max_depth*, *max_breadth*, *min_improvement_significance*, *max_checks* and *max_pop_complexity*. The *min_support_proportion* describes the minimum proportion of the subgroup rows that a pattern must be present in and was set to 0.05. The *max_depth* parameter represents the maximum

pattern length and was adjusted to five. The *max_breadth* stands for maximum number of candidates at each search level and was modified to 50. The *min_improvement_significance* reflects the p-value needed to add any new element to a pattern and was set to 0.05. The *max_checks* parameter describes a hard cap on the max number of patterns checked and was adjusted to 1000. Finally, the *max_pop_complexity* describes the maximum number of phenotypic variables that can define a subgroup and was modified to three.

3.2.3 Immuno-Targeted Combination Therapies Discovery

Our goal was to find significant drug targets and their corresponding compounds that can be used in immuno-targeted combination therapies for cancer patients with no targetable mutations. Specifically, we were interested in identifying significant drug targets that increase the likelihood of stable disease versus progressive disease in four cancer types: HNSC, LUAD, LUSC and SKCM. For this purpose, we used the proportional odds model[81] on the Prat A. et al. dataset[82]. The dataset consists of 65 cancer patients, 15 phenotypic and 770 genotypic variables. The outcome variable was categorical and had four levels: partial response, complete response, progressive disease, and stable disease. The continuous covariates were represented as genes with corresponding normalized gene expression values. We used MASS R package[83] to fit the proportional odds model with logit link function. We computed p-values via two-tailed z-test to identify significant predictors of response to immunotherapy.

3.3 Results

3.3.1 Identification of Candidate Subgroups for Immuno-Targeted Combination Therapies

The overview of our informatic pipeline is presented in Figure 4. The Subgroup Discovery module generated 9887 subgroups. To reduce the search space, we focused on

subgroups that had Tumor Type (HNSC, LUAD, LUSC or SKCM) as one of the phenotypic variables. The filtering procedure resulted in 1207 subgroups. To further focus on more specific subgroups, we selected those that had at least three phenotypic variables. In total, there were 1129 subgroups used in our computational experiments.

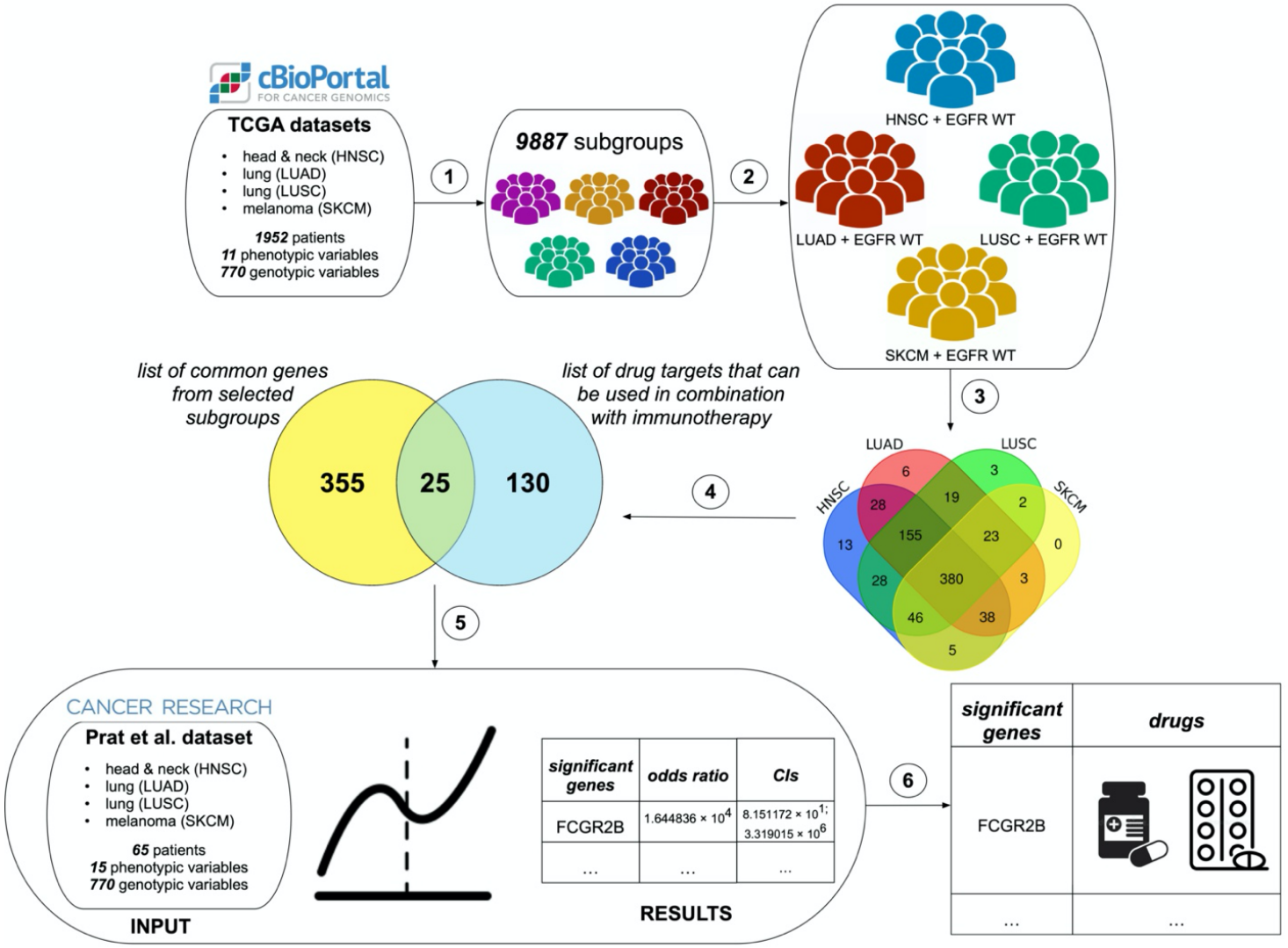

Figure 4: The Immuno-Targeted Combination Therapies Discovery module. (1) We run Subgroup Discovery algorithm on the combined TCGA dataset. The algorithm outputted 9887 subgroups (2) We retained EGFR WT subgroups that covered at least

20% of the initial dataset. (3) For these subgroups, we determined common DE genes (n = 380). (4) We then mapped these 380 genes to the list of 155 drug targets. Thus, we identified 25 targets that can be used in immuno-targeted combination therapies. (5) We determined significant drug targets that increase the likelihood of SD versus PD using proportional odds model. (6) We matched significant genes to drugs that can we used in immuno-targeted combination therapies using the Drug Gene Interaction Database (DGIdb).

First, we investigated subgroup coverage in the whole dataset of 1948 cancer patients. The subgroup coverage ranged from 1.23% to 25.97%. Therefore, we decided to target subgroups or unions of subgroups that covered at least 20% of the whole dataset, which resulted in eleven subgroup unions. Secondly, we narrowed down our search to four subgroup unions that had both Tumor Type (HNSC, LUAD, LUSC or SKCM) and were EGFR WT. We hypothesized that cancer patients from these four subgroup unions were unlikely to produce a favorable outcome on treatment with EGFR inhibitors due to lack of targetable mutations in this gene. However, these patients may have other mutations that can be sensitive to FDA-approved targeted treatments. Therefore, it is important to identify these additional drug targets to improve patients' outcomes. In addition, immunotherapy alone may not produce durable responses for these patients. In the ongoing phase III KEYNOTE-042 trial of patients with treatment-naïve, advanced, EGFR/ALK WT NSCLC and at least 1% tumor PD-L1 expression, there was no statistically significant PFS benefit among patients receiving Pembrolizumab compared with those receiving chemotherapy, except for those with the highest level of PD-L1 expression[84]. However, EGFR WT cancer patients may benefit from compounds that have been used for immuno-targeted combination therapies, as has previously been shown in EGFR WT NSCLC patients[85]. Thirdly, we identified common differentially expressed (DE) genes for these four subgroup unions. We started with identifying unique genotypic patterns for each subgroup union and then determined unique elements from these genotypic patterns. The intersection of unique genes from four subgroup unions produces 380 DE genes. The summary for each subgroup union is presented in Table 3.

Subgroup union	# patients	% of patients in cancer type	% of patients in whole dataset	# subgroups size three	# unique genotypic patterns	# unique DE genes
HNSC	500	97.08	25.66	16	8448	693
LUSC	466	96.68	23.92	16	10005	652
LUAD	444	87.06	22.79	20	10216	656
SKCM	407	92.29	20.89	15	7428	497

Table 3: A subgroup summary for selected subgroup unions.

There were also some unique features for each subgroup union. Our analysis revealed that HNSC EGFR WT subgroups consisted of patients with low TMB. The role of TMB in predicting the outcome of immunotherapy for advanced HNSCC remains unclear. For example, one study had showed that immunotherapy was more effective in metastatic HNSCC patients with high TMB, and the median OS of these patients was 2.5 times as long as that of the patients with low TMB (25 vs. 10 months)[86]. However, data from over 10 000 patient tumors included in The Cancer Genome Atlas failed to support application of high TMB as a biomarker for treatment with ICIs in all solid cancer types[87]. The LUAD EGFR WT subgroups had substantial number of patients with BRAF WT gene. The clinical impact of BRAF mutational subtypes in lung adenocarcinoma remains to be defined. For example, the data from two large German lung cancer centers showed that patients with BRAF-mutated NSCLC have an inferior prognosis, which is not determined by BRAF mutation functional class. However, in contrast to NSCLC with other driver mutations, BRAF-mutated NSCLC exhibit high susceptibility to immune checkpoint inhibitors[88]. The LUSC union consisted of patients who are former smokers. For example, LUSC patients with history of smoking and high TMB had longer PFS after treatment with chemoimmunotherapy or anti-angiogenesis therapy[89]. There were no distinctive phenotypic

features for SKCM union. These identified unique features should be considered when designing combinatorial treatment regimens for EGFR WT cancer patients.

3.3.2 Drug Target Prediction for EGFR Wild-Type Subgroups

We examined the Drug Gene Interaction Database (DGIdb)[90] to map therapeutic compounds that can be used in immuno-targeted combination therapies. This search produced 155 potential drug targets and 36 targeted treatments. We then mapped the 380 common DE genes from the subgroup unions to the list of 155 drug targets. In this analysis we identified 25 targets and 16 compounds that may be used in combination with ICIs.

The Immuno-Targeted Combination Therapies Discovery module identified six significant drug targets: CDH5, FCGR2B, IGF1R, ITK, JAK2 and KIT (Supplementary Table 1). The estimates in the output are given in units of ordered logits or ordered log odds. For example, for FCGR2B gene the likelihood of stable disease versus partial response, complete response and progressive disease increases by 9.70 on the log odds scale. We also calculated odd ratios and confidence intervals (Supplementary Table 2). These can be obtained either by profiling the likelihood function or by using the standard errors and assuming a normal distribution (Table 4). For example, for every unit increase in expression of KIT gene the odds of having stable disease is multiplied 3.57 times, with all other variables constant. We had also matched significant drug targets to their corresponding therapeutic compounds along with cancer type (Table 5)

	OR	2.5 %	97.5 %	p-values
CDH5	2.875372×10^{-2}	8.965756×10^{-4}	9.221493×10^{-1}	0.0448808290
FCGR2B	1.644836×10^{4}	8.151472×10^{1}	3.319015×10^{6}	0.0003368373
IGF1R	2.238631×10^{3}	1.110336×10^{0}	4.513470×10^{6}	0.0469308638
ITK	1.274047×10^{-1}	2.807962×10^{-2}	5.780693×10^{-1}	0.0075795133
JAK2	1.047600×10^{-5}	1.525653×10^{-10}	7.193414×10^{-1}	0.0435977859
KIT	3.571475×10^{0}	1.543342×10^{0}	8.264815×10^{1}	0.0029426309

Table 4: The odd ratios and confidence intervals for significant genes.

Drug target	Compound	Cancer type
CDH5	Ruxolitinib, Lenalidomide	Lung Squamous Carcinoma, Skin Cutaneous Melanoma
FCGR2B	Bevacizumab, Cetuximab, Trastuzumab	Lung Adenocarcinoma, Head and Neck Squamous Carcinoma
IGF1R	Gefitinib	Lung Adenocarcinoma
ITK	Pazopanib, Ibrutinib	Skin Cutaneous Melanoma
JAK2	Bortezomib	Lung Adenocarcinoma
KIT	Axitinib, Cabozantinib, Pazopanib, Sunitinib	Head and Neck Squamous Carcinoma

Table 5: Significant genes and potential compounds that can be used in immuno-targeted combination therapies.

To address the question about potential drug interaction effects and toxicity of identified combinations, we employed the Drug Interaction Checker from the DrugBank database[91]. Based on our search, the administration of PD-1 or PD-L1 inhibitors in combination with FCGR2B targeted therapies can increase the risk of adverse effects. Specifically, these therapeutic combinations carry a risk of immunogenicity which can produce a wide array of adverse effects the most serious of which include anaphylaxis and serum sickness-type reactions[92]. A few studies suggest the use of multiple immunoglobulin agents is relatively safe and may be more effective than monotherapy for certain conditions[93, 94].

From clinicaltrials.gov we were able to identify clinical trials for predicted immuno-targeted combination treatments (Table 6):

Trial ID	Treatment combination	Condition	Results/Conclusions	References
MC1534, NCT03012230	Pembrolizumab and Ruxolitinib	Stage IV Triple Negative Breast Cancer	Estimated Primary Completion Date: April 1, 2023	[95]
BTCRC-HEM15-027, NCT03681561	Nivolumab and Ruxolitinib	Relapsed or Refractory Classical Hodgkin Lymphoma	Therapy combining Ruxolitinib with Nivolumab is well tolerated and yield encouragingly high remission rates and durable responses in patients who had all failed previous check-point inhibitor (CPI)	[96]
NCI-2020-08331, NCT04609046	Nivolumab and Lenalidomide	Primary CNS Lymphoma	Estimated Primary Completion Date: May 31, 2024	[97]
MK-3475-021/KEYNOTE-021, NCT02039674	Pembrolizumab and Gefitinib	Non-small Cell Lung Cancer	First-line Pembrolizumab plus Pemetrexed-Carboplatin continued to show improved response and survival versus chemotherapy alone in advanced NSCLC, with durable clinical benefit in patients who completed 2 years of therapy. No new safety signals were observed with longer follow-up.	[98]
MC1577, NCT03021460	Pembrolizumab and Ibrutinib	Stage III-IV Melanoma	Estimated Primary Completion Date: February 1, 2023	[99]
OSU-18015, NCT03525925	Nivolumab and Ibrutinib	Metastatic Solid Tumors	Estimated Primary Completion Date: December 31, 2021	[100]

020-008, NCT04265872	Pembrolizumab and Bortezomib	Metastatic Triple Negative Breast Cancer	Estimated Primary Completion Date: October 1, 2023	[101]
PANDORA 001, NCT04995016	Pembrolizumab and Axitinib	Locally Advanced Non-metastatic Clear Cell Renal Cell Carcinoma	Estimated Primary Completion Date: August 20, 2022	[102]
Winship4234-17, NCT03468218	Pembrolizumab and Cabozantinib	Head and Neck Squamous Cell Cancer	This phase II trial of Pembrolizumab + Cabozantinib met its primary endpoint of overall response rate (ORR). The regimen is well-tolerated with very encouraging clinical activity in relapsed metastatic HNSCC and warrants further exploration in this disease.	[103]
CheckMate 016, NCT01472081	Nivolumab, Pazopanib and Sunitinib	Metastatic Renal Cell Carcinoma	The addition of standard doses of sunitinib or pazopanib to nivolumab resulted in a high incidence of high-grade toxicities limiting future development of either combination regimen.	[104]
16-2300.cc, NCT03149822	Pembrolizumab and Cabozantinib	Metastatic Renal Cell Carcinoma	This study of the combination of Pembrolizumab and Cabozantinib met the primary endpoint of ORR. Benefit was seen in first- and subsequent-line therapy. The safety profile was manageable.	[105]

Table 6: Clinical trials for predicted immuno-targeted combinations.

Some of the clinical trials from Table 6 are still ongoing, therefore they don't have information about primary and secondary outcomes. However, we were able to retrieve such information from other clinical trials, such as NCT02039674 and NCT01472081. In NCT02039674 trial[98], Cohort F received Pembrolizumab (Pembro 2 mg/kg) via IV infusion on Day 1 of each 3-week cycle plus Gefitinib (G 250 mg) via oral tablet once a day on every day of each 3-week cycle. Overall, seven participants were treated with this regiment, but none of these patients were able to complete the treatment due to death (n = 1), excluded medication (n = 4) or withdrawal from the study (n = 2).

In NCT01472081 study[104], Arm S was treated with combination of Nivolumab and Sunitinib and Arm P was treated with combination of Nivolumab and Pazopanib. Arm S received two different doses of Nivolumab: 2 mg/kg (SUN + NIV2, n = 7) and 5 mg/kg (SUN + NIV5, n = 26). Arm P received only 2 mg/kg dose of Nivolumab (PAZ + NIV2, n = 20). The all-causality severe adverse effects (SAEs) of any grade were observed in 42% of SUN + NIV2 cohort, 61% of SUN + NIV5 cohort and 65% of PAZ + NIV2 cohort. The drug-related SAEs of any grade were observed in 28% of SUN + NIV2 cohort, 46% of SUN + NIV5 cohort and 10% of PAZ + NIV2 cohort. The all-cause adverse effects (AEs) that led to discontinuation of treatment were observed in 42% of SUN + NIV2 cohort, 38% of SUN + NIV5 cohort and 25% of PAZ + NIV2 cohort. These results suggest that higher concentration of Nivolumab in combination with Sunitinib (SUN + NIV5 cohort) lead to higher chance of SAEs. By analyzing secondary responses, SUN + NIV5 cohort achieved lower rate of partial response to the treatment at 42%, in comparison to 71% in SUN + NIV2 cohort. Therefore, lower concentrations of Nivolumab in combination with Sunitinib lead not only to less adverse effects, but also to

better patients' outcomes. Ideally, more trial settings are needed to study synergetic effect of two or more drugs. Unfortunately, we cannot assess the synergetic effects of these treatments from the NCT01472081 study because there were no cohorts that undergone solely Nivolumab or Sunitinib treatments.

3.4 Discussion

One of the aims of this study was to find phenotypic commonalities for EGFR WT cancer patient subgroup unions that might be helpful in selecting responders for immuno-targeted combination therapies. Unlike existing "black-box" models, our approach interprets a combination of phenotypic characteristics by assessing the frequency of significant genomic patterns in investigated subgroups. For this purpose, we examined the third phenotypic feature for each subgroup that was a part of the unions and covered at least 15% of the dataset. For all four unions – HNSC, LUAD, LUSC and SKCM - progression free survival (PFS) more than 6 months was a common feature. Interestingly, in Impower150, clinical trial patients with EGFR WT metastatic lung adenocarcinoma had longer PFS on combinatorial treatment with *Atezolizumab* (immunotherapy that targets PD-L1), *Bevacizumab* (targeted therapy against VEGF-A), *Paclitaxel* and *Carboplatin* (both are chemotherapies)[106]. Finally, in the KEYNOTE-048 clinical trial *Pembrolizumab* (immunotherapy that targets PD-L1) with chemotherapy improved overall survival versus *Cetuximab* (EGFR targeted therapy) with chemotherapy in patients with head and neck squamous cell carcinoma[107]. This demonstrates that combinatorial treatments with anti-PD-1 or anti-PD-L1 inhibitors with targeted therapy substantially prolonged PFS in cancer patients with EGFR WT gene.

Analysis of the dataset from Prat A., et al. revealed that three out of six significant drug targets – FCGR2B, IGF1R and KIT – substantially increased the odds of having stable disease versus progressive disease in EGFR WT cancer patients. The importance of this finding is further supported by the fact there is an ongoing clinical trial of *BI-1206*, a monoclonal antibody to FCGR2B, in combination with *Rituximab* (chemotherapy that targets CD20) in patients with indolent B-cell non-Hodgkin lymphoma that has relapsed or is refractory to *Rituximab*[108]. From another study, LUAD patients with high plasma levels of IGF-1 or high IGF-1R expression in tumors were associated with resistance to anti-PD-1-programmed death-ligand 1 immunotherapy, which support the need for clinical evaluation of IGF-1 modulators in combination with PD-1 blockade[109]. Finally, there is an ongoing clinical trial of *Ipilimumab* (immunotherapy that targets CTLA-4) and *Imatinib Mesylate* (KIT inhibitor) in treating patients with solid tumors that have spread to other places in the body or cannot be removed by surgery[110]. Therefore, these genes can be useful therapeutic targets for immuno-targeted combination therapies.

Based on our knowledge, there is no computational pipeline that can evaluate the synergetic effect from immuno-targeted combination therapies. Our evaluations based on primary and secondary outcomes from existing clinical trials. The limitation of these trials is that they are lacking the data on monotherapy effects. For example, NCT01472081 trial evaluates combinations of Nivolumab + Sunitinib and Nivolumab + Pazopanib. However, there is no information on treatment with Nivolumab, Sunitinib or Pazopanib alone. Therefore, better clinical trial design is required to objectively evaluate synergetic effect from immuno-targeted combination therapies.

Chapter 4 Immune-Related Gene Signatures to Predict the Effectiveness of Chemoimmunotherapy in Triple-Negative Breast Cancer Using Explanatory Subgroup Discovery

4.1 Introduction

Triple-negative breast cancer (TNBC) is an aggressive type of breast cancer that represents 10 to 20% of all breast cancer cases[111]. TNBC is characterized by the absence of detectable estrogen receptor (ER), progesterone receptor (PR) and epidermal growth factor receptor 2 (HER2)[112]. Depending on their response to initial chemotherapy, one in three TNBC patients will undergo tumor recurrence, which often occurs within the first three years of primary diagnosis[113]. Although numerous therapeutic agents have been evaluated for the treatment of early TNBC[114], only Olaparib had been approved for the adjuvant treatment of high-risk TNBC patients harboring germline BRCA1 or BRCA2 pathogenic variants[115]. However, the emergence of cancer immunotherapy may alter the paradigm in TNBC treatment.

TNBC, unlike other breast cancer subtypes, has high tumor mutational burden (TMB), which has been correlated with responsiveness to immune-checkpoint inhibitors (ICIs)[116]. Indeed, checkpoint inhibition with Pembrolizumab has been approved for advanced-stage, PD-L1 positive TNBC due to improved outcomes when combined with frontline chemotherapy in the KEYNOTE-355 trial[117]. Interestingly, ICIs are more effective in treating TNBC when given early in the course of the disease; this may be as a result of immune escape mechanisms emerging as the condition progresses[118]. More recently, results from the KEYNOTE-522 trial indicate that adding checkpoint inhibition in the early-stage setting does in fact improve long-term outcomes[119]. However, subgroup analyses in KEYNOTE-522 did not pinpoint any

biomarker strongly predicting the benefit of Pembrolizumab. For example, PD-L1 expression did not distinguish responders from non-responders in the early setting, with both PD-L1-negative and PD-L1-positive patients obtaining a benefit from Pembrolizumab addition. Moreover, the addition of immunotherapy increased adverse effects (AEs) beyond the toxicities of traditional chemotherapy[120]. The most common AEs discovered in KEYNOTE-522 were infusion reactions (18%), thyroid impairment (15.1%, hypothyroidism; 5.2% hyperthyroidism), skin toxicities (5.7%), pneumonitis (2.2%) and hepatitis (1.4%)[119]. In another study – IMPASSION131 trial – the combination of Paclitaxel with another PD-L1 inhibitor Atezolizumab failed to improve progression-free survival (PFS) or overall survival (OS) in TNBC patients[121]. These findings could be the outcome of imbalances in prognostic features or accidental discoveries in a relatively small trial. Therefore, the exploration of immune-related biomarkers is needed to optimally identify patients requiring the addition of ICIs to chemotherapy, including different thresholds of PD-L1 expression[122], the presence of tumor infiltrated lymphocytes (TILs)[123], TMB[124] and immune gene expression profiles.

In this work we determined homogenous TNBC subgroups based on both phenotypic and genotypic parameters using exploratory subgroup mining. We have also identified significant predictors that increase chances of partial remission in TNBC patients on chemoimmunotherapy treatment using multinomial regression model on TNBC scRNA-seq dataset. Lastly, we uncovered distinct immune cell populations (T-cells, B-cells, Myeloid, NK-cells) for TNBC patients with various treatment outcomes. We interpreted our results using biomedical knowledge, including findings from existing clinical trials, immunohistochemistry experiments and functional characterization of specific genes. The proposed informatics pipeline may assist

health care professionals in the selection of chemoimmunotherapy responders, as well as determine the underlying causes of drug resistance in TNBC patients at a single-cell level and resolution.

4.2 Materials and Methods

4.2.1 Data Mapping

In this study we employed two datasets. Each dataset consisted of multiple phenotypic (either categorical or continuous) and genotypic (continuous only) variables. Each categorical variable was labeled based on the National Comprehensive Cancer Network (NCCN) Guidelines in Oncology[125]. For example, relapse-free status was categorized as (1) disease-free or (2) recurred. Continuous variables were converted into categoric variables by grouping values into several categories. For example, normalized gene expression values were categorized as (1) downregulated, (2) upregulated, or (3) non-differentially expressed.

The first TNBC dataset comprised of 422 patients. These patients were selected from 24 breast cancer studies available at the cBioPortal platform[6]. The final dataset included breast cancer patients that lack expression of the estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2). This dataset consisted of 12 phenotypic variables, including clinical-pathologic data (age at diagnosis, menopausal status, tumor type, tumor stage, tumor cellularity, histologic grade, TMB), treatment regimen (chemotherapy, radiotherapy, hormone therapy) and survival status (overall survival status, relapse-free status). There were 1067 genotypic variables in the form of normalized gene expression values derived from human immunome (immune-related genes) and human kinome (protein kinase genes).

The second TNBC dataset consisted of scRNA-seq profiles for 22 TNBC patients that underwent chemotherapy (*Pactilaxel*) or chemoimmunotherapy treatment (*Paclitaxel* with *Atezolizumab*)[126]. For this study we selected six phenotypic variables, including information about treatment timeline (pre-, post-treatment, progression), tissue type (tumor or blood), tumor site (brain, breast, chest wall, liver, lymph nodes), treatment type (anti-PD-L1+chemotherapy or chemotherapy only), treatment response (partial response (PR), stable disease (SD), progressive disease (PD)) and cell cluster (T-cells, B-cells, Myeloid, NK-cells). We used the same genotypic variables as in the TNBC subgroup discovery dataset.

4.2.2 The Informatics Pipeline

Our informatics pipeline had three modules: (1) exploratory subgroup discovery, (2) inference module based on multinomial regression model and (3) immune cell populations discovery. Our goal was two-fold: (1) to identify significant genes from exploratory subgroup discovery that increase odds of having partial remission after anti-PD-L1+chemotherapy treatment and (2) to uncover unique immune cell populations for TNBC subgroups with various treatment outcomes.

The key objective of exploratory subgroup discovery module was to find a large number of existing subgroups based on phenotypic parameters (Module A on Figure 5), where most patients in that subgroup share distinctive genotypic patterns (Module B on Figure 5). Each genotypic pattern had been represented as a combination of differentially expressed genes. For example, the genotypic pattern may consist of three genes: upregulated *EGFR*, downregulated *MTOR* and upregulated *MAPK1* genes. The algorithm starts by choosing a single phenotypic variable (e.g. *Chemotherapy = Yes*) as a base subgroup, which has the most significant contrast

against the rest of the population. On the next inclusion step it adds a new phenotypic variable, e.g. *TMB = High*, to the previous subgroup to generate a more focused subgroup (*Chemotherapy = Yes* AND *TMB = High*). After each inclusion step, the exclusion step is employed to eliminate a less significant inclusion move by removing a variable from the "greedy" subgroup if the result of the exclusion process has improved performance. The exploratory search selects multiple paths that have equally relevant genotypic patterns within each subgroup. It will be terminated in each path when the algorithm gets into a most highly specific subgroup with the highest contrast score that cannot be improved further. The algorithm identified the genotypic patterns that were frequently occur within each candidate subgroup but rare in the rest of the population. To evaluate the frequency for a specific genotypic pattern in the homogenous subgroup the support[77] and growth rate metrics[78] were employed. We applied a *J*-value[79] to calculate the contrast for all patterns in each subgroup. For more details on the algorithm, please refer to our previous publication[70].

To find significant predictors of partial remission on anti-PD-L1+chemotherapy regimen, we employed a multinomial regression model (Figure 5C) on the scRNA-seq TNBC dataset. The outcome variable was categorical and represented as a combination of treatment response, treatment timeline and treatment type. For example, the level of outcome variable can be encoded as *SD-Post_treatment-Chemo* meaning that a fraction of TNBC patients achieved stable disease after treatment with chemotherapy only. Overall, there were ten levels of outcome variable. We set *PD-Post_treatment-Chemo* – progressive disease after chemotherapy – as a baseline for the model. The continuous covariates were represented as genes with corresponding normalized gene expression values identified as the top 10% of genotypic patterns in the

exploratory subgroup discovery stage. We used the multinom function from the nnet package[127] to estimate a multinomial logistic regression model. We computed p-values via two-tailed z-test to identify significant predictors of response to anti-PD-L1+chemotherapy treatment.

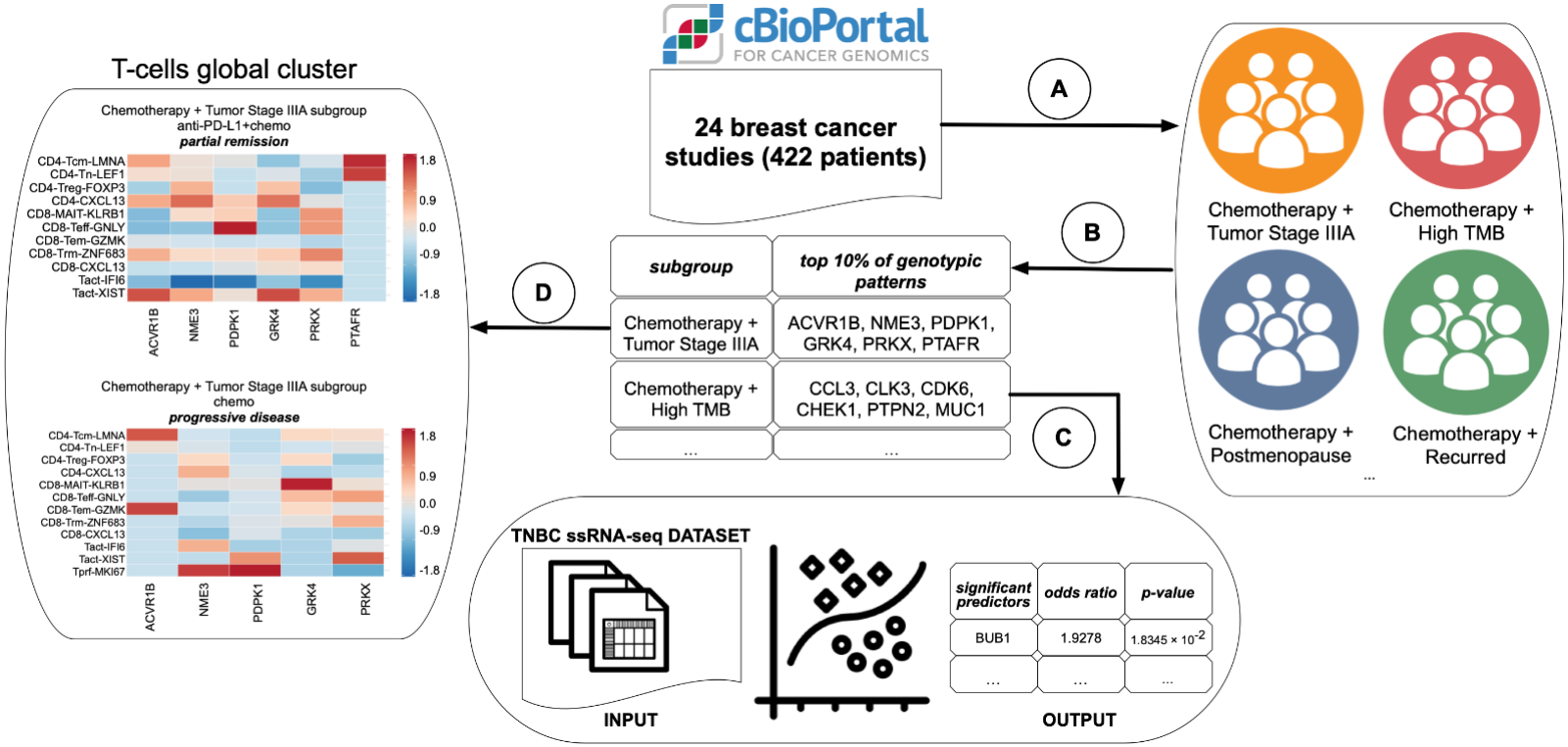

Figure 5: The informatics pipeline for deciphering the efficacy of chemoimmunotherapy. Modules (A) and (B) – the exploratory subgroup discovery process, Module (C) – the inference module based on multinomial regression model, Module (D) – immune cell populations discovery.

The immune cell populations discovery module (Figure 5D) determined distinct immune cell populations (T-cells, B-cells, Myeloid, NK-cells) for TNBC patients with various treatment outcomes. Each TNBC subgroup had two conditions: (1) an-ti-PD-L1+chemotherapy regimen, post treatment, PR and (2) chemotherapy, post treatment, PD. Using the top 10% of genotypic patterns from the exploratory mining stage as an input, we generated heatmap plots for each condition in every TNBC subgroup of interest. For example, NME3 gene was represented as a geometric mean of NME3 expression values in CD4-Tcm-LMNA cells[128]. Finally, we compared immune cell populations in these two conditions to identify mutually exclusive cell populations that were associated with either PR after anti-PD-L1+chemotherapy treatment or PD after chemotherapy treatment.

4.3 Results

4.3.1 Identification of Homogenous TNBC Subgroups

First, we ran the exploratory subgroup discovery algorithm on the TNBC dataset described in Section 4.2.1. The algorithm outputted 11944 subgroups. We focused on analysis of 460 subgroups, where TNBC patients underwent chemotherapy treatment. On the next step, we narrowed down the search to twelve homogenous subgroups based on TMB (low or high), relapse status (disease-free or recurred), tumor cellularity (high, low and moderate), menopausal status (pre- or post) and tumor stage (I, II and III). Since the lengths of genotypic patterns vary (up to 5 genes), we decided to make a union of top 10% of genotypic patterns for each subgroup of interest. Let's assume that each genotypic pattern is a set of elements, where each element is a unique differentially expressed gene (e.g. upregulated *MTOR* gene). The union would represent a set of a collection of genotypic patterns, where each element would not be repetitive. These

genotypic patterns were used as covariates for the multinomial regression model in the next section.

4.3.2 Significant Predictors of Partial Remission After Anti-PD-L1+Chemotherapy

The multinomial regression model on scRNA-seq TNBC dataset was able to identify significant predictors from the exploratory subgroup discovery results that increase odds of having PR after anti-PD-L1+chemotherapy treatment versus PD after chemotherapy regimen (Table 7).

Subgroups	Predictors	Coefficients	p-values	Odds Ratio
Chemotherapy (Yes) TMB (High)	ACVR1B	0.4506	0.0136	1.5692
	PDPK1	0.1271	0.0169	1.1355
Chemotherapy (Yes) TMB (Low)	CLK3	0.1402	0.0017	1.1505
	TAOK2	0.6603	8.7493×10^{-7}	1.9354
Chemotherapy (Yes) Relapse Status (Disease Free)	CDK9	0.4152	2.1356×10^{-10}	1.5146
	CFP	0.2368	0.0438	1.2673
	VRK3	0.1591	0.0011	1.1725
Chemotherapy (Yes) Relapse Status (Recurred)	BUB1	0.6563	0.0183	1.9278
	BAZ1B	0.3529	0	1.4232
Chemotherapy (Yes) Tumor Cellularity (High)	PDIK1L	0.6405	1.3243×10^{-5}	1.8974
	KIR2DL4	1.8931	0.0094	6.6403
	MAPK3	0.8886	0	2.4319
	STK24	0.2842	1.2534×10^{-8}	1.3287
Chemotherapy (Yes) Tumor Cellularity (Low)	IFI16	0.0665	0.0205	1.0688
	CSK	0.2397	0	1.2709

	TAP2	0.2165	0.0185	1.2417
	TIGIT	0.5160	0	1.6754
Chemotherapy (Yes) Menopausal Status (Pre)	CCR6	0.1307	0.0023	1.1396
	BCL10	0.1222	0.0064	1.1300
	PRKCA	0.9982	0	2.7135
	EPHB6	0.8559	0	2.3536
	IFNAR2	0.4546	6.4298×10^{-6}	1.5756
Chemotherapy (Yes) Menopausal Status (Post)	PDIK1L	0.4952	0.0003	1.6409
	RPS6KA5	0.2836	6.7817×10^{-9}	1.3279
	IKZF2	0.4714	5.9285×10^{-14}	1.6023
Chemotherapy (Yes) Tumor Stage (I)	RIOK3	0.1683	2.4615×10^{-5}	1.1833
Chemotherapy (Yes) Tumor Stage (II)	PRKCA	1.0213	0	2.7770
Chemotherapy (Yes) Tumor Stage (III)	IFIH1	0.3366	1.2977×10^{-5}	0.1773
	CDKL5	0.7786	0.0066	2.1785

Table 7: Significant predictors that increase odds of partial remission after anti-PD-L1 + chemotherapy

Next, we highlight the importance of identified phenotypic features from Table 7 for TNBC patient outcomes. Using literature, high-TMB TNBC cases may derive particular benefit from ICIs in combination with chemotherapy (IMpassion130 trial)[129] or even ICIs alone (KEYNOTE-119 trial)[130]. The IMpassion130 trial included 149 early-stage TNBC receiving immunotherapy agent Durvalumab in combination with chemotherapy or chemotherapy alone. Median TMB was significantly higher in patients with pathologic complete response (pCR) (median 1.87 versus 1.39, p = 0.005), and odds ratios for pCR per mut/MB were 2.06 (95% CI 1.33–3.20) among all patients, 1.77 (95% CI 1.00–3.13) in the Durvalumab arm, and 2.82 (95%

CI 1.21–6.54) in the chemotherapy arm. Although the association between pCR and TMB was present in both arms, it was stronger in the cohort treated with chemotherapy alone. The KEYNOTE-119 trial compared Pembrolizumab monotherapy to chemotherapy in 622 patients with pre-treated, metastatic TNBC. Of the 253/601 treated patients with available TMB data (n = 132 Pembrolizumab, n = 121 chemotherapy arm), 26 patients (10.3%) were TMB-high (TMB ≥ 10 mut/MB). There was a positive association between TMB and clinical response to Pembrolizumab (ORR p = 0.154, PSF p = 0.014, OS p = 0.018) but not to chemotherapy (ORR p = 0.114, PFS p = 0.478, OS p = 0.906). In TMB-high cases, ORR and hazard ratio (HR) for OS also suggested a trend towards increased benefit with Pembrolizumab versus chemotherapy. Although suggestive for the clinical utility of TMB in TNBC, this study was limited by the small sample size and low number of TMB-high cases.

In terms of relapse status, one study hypothesized that rapid versus late relapse in TNBC is also defined by distinct clinical and genomic features of primary tumors[131]. Relative to 'no relapse' group (nrTNBC), both 'rapid relapse' (rrTNBC) and 'late relapse' (lrTNBC) groups had significantly lower immune signatures and immune signatures were highly correlated to anti-tumor CD8 T-cell, M1 macrophage, and gamma-delta T-cell CIBERSORT inferred immune subsets. Intriguingly, lrTNBCs were enriched for luminal signatures. There was no difference in TMB or percent genome altered across groups.

In connection to menopausal status, TNBC was frequently seen in postmenopausal patients[132]. Premenopausal patients with TNBC frequently had an overexpression of the p53 protein, a significantly higher Ki-67 index value, and a higher nuclear grade. A multivariate

analysis revealed that menopausal status, nodal status, and tumor size were significant factors for disease-free survival (DFS) in TNBC cases.

We had also discovered novel phenotypic features in TNBC subgroups, such as tumor cellularity and tumor stage. The evaluation of tumor cellularity, defined as the percentage of invasive tumor comprised of tumor cells, may represent an informative histologic measure of the differential response of TNBC to chemoimmunotherapy. The tumor staging system is commonly used to categorize the extent of a malignant disease in a given patient at various time points at or after diagnosis. Tumor staging plays a critical role in optimizing cancer treatment for better patient care. Therefore, these features can be important in the design and analysis of intervention studies, including randomized clinical trials, to better assess their prognostic utility for TNBC patients.

4.3.3 Differences in Immune Cell Populations for Discovered TNBC Subgroups

This section described immune cell populations that were discovered in scRNA-seq TNBC data based on genotypic patterns from exploratory mining stage. We interpreted our results using biomedical knowledge, including findings from existing clinical trials, immunohistochemistry experiments and functional characterization of specific genes. The summary of our findings is presented in Table 8.

Condition	T-cells	B-cells	Myeloid cells	NK-cells
anti-PD-L1 post treatment partial remission	CD4-Tn-LEF1 CD4-CXCL13	-	-	ILC3-AREG ILC3-IL7R
chemo post treatment	Tact-IFI6	pB-IGHG1	cDC1-CLEC9A macro-CFD	ILC1-VCAM1

progressive disease	macro-FOLR2
	macro-MKI67
	macro-SPP1
	macro-TUBA1B
	mono-FCN1
	mono-S100A89
	mono-SMIM25

Table 8: Immune cell populations that are linked to the specific TNBC outcome determined by our informatics pipeline

T-cells global cluster

The proliferative MKI67$^+$ T-cells (Tprf-MKI67) were exclusively present in TNBC patients with PD after chemotherapy. Interestingly, the expression of MKI67 gene was significantly correlated with lymph node metastases, tumor invasion and adverse survival outcome in TNBC[133]. In addition, high Ki-67 immunohistochemical expression levels in distant metastatic lesions were independently associated with poorer overall survival outcomes after biopsy of recurrence lesion in breast cancer patients (hazard ratio 2.307; 95% confidence interval 1.207–4.407, p-value = 0.011)[134]. Therefore, MKI67 could be an important biomarker of predictive and prognostic value in TNBC.

CD4-Tn-LEF1 and CD4-CXCL13 T-cells were linked to PR after anti-PD-L1+chemotherapy treatment. Importantly, these CD4$^+$ T-cells express very high amounts of PD-1 and other co-stimulatory and inhibitory receptors. Therefore, they provide essential help to B-cells for potent antibody responses and their tumor tissue presence is often correlated with a better outcome in several solid tumor entities[135]. Furthermore, immunohistochemistry staining

revealed the presence of CD4-CXCL13 T-cells in TNBC tumors responsive to chemoimmunotherapy[126, 136]. In addition to CXCL13$^+$ T-cells, naïve LEF1$^+$ T-cells (Tn-LEF1) were also linked to a favorable response to both anti-PD-L1+chemotherapy and chemotherapy. In a recent study, a four-gene signature – HLF, CXCL13, SULT1E1 and GBP1 – predicted the extent of lymphocytic infiltration after neoadjuvant therapy in TNBC. This signature was associated with favorable outcome, adding novel prognostic information for this aggressive breast cancer subtype[120].

The activated IFI6$^+$ T-cells (Tact-IFI6) were linked to PD after chemotherapy. The upregulation of mitochondrial antiapoptotic protein IFI6 was associated with poor distance metastasis-free survival and suggested to facilitate breast cancer metastasis by modulating mitochondrial ROS production[137]. Therefore, interrupting mitochondrial functions of IFI6 could potentially improve clinical outcomes in breast cancer patients.

B-cells global cluster

The MKI67$^+$ follicular B-cells (Bfoc-MKI67), NEIL1$^+$ follicular B-cells (Bfoc-NEIL1) and MKI67$^+$ memory B-cells (Bmem-MKI67) were exclusively present in TNBC patients with PR after anti-PD-L1+chemotherapy treatment. Follicular B-cells was the most significant B cell subtype associated with favorable clinical outcomes for TCGA patients with breast cancer[138]. The naïve B-cells, memory B-cells and follicular B-cells were mainly enriched in tumors responsive to chemoimmunotherapy but not in tumors responsive to chemotherapy treatment[126]. In regard with Bfoc-NEIL1 cell population, NEIL1 implicated in repair of oxidative damage associated with DNA replication or transcription[139]. Reduction in NEIL1 expression was associated with a poorer outcome in patients with breast invasive carcinoma[140]. Hence, NEIL1

could be a promising biomarker for TNBC patients that consider chemoimmunotherapy treatment.

Plasma IGHG1$^+$ B-cells (pB-IGHG1) were linked to PD after chemotherapy treatment. In TNBC, the expression of IGHG1 offered the most significant increases in prognostic value compared to classical clinicopathological parameters[141]. Intriguingly, IGHG1 expression in B-cells and plasma cells could be linked with tumor cell proliferation and immune evasion in breast cancer[142]. These data may imply that B cells or plasma cells could assume pro-tumoral roles under certain conditions; however, the factors driving the emergence of this putative pathologic phenotype and the roles played by B cells and plasma cells in these circumstances have yet to be discovered.

Myeloid cells global cluster

The MMP9$^+$ macrophages (macro-MMP9) were exclusively present in TNBC patients with PR after anti-PD-L1+chemotherapy treatment. In general, MMPs have a complex role in cancer progression and may exert both pro- and antitumorigenic activities[143]. Although MMP expression generally has been associated with tumor progression in several cancer forms including breast cancer[144], clinical trials with broad-spectrum MMP inhibitors have failed, and in some cases, patients treated with the inhibitors even showed significantly poorer survival than patients receiving placebo[145]. Indeed, the overexpression of MMP9 results in increased generation of antiangiogenic fragments, decreased angiogenesis, and therapeutic effects of established breast cancer[146]. In another study, gene transfer of MMP-9 to established breast cancer tumors induced tumor regression via increased neutrophil infiltration and an activation of

tumor-associated macrophages (TAMs) into antitumorigenic properties[147]. Therefore, MMP9 can serve as a biomarker for predicting tumor regression in TNBC.

The macro-CCL2, macro-CX3CR1, macro-IFI27, macro-IGFBP7, macro-IL1B9, macro-MGP and macro-SLC40A1 cells were exclusively present in TNBC patients achieving PD after chemotherapy. Interestingly, CCL2 expression in breast carcinomas was highly associated with macrophage infiltration, and its expression was correlated with poor prognosis in breast cancer patients[148]. In another study, chemokine receptor CX3CR1 showed a role in angiogenic macrophage survival in the tumor microenvironment contributing to tumor metastasis[149]. In a similar fashion, IFI27 overexpression was shown to impair the tamoxifen-induced apoptosis in breast cancer cells[150]. Finally, IL1B signaling contributed to breast cancer metastasis by enhancing tumor cell motility and inhibiting cell proliferation[151]. These findings highlight the importance of CCL2, CX3CR1, IFI27 and IL1B expressed in macrophages in progression of TNBC.

The CLEC9A$^+$ dendritic cells (cDC1-CLEC9A), macro-CFD, macro-FOLR2, macro-MKI67, macro-SPP1, macro-TUBA1B, FCN1$^+$ monocytes (mono-FCN1), mono-S100A89 and mono-SMIM25 cells were linked to PD after chemotherapy treatment. Notably, CFD functioned as an enhancer of tumor proliferation and cancer stem cell properties in breast cancers[152]. In another study, SPP1-associated macrophages in the tumor-adipose microenvironment facilitate breast cancer progression[153]. Interestingly, S100A8/A9, which are calcium-binding proteins that are secreted primarily by granulocytes and monocytes, may be associated with the loss of estrogen receptor, and may be involved in the poor prognosis of Her2$^+$/basal-like subtypes of

breast cancer. Therefore, myeloid cell populations expressing CFD, SPP1 and S100A89 might be crucial biomarkers of poor treatment response in TNBC.

NK-cells global cluster

The CNOT2^{+} group 2 innate lymphoid cells (ILC2-CNOT2) were exclusively present in TNBC patients achieving PR after anti-PD-L1+chemotherapy treatment. Indeed, ILC2s involved in both anti-tumor and pro-tumoral immunity in a variety of human cancers[154]. In terms of pro-tumoral immunity, ILC2s interact with tumor microenvironment (TME) and tumor cells by recruiting and amplifying cells of type 2 inflammation which favor tumor growth and metastasis[155]. Regarding anti-tumor immunity, ILC2s interact with TME and tumor cells by recruiting and activating eosinophils[156], macrophages with M1 profile and other cells of type 1 response (CD8^{+} T-cells) and molecules (CXCL1L/CXCL2L) promoting the lysis/apoptosis of tumor cells[157].

The ZNF683^{+} group 1 innate lymphoid cells (ILC1-ZNF683) cells were exclusively present in TNBC patients achieving PD after chemotherapy. Interestingly, ILC1 cells appear to play an active role in inhibiting the antitumoral immune response, prompting to evaluate the differential tumor infiltration of ILC1 cells in patients to optimize the harnessing of immunity in cancer therapies[158]. However, the role of ZNF683 gene remains elusive.

ILC3-AREG and ILC3-IL7R cells were linked to PR after anti-PD-L1+chemotherapy treatment. It had been shown that ILC3-IL7R could predict a favorable response to both treatment regimens, indicating its potential role in effective antitumor immunity[126]. In contrary, ILC1-VCAM1 cells were linked to PD after chemotherapy treatment. Recent studies had shown that vascular cell adhesion molecule-1 (VCAM1) was aberrantly expressed in breast cancer cells

and mediates prometastatic tumor-stromal interactions[159]. Therefore, AREG$^+$, IL7R$^+$ and VCAM1$^+$ innate lymphoid cells can help determined prognosis for breast cancer patients.

4.4 Discussion

One of the aims of this study was to find clinical predictors among TNBC subgroups that might be helpful in selecting responders for chemoimmunotherapy. Unlike existing "black-box" models, our approach interpreted a combination of phenotypic characteristics by assessing the frequency of significant genomic patterns in the investigated subgroups. For example, menopausal status was a significant factor in DFS of TNBC patients[132]. However, another clinical predictor – high TMB – had shortcomings as a marker of therapeutic effects in chemoimmunotherapy. For example, IMpassion130 trial[129] demonstrated that association with favorable outcome was stronger in the chemotherapy arm rather than in the chemoimmunotherapy arm. Conversely, KEYNOTE-119 trial[130] indicated positive association between TMB and clinical response to immunotherapy, but not to chemotherapy. Therefore, further studies are needed to better understand the role of TMB in efficacy of currently available chemoimmunotherapy combinations.

The analysis of the TNBC scRNA-seq data revealed distinct immune cell populations that are linked to either PR after anti-PD-L1+chemotherapy or PD after chemotherapy only. In terms of T-cells, CD4-Tn-LEF1 and CD4-CXCL13 T-cells were linked to PR after anti-PD-L1+chemotherapy treatment, while Tact-IFI6 T-cells were linked to PD after chemotherapy. The naïve B-cells, memory B-cells and follicular B-cells were mainly enriched in tumors responsive to chemoimmunotherapy but not in tumors responsive to chemotherapy treatment. The MMP9$^+$ macrophages (macro-MMP9) were exclusively present in TNBC patients with PR after anti-PD-

L1+chemotherapy treatment, while heterogenous population of macrophages, including macro-CCL2, macro-CX3CR1, macro-IFI27, macro-IGFBP7, macro-IL1B9, macro-MGP and macro-SLC40A1 cells were exclusively present in TNBC patients achieving PD after chemotherapy. Finally, group 3 innate lymphoid cells (ILC3-AREG and ILC3-IL7R) were linked to PR after anti-PD-L1+chemotherapy treatment, while ZNF683^{+} group 1 innate lymphoid cells (ILC1-ZNF683) cells were exclusively present in TNBC patients achieving PD after chemotherapy. Each of these cell populations have distinctive genetic markers that could be useful therapeutic targets for chemoimmunotherapy.

The role of T follicular helper and B-cell crosstalk in tumor immunity has been extensively studied over the last decade. Accumulating evidence suggests that tumor infiltrated lymphocyte (TIL) subpopulations (CD4, CD8, and CD19/20) constitute of both suppressive (pro-tumor) or effector (anti-tumor) phenotypes whose functions are influenced by the surrounding TME[160]. Natural or treatment-induced immune activation or suppression may determine the balance between pro- or anti-tumor immune cell crosstalk within a given tumor. Key anti-tumor effector activities include antibody-dependent cell cytotoxicity, complement activation, antibody-mediated tumor cell phagocytosis, antigen presentation, T cell activation, cytokine secretion, and direct tumor killing by TIL, including CD8, NK, B cells, and/or macrophages[161].

Despite of significant survival advantages that could be achieved after treatment with chemoimmunotherapy, most TNBC patients would not benefit. Therefore, more and more attention has been paid to the identification and development of biomarkers for the response of chemoimmunotherapy in recent years. Our informatics pipeline identified novel phenotypic and genotypic predictors in unsupervised manner that indicative of favorable outcome after

chemoimmunotherapy. These predictors could be important biomarkers in the design and analysis of intervention studies and ultimately could help to optimize therapy decisions for TNBC patients. In addition, it may help to select better responders to chemoimmunotherapy, as well as pinpoint the underlying mechanisms of drug resistance in TNBC patients at single-cell resolution.

CONCLUSIONS AND FUTURE WORK

In terms of probabilistic framework based on Mutational Forks Formalism we conclude that this tool has a great chance to enable an assessment of patient-specific flow by leveraging information from multiple single-nucleotide alterations to adjust the transitional likelihoods that are solely based on the canonical view of a disease. In the future, we plan to incorporate more signaling pathways and to test the pipeline on a larger cohort of patients. It is possible to extend the analysis to mutational forks that consist of a larger number of genes, given enough computing resources. To reduce the search space for potential gene candidates in a particular mutational fork, we will employ heuristic approaches, including greedy and semi-greedy search, throughout the interactome. For example, the EGFR gene has more than a hundred potential gene-interaction partners, and by applying heuristic searches the analysis could be narrowed down to the most relevant nodes, which in turn would help to address the combinatorial complexity of TC calculations. We also plan to redefine the activation status of the gene (e.g. two standard deviations from the mean value of the expression) to account for highly differentially expressed genes. Moreover, we plan to incorporate other types of genomic data (e.g. deletions, insertions, etc.) and use TREC collections and MEDLINE for non-canonical literature findings. Finally, we will conduct studies outside of the scope of cancer research to test the generalizability of the proposed method.

Regarding informatic pipeline from Chapter 3, we augment our explanatory subgroup discovery algorithm to identify patient subpopulations that may benefit from immuno-targeted combination therapy. Specifically, we identify drug targets that *increase* the likelihood of stable versus progressive disease in cancer patients with HNSC, LUAD, LUSC and SKCM. Our novel

informatic pipeline identified six significant drug targets and thirteen specific compounds for EGFR WT cancer patients. Three out of six drug targets – FCGR2B, IGF1R and KIT – have previously been shown to substantially increase the odds of having a stable disease in other studies. We also show that PFS of more than 6 months was a characteristic feature among investigated EGFR WT subgroups. Moreover, the literature demonstrates immuno-targeted combination therapies with anti-PD-1 or anti-PD-L1 inhibitors substantially prolonged PFS in EGFR WT cancer patients. Further validation of our findings in wet lab experiments will be a significant step toward improving healthcare for cancer patients without targetable mutations.

In terms of informatic framework from Chapter 4, we extend our exploratory subgroup discovery algorithm to identify TNBC subpopulations that may benefit from treatments with chemoimmunotherapy. Specifically, we identify immune-related gene signatures that *increase* the likelihood of partial remission after anti-PD-L1+chemotherapy regimen versus progressive disease after chemotherapy in TNBC patients. Our informatic approach uncover distinct immune cell populations that associated with various treatment outcomes in TNBC. Each of these cell populations have distinctive genetic markers that could be useful therapeutic targets for chemoimmunotherapy. We also show the importance of TMB and menopausal status among the investigated TNBC subgroups. The potential limitations include the usage of two disjoint datasets and the absence of outcome variable for immunotherapy outcomes in TCGA datasets. Due to the computational nature of these studies, there will remain gaps between the wet- and dry-labs for further validating the results.

BIBLIOGRAPHY

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin*. 2022;72(1):7-33. doi:10.3322/caac.21708

2. Sambi M, Bagheri L, Szewczuk MR. Current Challenges in Cancer Immunotherapy: Multimodal Approaches to Improve Efficacy and Patient Response Rates. *J Oncol*. 2019;2019:4508794. doi:10.1155/2019/4508794

3. Ventola CL. Cancer Immunotherapy, Part 3: Challenges and Future Trends. *Pharm Ther*. 2017;42(8):514-521.

4. Frisone D, Friedlaender A, Addeo A, Tsantoulis P. The Landscape of Immunotherapy Resistance in NSCLC. *Front Oncol*. 2022;12. Accessed October 10, 2022. https://www.frontiersin.org/articles/10.3389/fonc.2022.817548

5. Liu D, Baskett W, Beversdorf D, Shyu CR. Exploratory Data Mining for Subgroup Cohort Discoveries and Prioritization. *IEEE J Biomed Health Inform*. 2020;24(5):1456-1468. doi:10.1109/JBHI.2019.2939149

6. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6(269):pl1. doi:10.1126/scisignal.2004088

7. Sever R, Brugge JS. Signal transduction in cancer. *Cold Spring Harb Perspect Med*. 2015;5(4):a006098. doi:10.1101/cshperspect.a006098

8. Khatri P, Sirota M, Butte AJ. Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput Biol*. 2012;8(2):e1002375. doi:10.1371/journal.pcbi.1002375

9. Zhang Q, Li J, Xue H, Kong L, Wang Y. Network-based methods for identifying critical pathways of complex diseases: a survey. *Mol Biosyst*. 2016;12(4):1082-1089. doi:10.1039/c5mb00815h

10. Zhao X, Liu ZP. Analysis of Topological Parameters of Complex Disease Genes Reveals the Importance of Location in a Biomolecular Network. *Genes*. 2019;10(2):143. doi:10.3390/genes10020143

11. Erten S, Bebek G, Koyutürk M. Vavien: An Algorithm for Prioritizing Candidate Disease Genes Based on Topological Similarity of Proteins in Interaction Networks. *J Comput Biol*. 2011;18(11):1561-1574. doi:10.1089/cmb.2011.0154

12. Ning K, Ng HK, Srihari S, Leong HW, Nesvizhskii AI. Examination of the relationship between essential genes in PPI network and hub proteins in reverse nearest neighbor topology. *BMC Bioinformatics*. 2010;11:505. doi:10.1186/1471-2105-11-505

13. Lee D, Cho KH. Topological estimation of signal flow in complex signaling networks. *Sci Rep*. 2018;8(1):5262. doi:10.1038/s41598-018-23643-5

14. Lee D, Cho KH. Signal flow control of complex signaling networks. *Sci Rep*. 2019;9(1):14289. doi:10.1038/s41598-019-50790-0

15. Santolini M, Barabási AL. Predicting perturbation patterns from the topology of biological networks. *Proc Natl Acad Sci U S A*. 2018;115(27):E6375-E6383. doi:10.1073/pnas.1720589115

16. Li D, Gao J. Towards perturbation prediction of biological networks using deep learning. *Sci Rep*. 2019;9(1):11941. doi:10.1038/s41598-019-48391-y

17. Kholod O, Lyu Z, Mitchem J, Tonellato P, Joshi T, Shin D. Mutational Forks: Inferring Deregulated Flow of Signal Transduction Based on Patient-Specific Mutations. In: *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. ; 2019:2063-2068. doi:10.1109/BIBM47256.2019.8983203

18. Creixell P, Schoof EM, Simpson CD, et al. Kinome-wide Decoding of Network-Attacking Mutations Rewiring Cancer Signaling. *Cell*. 2015;163(1):202-217. doi:10.1016/j.cell.2015.08.056

19. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6(269):pl1. doi:10.1126/scisignal.2004088

20. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*. 2017;45(D1):D353-D361. doi:10.1093/nar/gkw1092

21. Thanintorn N, Wang J, Ersoy I, et al. RDF SKETCH MAPS - KNOWLEDGE COMPLEXITY REDUCTION FOR PRECISION MEDICINE ANALYTICS. *Pac Symp Biocomput Pac Symp Biocomput*. 2016;21:417-428.

22. Shin D, Arthur G, Popescu M, Korkin D, Shyu CR. Uncovering influence links in molecular knowledge networks to streamline personalized medicine. *J Biomed Inform*. 2014;52:394-405. doi:10.1016/j.jbi.2014.08.003

23. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res*. 2019;47(D1):D330-D338. doi:10.1093/nar/gky1055

24. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*. 2019;47(D1):D506-D515. doi:10.1093/nar/gky1049

25. Goel R, Harsha HC, Pandey A, Keshava Prasad TS. Human Protein Reference Database and Human Proteinpedia as Resources for Phosphoproteome Analysis. *Mol Biosyst*. 2012;8(2):453-463. doi:10.1039/c1mb05340j

26. Guan J, Gupta R, Filipp FV. Cancer systems biology of TCGA SKCM: efficient detection of genomic drivers in melanoma. *Sci Rep*. 2015;5:7857. doi:10.1038/srep07857

27. Cheng N, Li M, Zhao L, et al. Comparison and integration of computational methods for deleterious synonymous mutation prediction. *Brief Bioinform*. 2020;21(3):970-981. doi:10.1093/bib/bbz047

28. Sever R, Glass CK. Signaling by Nuclear Receptors. *Cold Spring Harb Perspect Biol*. 2013;5(3):a016709. doi:10.1101/cshperspect.a016709

29. Liu TC, Jin X, Wang Y, Wang K. Role of epidermal growth factor receptor in lung cancer and targeted therapies. *Am J Cancer Res*. 2017;7(2):187-202.

30. Brennan DF, Dar AC, Hertz NT, et al. A Raf-induced allosteric transition of KSR stimulates phosphorylation of MEK. *Nature*. 2011;472(7343):366-369. doi:10.1038/nature09860

31. Zhao L, Vogt PK. Class I PI3K in oncogenic cellular transformation. *Oncogene*. 2008;27(41):5486-5496. doi:10.1038/onc.2008.244

32. MacAulay K, Woodgett JR. Targeting glycogen synthase kinase-3 (GSK-3) in the treatment of Type 2 diabetes. *Expert Opin Ther Targets*. 2008;12(10):1265-1274. doi:10.1517/14728222.12.10.1265

33. Zimmermann S, Moelling K. Phosphorylation and regulation of Raf by Akt (protein kinase B). *Science*. 1999;286(5445):1741-1744. doi:10.1126/science.286.5445.1741

34. Matsuura H, Nishitoh H, Takeda K, et al. Phosphorylation-dependent Scaffolding Role of JSAP1/JIP3 in the ASK1-JNK Signaling Pathway: A NEW MODE OF REGULATION OF THE MAP KINASE CASCADE*. *J Biol Chem*. 2002;277(43):40703-40709. doi:10.1074/jbc.M202004200

35. Yao K, Cho YY, Bergen HR, et al. Nuclear factor of activated T3 is a negative regulator of Ras-JNK1/2-AP-1 induced cell transformation. *Cancer Res*. 2007;67(18):8725-8735. doi:10.1158/0008-5472.CAN-06-4788

36. Nihalani D, Wong HN, Holzman LB. Recruitment of JNK to JIP1 and JNK-dependent JIP1 phosphorylation regulates JNK module dynamics and activation. *J Biol Chem*. 2003;278(31):28694-28702. doi:10.1074/jbc.M304212200

37. Canon J, Rex K, Saiki AY, et al. The clinical KRAS(G12C) inhibitor AMG 510 drives anti-tumour immunity. *Nature*. 2019;575(7781):217-223. doi:10.1038/s41586-019-1694-1

38. Forschner A, Battke F, Hadaschik D, et al. Tumor mutation burden and circulating tumor DNA in combined CTLA-4 and PD-1 antibody therapy in metastatic melanoma - results of a prospective biomarker study. *J Immunother Cancer*. 2019;7(1):180. doi:10.1186/s40425-019-0659-0

39. Alexander M, Galeas J, Cheng H. Tumor mutation burden in lung cancer: a new predictive biomarker for immunotherapy or too soon to tell? *J Thorac Dis*. 2018;10(Suppl 33):S3994-S3998. doi:10.21037/jtd.2018.09.35

40. Sharma Y, Miladi M, Dukare S, et al. A pan-cancer analysis of synonymous mutations. *Nat Commun*. 2019;10(1):2569. doi:10.1038/s41467-019-10489-2

41. Leek JT, Scharpf RB, Bravo HC, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet*. 2010;11(10):733-739. doi:10.1038/nrg2825

42. Taefehshokr N, Baradaran B, Baghbanzadeh A, Taefehshokr S. Promising approaches in cancer immunotherapy. *Immunobiology*. 2020;225(2):151875. doi:10.1016/j.imbio.2019.11.010

43. Gettinger SN, Horn L, Gandhi L, et al. Overall Survival and Long-Term Safety of Nivolumab (Anti-Programmed Death 1 Antibody, BMS-936558, ONO-4538) in Patients With Previously Treated Advanced Non-Small-Cell Lung Cancer. *J Clin Oncol Off J Am Soc Clin Oncol*. 2015;33(18):2004-2012. doi:10.1200/JCO.2014.58.3708

44. Herbst RS, Baas P, Kim DW, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. *Lancet Lond Engl*. 2016;387(10027):1540-1550. doi:10.1016/S0140-6736(15)01281-7

45. Blons H, Garinet S, Laurent-Puig P, Oudart JB. Molecular markers and prediction of response to immunotherapy in non-small cell lung cancer, an update. *J Thorac Dis*. 2019;11(Suppl 1):S25-S36. doi:10.21037/jtd.2018.12.48

46. Haslam A, Prasad V. Estimation of the Percentage of US Patients With Cancer Who Are Eligible for and Respond to Checkpoint Inhibitor Immunotherapy Drugs. *JAMA Netw Open*. 2019;2(5):e192535. doi:10.1001/jamanetworkopen.2019.2535

47. Créquit P, Chaimani A, Yavchitz A, et al. Comparative efficacy and safety of second-line treatments for advanced non-small cell lung cancer with wild-type or unknown status for

epidermal growth factor receptor: a systematic review and network meta-analysis. *BMC Med*. 2017;15(1):193. doi:10.1186/s12916-017-0954-x

48. Formenti SC, Demaria S. Systemic effects of local radiotherapy. *Lancet Oncol*. 2009;10(7):718-726. doi:10.1016/S1470-2045(09)70082-8

49. Coffelt SB, de Visser KE. Immune-mediated mechanisms influencing the efficacy of anticancer therapies. *Trends Immunol*. 2015;36(4):198-216. doi:10.1016/j.it.2015.02.006

50. Guan Y, Kraus SG, Quaney MJ, Daniels MA, Mitchem JB, Teixeiro E. FOLFOX Chemotherapy Ameliorates CD8 T Lymphocyte Exhaustion and Enhances Checkpoint Blockade Efficacy in Colorectal Cancer. *Front Oncol*. 2020;10:586. doi:10.3389/fonc.2020.00586

51. Wilmott JS, Long GV, Howle JR, et al. Selective BRAF inhibitors induce marked T-cell infiltration into human metastatic melanoma. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2012;18(5):1386-1394. doi:10.1158/1078-0432.CCR-11-2479

52. Ott PA, Hodi FS, Buchbinder EI. Inhibition of Immune Checkpoints and Vascular Endothelial Growth Factor as Combination Therapy for Metastatic Melanoma: An Overview of Rationale, Preclinical Evidence, and Initial Clinical Data. *Front Oncol*. 2015;5:202. doi:10.3389/fonc.2015.00202

53. Yan Y, Kumar AB, Finnes H, et al. Combining Immune Checkpoint Inhibitors With Conventional Cancer Therapy. *Front Immunol*. 2018;9:1739. doi:10.3389/fimmu.2018.01739

54. Corke L, Sacher A. New Strategies and Combinations to Improve Outcomes in Immunotherapy in Metastatic Non-Small-Cell Lung Cancer. *Curr Oncol Tor Ont*. 2021;29(1):38-55. doi:10.3390/curroncol29010004

55. Sharma P, Hu-Lieskovan S, Wargo JA, Ribas A. Primary, Adaptive, and Acquired Resistance to Cancer Immunotherapy. *Cell*. 2017;168(4):707-723. doi:10.1016/j.cell.2017.01.017

56. Lim AR, Rathmell WK, Rathmell JC. The tumor microenvironment as a metabolic barrier to effector T cells and immunotherapy. *eLife*. 2020;9:e55185. doi:10.7554/eLife.55185

57. Lee M, Samstein RM, Valero C, Chan TA, Morris LGT. Tumor mutational burden as a predictive biomarker for checkpoint inhibitor immunotherapy. *Hum Vaccines Immunother*. 2020;16(1):112-115. doi:10.1080/21645515.2019.1631136

58. Lawlor RT, Mattiolo P, Mafficini A, et al. Tumor Mutational Burden as a Potential Biomarker for Immunotherapy in Pancreatic Cancer: Systematic Review and Still-Open Questions. *Cancers*. 2021;13(13):3119. doi:10.3390/cancers13133119

59. Sade-Feldman M, Jiao YJ, Chen JH, et al. Resistance to checkpoint blockade therapy through inactivation of antigen presentation. *Nat Commun*. 2017;8(1):1136. doi:10.1038/s41467-017-01062-w

60. Miao D, Margolis CA, Vokes NI, et al. Genomic correlates of response to immune checkpoint blockade in microsatellite-stable solid tumors. *Nat Genet*. 2018;50(9):1271-1281. doi:10.1038/s41588-018-0200-2

61. Dotsu Y, Horiike A, Yoshizawa T, et al. Programmed death-ligand 1 expression after progressive disease with EGFR-TKI and efficacy of anti-programmed death-1 antibody in non-small cell lung cancer(NSCLC) harboring EGFR mutation. *J Clin Oncol*. 2018;36(15_suppl):e21232-e21232. doi:10.1200/JCO.2018.36.15_suppl.e21232

62. Garassino MC, Cho BC, Kim JH, et al. Durvalumab as third-line or later treatment for advanced non-small-cell lung cancer (ATLANTIC): an open-label, single-arm, phase 2 study. *Lancet Oncol*. 2018;19(4):521-536. doi:10.1016/S1470-2045(18)30144-X

63. Al-Taie Z, Hannink M, Mitchem J, Papageorgiou C, Shyu CR. Drug Repositioning and Subgroup Discovery for Precision Medicine Implementation in Triple Negative Breast Cancer. *Cancers*. 2021;13(24):6278. doi:10.3390/cancers13246278

64. Park JV, Park SJ, Yoo JS. Finding characteristics of exceptional breast cancer subpopulations using subgroup mining and statistical test. *Expert Syst Appl*. 2019;118:553-562. doi:10.1016/j.eswa.2018.10.016

65. Sheets L, Petroski GF, Zhuang Y, et al. Combining Contrast Mining with Logistic Regression To Predict Healthcare Utilization in a Managed Care Population. *Appl Clin Inform*. 2017;8(2):430-446. doi:10.4338/ACI-2016-05-RA-0078

66. Rodriguez MZ, Comin CH, Casanova D, et al. Clustering algorithms: A comparative approach. *PLOS ONE*. 2019;14(1):e0210236. doi:10.1371/journal.pone.0210236

67. Linzer DA, Lewis JB. poLCA: An R Package for Polytomous Variable Latent Class Analysis. *J Stat Softw*. 2011;42:1-29. doi:10.18637/jss.v042.i10

68. Novak PK, Lavra&#269 N, Webb GI. Supervised Descriptive Rule Discovery: A Unifying Survey of Contrast Set, Emerging Pattern and Subgroup Mining. *J Mach Learn Res*. 2009;10(14):377-403.

69. Lavrač N, Kavšek B, Flach P, Todorovski L. Subgroup Discovery with CN2-SD. *J Mach Learn Res*. 2004;5(Feb):153-188.

70. Liu D, Baskett W, Beversdorf D, Shyu CR. Exploratory Data Mining for Subgroup Cohort Discoveries and Prioritization. *IEEE J Biomed Health Inform*. 2020;24(5):1456-1468. doi:10.1109/JBHI.2019.2939149

71. Rosell R, Carcereny E, Gervais R, et al. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. *Lancet Oncol*. 2012;13(3):239-246. doi:10.1016/S1470-2045(11)70393-X

72. Sequist LV, Yang JCH, Yamamoto N, et al. Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. *J Clin Oncol Off J Am Soc Clin Oncol*. 2013;31(27):3327-3334. doi:10.1200/JCO.2012.44.2806

73. Liu J, Lichtenberg T, Hoadley KA, et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell*. 2018;173(2):400-416.e11. doi:10.1016/j.cell.2018.02.052

74. Borghaei H, Paz-Ares L, Horn L, et al. Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer. *N Engl J Med*. 2015;373(17):1627-1639. doi:10.1056/NEJMoa1507643

75. Cesano A. nCounter(®) PanCancer Immune Profiling Panel (NanoString Technologies, Inc., Seattle, WA). *J Immunother Cancer*. 2015;3:42. doi:10.1186/s40425-015-0088-7

76. Pudil P, Novovičová J, Kittler J. Floating search methods in feature selection. *Pattern Recognit Lett*. 1994;15(11):1119-1125. doi:10.1016/0167-8655(94)90127-9

77. Agrawal R, Srikant R, Road H, Jose S. Fast Algorithms for Mining Association Rules. *Proc 20th Int Conf Very Large Data Bases*. Published online September 12, 1994:487-499.

78. Dong G, Li J. Efficient Mining of Emerging Patterns: Discovering Trends and Differences. *Proc Fifth ACM SIGKDD Int Conf Knowl Discov Data Min*. Published online August 1, 1999:43-52. doi:10.1145/312129.312191

79. Egghe L. Theory and practise of the g-index. *Scientometrics*. 2006;69(1):131-152.

80. Liu Z, Zhao Q, Zuo ZX, et al. Systematic Analysis of the Aberrances and Functional Implications of Ferroptosis in Cancer. *iScience*. 2020;23(7):101302. doi:10.1016/j.isci.2020.101302

81. Agresti A. *Categorical Data Analysis*. 2nd edition. Wiley-Interscience; 2002.

82. Prat A, Navarro A, Paré L, et al. Immune-Related Gene Expression Profiling After PD-1 Blockade in Non-Small Cell Lung Carcinoma, Head and Neck Squamous Cell Carcinoma, and Melanoma. *Cancer Res*. 2017;77(13):3540-3550. doi:10.1158/0008-5472.CAN-16-3556

83. Venables WN, Ripley BD. *Modern Applied Statistics with S*. Springer; 2002. doi:10.1007/978-0-387-21706-2

84. Mok TSK, Wu YL, Kudaba I, et al. Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): a randomised, open-label, controlled, phase 3 trial. *Lancet Lond Engl*. 2019;393(10183):1819-1830. doi:10.1016/S0140-6736(18)32409-7

85. Gandhi L, Rodríguez-Abreu D, Gadgeel S, et al. Pembrolizumab plus Chemotherapy in Metastatic Non-Small-Cell Lung Cancer. *N Engl J Med*. 2018;378(22):2078-2092. doi:10.1056/NEJMoa1801005

86. Peng YP, Wang R, Liu QD, et al. Combination of Tumor Mutational Burden and Specific Gene Mutations Stratifies Outcome to Immunotherapy Across Recurrent and Metastatic Head and Neck Squamous Cell Carcinoma. *Front Genet*. 2021;12:756506. doi:10.3389/fgene.2021.756506

87. McGrail DJ, Pilié PG, Rashid NU, et al. High tumor mutation burden fails to predict immune checkpoint blockade response across all cancer types. *Ann Oncol Off J Eur Soc Med Oncol*. 2021;32(5):661-672. doi:10.1016/j.annonc.2021.02.006

88. Wiesweg M, Preuß C, Roeper J, et al. BRAF mutations and BRAF mutation functional class have no negative impact on the clinical outcome of advanced NSCLC and associate with susceptibility to immunotherapy. *Eur J Cancer Oxf Engl 1990*. 2021;149:211-221. doi:10.1016/j.ejca.2021.02.036

89. Sun LY, Cen WJ, Tang WT, et al. Smoking status combined with tumor mutational burden as a prognosis predictor for combination immune checkpoint inhibitor therapy in non-small cell lung cancer. *Cancer Med*. 2021;10(19):6610-6617. doi:10.1002/cam4.4197

90. Freshour SL, Kiwala S, Cotto KC, et al. Integration of the Drug-Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts. *Nucleic Acids Res*. 2021;49(D1):D1144-D1151. doi:10.1093/nar/gkaa1084

91. Wishart DS, Feunang YD, Guo AC, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res*. 2018;46(D1):D1074-D1082. doi:10.1093/nar/gkx1037

92. Corominas M, Gastaminza G, Lobera T. Hypersensitivity reactions to biological drugs. *J Investig Allergol Clin Immunol*. 2014;24(4):212-225; quiz 1p following 225.

93. Strauss SJ, Morschhauser F, Rech J, et al. Multicenter phase II trial of immunotherapy with the humanized anti-CD22 antibody, epratuzumab, in combination with rituximab, in refractory or recurrent non-Hodgkin's lymphoma. *J Clin Oncol Off J Am Soc Clin Oncol*. 2006;24(24):3880-3886. doi:10.1200/JCO.2006.05.6291

94. Genovese MC, Breedveld FC, Emery P, et al. Safety of biological therapies following rituximab treatment in rheumatoid arthritis patients. *Ann Rheum Dis*. 2009;68(12):1894-1897. doi:10.1136/ard.2008.101675

95. Mayo Clinic. *A Phase 1 Study of PD-1 Inhibition With Pembrolizumab Combined With JAK2 Inhibition in Triple Negative Breast Cancer*. clinicaltrials.gov; 2022. Accessed September 22, 2022. https://clinicaltrials.gov/ct2/show/NCT03012230

96. Bachanova V, Hegerova L, Cao Q, et al. Ruxolitinib Plus Nivolumab in Patients with R/R Hodgkin Lymphoma after Failure of Check-Point Inhibitors: Preliminary Report on Safety and Efficacy. *Blood*. 2021;138:230. doi:10.1182/blood-2021-149675

97. National Cancer Institute (NCI). *Phase I Trial of Methotrexate, Rituximab, Lenalidomide, and Nivolumab (Nivo-MR2) Induction Followed by Lenalidomide and Nivolumab Maintenance in Primary CNS Lymphoma*. clinicaltrials.gov; 2022. Accessed September 22, 2022. https://clinicaltrials.gov/ct2/show/NCT04609046

98. Awad MM, Gadgeel SM, Borghaei H, et al. Long-Term Overall Survival From KEYNOTE-021 Cohort G: Pemetrexed and Carboplatin With or Without Pembrolizumab as First-Line Therapy for Advanced Nonsquamous NSCLC. *J Thorac Oncol*. 2021;16(1):162-168. doi:10.1016/j.jtho.2020.09.015

99. Mayo Clinic. *Phase I Study of Pembrolizumab in Combination With Ibrutinib in the Treatment of Unresectable or Metastatic Melanoma*. clinicaltrials.gov; 2022. Accessed September 22, 2022. https://clinicaltrials.gov/ct2/show/NCT03021460

100. Wesolowski R. *Pilot Study Testing the Effects of BTK Inhibitor Ibrutinib on Levels and Function of Myeloid Derived Suppressor Cells and Other Immune Subsets in Patients With Metastatic Solid Tumors*. clinicaltrials.gov; 2021. Accessed September 22, 2022. https://clinicaltrials.gov/ct2/show/NCT03525925

101. Baylor Research Institute. *Pilot Clinical Trial of Treatment With Bortezomib to Inhibit Homologous Recombination (HR) Followed by Pembrolizumab and Cisplatin in Patients With Chemotherapy-Pretreated Metastatic Triple Negative Breast Cancer*. clinicaltrials.gov; 2021. Accessed September 22, 2022. https://clinicaltrials.gov/ct2/show/NCT04265872

102. Tianjin Medical University Second Hospital. *A Single-Center, Non-Randomized Controlled, Single-Arm, Phase II Clinical Trial of Pembrolizumab and Axitinib as Neoadjuvant Therapy for Locally Advanced Non-Metastatic Clear Cell Renal Cell Carcinoma*. clinicaltrials.gov; 2021. Accessed September 22, 2022. https://clinicaltrials.gov/ct2/show/NCT04995016

103. Saba NF, Ekpenyong A, McCook-Veal A, et al. A phase II trial of pembrolizumab and cabozantinib in patients (pts) with recurrent metastatic head and neck squamous cell carcinoma (RMHNSCC). *J Clin Oncol*. 2022;40(16_suppl):6008-6008. doi:10.1200/JCO.2022.40.16_suppl.6008

104. Amin A, Plimack ER, Ernstoff MS, et al. Safety and efficacy of nivolumab in combination with sunitinib or pazopanib in advanced or metastatic renal cell carcinoma: the CheckMate 016 study. *J Immunother Cancer*. 2018;6(1):109. doi:10.1186/s40425-018-0420-0

105. Kessler ER, Hu J, Srivastava G, et al. Phase I/II trial of pembrolizumab and cabozantinib in the treatment of metastatic renal cell carcinoma (mRCC). *J Clin Oncol*. 2021;39(15_suppl):4544-4544. doi:10.1200/JCO.2021.39.15_suppl.4544

106. Socinski MA, Nishio M, Jotte RM, et al. IMpower150 Final Overall Survival Analyses for Atezolizumab Plus Bevacizumab and Chemotherapy in First-Line Metastatic Nonsquamous NSCLC. *J Thorac Oncol Off Publ Int Assoc Study Lung Cancer*. 2021;16(11):1909-1924. doi:10.1016/j.jtho.2021.07.009

107. Burtness B, Harrington KJ, Greil R, et al. Pembrolizumab alone or with chemotherapy versus cetuximab with chemotherapy for recurrent or metastatic squamous cell carcinoma of the head and neck (KEYNOTE-048): a randomised, open-label, phase 3 study. *Lancet Lond Engl*. 2019;394(10212):1915-1928. doi:10.1016/S0140-6736(19)32591-7

108. BioInvent International AB. *Phase 1/2a Clinical Trial of BI-1206, a Monoclonal Antibody to CD32b (FcYRIIB), in Combination With Rituximab in Subjects With Indolent B-Cell Non-Hodgkin Lymphoma That Has Relapsed or Is Refractory to Rituximab*. clinicaltrials.gov; 2022. Accessed September 22, 2022. https://clinicaltrials.gov/ct2/show/NCT03571568

109. Ajona D, Ortiz-Espinosa S, Lozano T, et al. Short-term starvation reduces IGF-1 levels to sensitize lung tumors to PD-1 immune checkpoint blockade. *Nat Cancer*. 2020;1(1):75-85. doi:10.1038/s43018-019-0007-9

110. Reilley M, Bailey AM, Subbiah V, et al. Phase I clinical trial of combination imatinib and ipilimumab in patients with advanced malignancies. *J Clin Oncol*. 2016;34(15_suppl):3054-3054. doi:10.1200/JCO.2016.34.15_suppl.3054

111. Moss JL, Tatalovich Z, Zhu L, Morgan C, Cronin KA. Triple-negative breast cancer incidence in the United States: ecological correlations with area-level sociodemographics, healthcare, and health behaviors. *Breast Cancer Tokyo Jpn*. 2021;28(1):82-91. doi:10.1007/s12282-020-01132-w

112. Lehmann BD, Pietenpol JA. Identification and use of biomarkers in treatment strategies for triple negative breast cancer subtypes. *J Pathol*. 2014;232(2):142-150. doi:10.1002/path.4280

113. Gupta GK, Collier AL, Lee D, et al. Perspectives on Triple-Negative Breast Cancer: Current Treatment Strategies, Unmet Needs, and Potential Targets for Future Therapies. *Cancers*. 2020;12(9):2392. doi:10.3390/cancers12092392

114. Tarantino P, Corti C, Schmid P, et al. Immunotherapy for early triple negative breast cancer: research agenda for the next decade. *Npj Breast Cancer*. 2022;8(1):1-7. doi:10.1038/s41523-022-00386-1

115. Tutt ANJ, Garber JE, Kaufman B, et al. Adjuvant Olaparib for Patients with BRCA1- or BRCA2-Mutated Breast Cancer. *N Engl J Med*. 2021;384(25):2394-2405. doi:10.1056/NEJMoa2105215

116. O'Meara TA, Tolaney SM. Tumor mutational burden as a predictor of immunotherapy response in breast cancer. *Oncotarget*. 2021;12(5):394-400. doi:10.18632/oncotarget.27877

117. Cortes J, Cescon DW, Rugo HS, et al. Pembrolizumab plus chemotherapy versus placebo plus chemotherapy for previously untreated locally recurrent inoperable or metastatic triple-negative breast cancer (KEYNOTE-355): a randomised, placebo-controlled, double-blind, phase 3 clinical trial. *Lancet Lond Engl*. 2020;396(10265):1817-1828. doi:10.1016/S0140-6736(20)32531-9

118. Hutchinson KE, Yost SE, Chang CW, et al. Comprehensive Profiling of Poor-Risk Paired Primary and Recurrent Triple-Negative Breast Cancers Reveals Immune Phenotype Shifts. *Clin Cancer Res*. 2020;26(3):657-668. doi:10.1158/1078-0432.CCR-19-1773

119. Schmid P, Cortes J, Dent R, et al. VP7-2021: KEYNOTE-522: Phase III study of neoadjuvant pembrolizumab + chemotherapy vs. placebo + chemotherapy, followed by adjuvant pembrolizumab vs. placebo for early-stage TNBC. *Ann Oncol*. 2021;32(9):1198-1200. doi:10.1016/j.annonc.2021.06.014

120. Criscitiello C, Bayar MA, Curigliano G, et al. A gene signature to predict high tumor-infiltrating lymphocytes after neoadjuvant chemotherapy and outcome in patients with triple-negative breast cancer. *Ann Oncol Off J Eur Soc Med Oncol*. 2018;29(1):162-169. doi:10.1093/annonc/mdx691

121. Miles D, Gligorov J, André F, et al. Primary results from IMpassion131, a double-blind, placebo-controlled, randomised phase III trial of first-line paclitaxel with or without

atezolizumab for unresectable locally advanced/metastatic triple-negative breast cancer. *Ann Oncol Off J Eur Soc Med Oncol*. 2021;32(8):994-1004. doi:10.1016/j.annonc.2021.05.801

122. Davis AA, Patel VG. The role of PD-L1 expression as a predictive biomarker: an analysis of all US Food and Drug Administration (FDA) approvals of immune checkpoint inhibitors. *J Immunother Cancer*. 2019;7(1):278. doi:10.1186/s40425-019-0768-9

123. Geukes Foppen MH, Donia M, Svane IM, Haanen JBAG. Tumor-infiltrating lymphocytes for the treatment of metastatic cancer. *Mol Oncol*. 2015;9(10):1918-1935. doi:10.1016/j.molonc.2015.10.018

124. Cristescu R, Aurora-Garg D, Albright A, et al. Tumor mutational burden predicts the efficacy of pembrolizumab monotherapy: a pan-tumor retrospective analysis of participants with advanced solid tumors. *J Immunother Cancer*. 2022;10(1):e003091. doi:10.1136/jitc-2021-003091

125. Gradishar WJ, Anderson BO, Abraham J, et al. Breast Cancer, Version 3.2020, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Cancer Netw JNCCN*. 2020;18(4):452-478. doi:10.6004/jnccn.2020.0016

126. Zhang Y, Chen H, Mo H, et al. Single-cell analyses reveal key immune cell subsets associated with response to PD-L1 blockade in triple-negative breast cancer. *Cancer Cell*. 2021;39(12):1578-1593.e8. doi:10.1016/j.ccell.2021.09.010

127. nnet: Feed-Forward Neural Networks and Multinomial Log-Linear Models version 7.3-17 from CRAN. Accessed September 2, 2022. https://rdrr.io/cran/nnet/

128. Azizi E, Carr AJ, Plitas G, et al. Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. *Cell*. 2018;174(5):1293-1308.e36. doi:10.1016/j.cell.2018.05.060

129. Karn T, Denkert C, Weber KE, et al. Tumor mutational burden and immune infiltration as independent predictors of response to neoadjuvant immune checkpoint inhibition in early TNBC in GeparNuevo. *Ann Oncol Off J Eur Soc Med Oncol*. 2020;31(9):1216-1222. doi:10.1016/j.annonc.2020.05.015

130. Winer EP, Lipatov O, Im SA, et al. Association of tumor mutational burden (TMB) and clinical outcomes with pembrolizumab (pembro) versus chemotherapy (chemo) in patients with metastatic triple-negative breast cancer (mTNBC) from KEYNOTE-119. *J Clin Oncol*. 2020;38(15_suppl):1013-1013. doi:10.1200/JCO.2020.38.15_suppl.1013

131. Zhang Y, Asad S, Weber Z, et al. Genomic features of rapid versus late relapse in triple negative breast cancer. *BMC Cancer*. 2021;21(1):568. doi:10.1186/s12885-021-08320-7

132. Nishimura R, Osako T, Okumura Y, et al. Triple Negative Breast Cancer: An Analysis of the Subtypes and the Effects of Menopausal Status on Invasive Breast Cancer. *J Clin Med*. 2022;11(9):2331. doi:10.3390/jcm11092331

133. Arafah MA, Ouban A, Ameer OZ, Quek KJ. KI-67 LI Expression in Triple-Negative Breast Cancer Patients and Its Significance. *Breast Cancer Basic Clin Res*. 2021;15:11782234211016976. doi:10.1177/11782234211016977

134. Inari H, Suganuma N, Kawachi K, et al. Clinicopathological and prognostic significance of Ki-67 immunohistochemical expression of distant metastatic lesions in patients with metastatic breast cancer. *Breast Cancer Tokyo Jpn*. 2017;24(6):748-755. doi:10.1007/s12282-017-0774-z

135. Baumjohann D, Brossart P. T follicular helper cells: linking cancer immunotherapy and immune-related adverse events. *J Immunother Cancer*. 2021;9(6):e002588. doi:10.1136/jitc-2021-002588

136. Ma Q, Chen Y, Qin Q, Guo F, Wang Y sheng, Li D. CXCL13 expression in mouse 4T1 breast cancer microenvironment elicits antitumor immune response by regulating immune cell infiltration. *Precis Clin Med*. 2021;4(3):155-167. doi:10.1093/pcmedi/pbab020

137. Cheriyath V, Kaur J, Davenport A, Khalel A, Chowdhury N, Gaddipati L. G1P3 (IFI6), a mitochondrial localised antiapoptotic protein, promotes metastatic potential of breast cancer cells through mtROS. *Br J Cancer*. 2018;119(1):52-64. doi:10.1038/s41416-018-0137-3

138. Hollern DP, Xu N, Thennavan A, et al. B cells and T follicular helper cells mediate response to checkpoint inhibitors in high mutation burden mouse models of breast cancer. *Cell*. 2019;179(5):1191-1206.e21. doi:10.1016/j.cell.2019.10.028

139. Mori H, Ouchida R, Hijikata A, et al. Deficiency of the oxidative damage-specific DNA glycosylase NEIL1 leads to reduced germinal center B cell expansion. *DNA Repair*. 2009;8(11):1328-1332. doi:10.1016/j.dnarep.2009.08.007

140. Shinmura K, Kato H, Kawanishi Y, et al. Abnormal Expressions of DNA Glycosylase Genes NEIL1, NEIL2, and NEIL3 Are Associated with Somatic Mutation Loads in Human Cancer. *Oxid Med Cell Longev*. 2016;2016:1546392. doi:10.1155/2016/1546392

141. Yeong J, Lim JCT, Lee B, et al. High Densities of Tumor-Associated Plasma Cells Predict Improved Prognosis in Triple Negative Breast Cancer. *Front Immunol*. 2018;9:1209. doi:10.3389/fimmu.2018.01209

142. Pelegrina LT, Lombardi MG, Fiszman GL, Azar ME, Morgado CC, Sales ME. Immunoglobulin g from breast cancer patients regulates MCF-7 cells migration and MMP-9 activity by stimulating muscarinic acetylcholine receptors. *J Clin Immunol*. 2013;33(2):427-435. doi:10.1007/s10875-012-9804-y

143. López-Otín C, Matrisian LM. Emerging roles of proteases in tumour suppression. *Nat Rev Cancer*. 2007;7(10):800-808. doi:10.1038/nrc2228

144. Pellikainen JM, Ropponen KM, Kataja VV, Kellokoski JK, Eskelinen MJ, Kosma VM. Expression of matrix metalloproteinase (MMP)-2 and MMP-9 in breast cancer with a special reference to activator protein-2, HER2, and prognosis. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2004;10(22):7621-7628. doi:10.1158/1078-0432.CCR-04-1061

145. Coussens LM, Fingleton B, Matrisian LM. Matrix metalloproteinase inhibitors and cancer: trials and tribulations. *Science*. 2002;295(5564):2387-2392. doi:10.1126/science.1067100

146. Bendrik C, Robertson J, Gauldie J, Dabrosin C. Gene transfer of matrix metalloproteinase-9 induces tumor regression of breast cancer in vivo. *Cancer Res*. 2008;68(9):3405-3412. doi:10.1158/0008-5472.CAN-08-0295

147. Leifler KS, Svensson S, Abrahamsson A, et al. Inflammation Induced by MMP-9 Enhances Tumor Regression of Experimental Breast Cancer. *J Immunol Author Choice*. 2013;190(8):4420-4430. doi:10.4049/jimmunol.1202610

148. Sun X, Glynn DJ, Hodson LJ, et al. CCL2-driven inflammation increases mammary gland stromal density and cancer susceptibility in a transgenic mouse model. *Breast Cancer Res BCR*. 2017;19(1):4. doi:10.1186/s13058-016-0796-z

149. Zheng J, Yang M, Shao J, Miao Y, Han J, Du J. Chemokine receptor CX3CR1 contributes to macrophage survival in tumor metastasis. *Mol Cancer*. 2013;12(1):141. doi:10.1186/1476-4598-12-141

150. Cervantes-Badillo MG, Paredes-Villa A, Gómez-Romero V, et al. IFI27/ISG12 Downregulates Estrogen Receptor α Transactivation by Facilitating Its Interaction With CRM1/XPO1 in Breast Cancer Cells. *Front Endocrinol*. 2020;11. Accessed September 27, 2022. https://www.frontiersin.org/articles/10.3389/fendo.2020.568375

151. Tulotta C, Lefley DV, Moore CK, et al. IL-1B drives opposing responses in primary tumours and bone metastases; harnessing combination therapies to improve outcome in breast cancer. *NPJ Breast Cancer*. 2021;7(1):95. doi:10.1038/s41523-021-00305-w

152. Mizuno M, Khaledian B, Maeda M, et al. Adipsin-Dependent Secretion of Hepatocyte Growth Factor Regulates the Adipocyte-Cancer Stem Cell Interaction. *Cancers*. 2021;13(16):4238. doi:10.3390/cancers13164238

153. Liu Z, Gao Z, Li B, et al. Lipid-associated macrophages in the tumor-adipose microenvironment facilitate breast cancer progression. *Oncoimmunology*. 2022;11(1):2085432. doi:10.1080/2162402X.2022.2085432

154. Howard E, Hurrell BP, Helou DG, et al. PD-1 Blockade on Tumor Microenvironment-Resident ILC2s Promotes TNF-α Production and Restricts Progression of Metastatic Melanoma. *Front Immunol*. 2021;12. Accessed September 29, 2022. https://www.frontiersin.org/articles/10.3389/fimmu.2021.733136

155. Halim TYF, Rana BMJ, Walker JA, et al. Tissue-Restricted Adaptive Type 2 Immunity Is Orchestrated by Expression of the Costimulatory Molecule OX40L on Group 2 Innate Lymphoid Cells. *Immunity*. 2018;48(6):1195-1207.e6. doi:10.1016/j.immuni.2018.05.003

156. Grisaru-Tal S, Itan M, Klion AD, Munitz A. A new dawn for eosinophils in the tumour microenvironment. *Nat Rev Cancer*. 2020;20(10):594-607. doi:10.1038/s41568-020-0283-9

157. Carretero R, Sektioglu IM, Garbi N, Salgado OC, Beckhove P, Hämmerling GJ. Eosinophils orchestrate cancer rejection by normalizing tumor vessels and enhancing infiltration of CD8(+) T cells. *Nat Immunol*. 2015;16(6):609-617. doi:10.1038/ni.3159

158. Vienne M, Etiennot M, Escalière B, et al. Type 1 Innate Lymphoid Cells Limit the Antitumoral Immune Response. *Front Immunol*. 2021;12. Accessed September 29, 2022. https://www.frontiersin.org/articles/10.3389/fimmu.2021.768989

159. Chen Q, Massagué J. Molecular pathways: VCAM-1 as a potential therapeutic target in metastasis. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2012;18(20):5520-5525. doi:10.1158/1078-0432.CCR-11-2904

160. Kalaora S, Nagler A, Wargo JA, Samuels Y. Mechanisms of immune activation and regulation: lessons from melanoma. *Nat Rev Cancer*. 2022;22(4):195-207. doi:10.1038/s41568-022-00442-9

161. Salemme V, Centonze G, Cavallo F, Defilippi P, Conti L. The Crosstalk Between Tumor Cells and the Immune Microenvironment in Breast Cancer: Implications for Immunotherapy. *Front Oncol*. 2021;11:610303. doi:10.3389/fonc.2021.610303

APPENDIX

Predictor	Value	Std. Error	t value	p value
C1QA	6.8918	6.7934	1.0144	0.3103
C1QB	1.5290	9.4436	0.1619	0.8713
C1R	-9.0004	9.1254	-0.9863	0.3239
C1S	8.7135	8.3106	1.0484	0.2944
CDH5	-3.5489	1.7693	-2.0057	0.0448
CSF1	-0.6283	3.0745	-0.2043	0.8380
CSF1R	2.01424	1.4266	1.4118	0.1579
CXCR2	0.4276	0.2879	1.4851	0.1375
FCGR1A	-0.2995	0.3761	-0.7963	0.4258
FCGR2A	-5.2699	6.3559	-0.8291	0.4070
FCGR2B	9.7079	2.7078	3.5851	0.0003
FCGR3A	-6.7021	3.8532	-1.7393	0.0819
IDO1	2.2873	1.1896	1.9226	0.0545
IGF1R	7.7136	3.8821	1.9869	0.0469
ITK	-2.0603	0.7716	-2.6702	0.0075
JAK2	-11.4664	5.6822	-2.0179	0.0435
KIT	1.2729	0.4280	2.9736	0.0029
MAP2K2	-10.8645	8.0887	-1.3431	0.1792
PBK	-0.0886	0.3955	-0.2241	0.8226
PDGFRB	6.8776	3.8400	1.7910	0.0732
PSMB10	2.1847	6.3441	0.3443	0.73056
PSMB9	-0.8971	4.1000	-0.2188	0.8267
PTGS2	-0.7541	0.4883	-1.5442	0.1225

PR|CR	-12.5307	32.4375	-0.3863	0.6992
CR|PD	-12.1570	32.4333	-0.3748	0.7077
PD|SD	-9.0625	32.4213	-0.2795	0.7798

Supplementary Table 1: The summary for proportional odds model.

Predictor	OR	2.5 %	97.5 %
C1QA	$9.8424 \times 10^{+2}$	1.6237×10^{-3}	$5.9661 \times 10^{+8}$
C1QB	4.6136	4.2224×10^{-8}	$5.0410 \times 10^{+8}$
C1R	1.2334×10^{-4}	2.1064×10^{-12}	$7.2232 \times 10^{+3}$
C1S	$6.0846 \times 10^{+3}$	5.1306×10^{-4}	$7.2161 \times 10^{+10}$
CDH5	2.8753×10^{-2}	8.9657×10^{-4}	9.2214×10^{-1}
CSF1	5.3346×10^{-1}	1.2884×10^{-3}	$2.2086 \times 10^{+2}$
CSF1R	7.4950	4.5753×10^{-1}	$1.2277 \times 10^{+2}$
CXCR2	1.5336	8.7221×10^{-1}	2.6967
FCGR1A	7.4113×10^{-1}	3.5456×10^{-1}	1.5491
FCGR2A	5.1436×10^{-3}	2.0001×10^{-8}	$1.3227 \times 10^{+3}$
FCGR2B	$1.6448 \times 10^{+4}$	$8.1514 \times 10^{+1}$	$3.3190 \times 10^{+6}$
FCGR3A	1.2282×10^{-3}	6.4481×10^{-7}	2.3396
IDO1	9.8487	9.5661×10^{-1}	$1.0139 \times 10^{+2}$
IGF1R	$2.2386 \times 10^{+3}$	1.1103	$4.5134 \times 10^{+6}$
ITK	1.2740×10^{-1}	2.8079×10^{-2}	5.7806×10^{-1}
JAK2	1.0476×10^{-5}	1.5256×10^{-10}	7.1934×10^{-1}
KIT	3.5714	1.5433	8.2648
MAP2K2	1.9124×10^{-5}	2.4912×10^{-12}	$1.4681 \times 10^{+2}$
PBK	9.1516×10^{-1}	4.2150×10^{-1}	1.9869
PDGFRB	$9.7035 \times 10^{+2}$	5.2271×10^{-1}	$1.8013 \times 10^{+6}$
PSMB10	8.8883	3.5371×10^{-5}	$2.2334 \times 10^{+6}$
PSMB9	4.0772×10^{-1}	1.3194×10^{-4}	$1.2598 \times 10^{+3}$
PTGS2	4.7040×10^{-1}	1.8062×10^{-1}	1.2251

Supplementary Table 2: Proportional odds model odds ratios and confidence intervals for the proportional odds model.

VITA

Olha is a PhD Candidate in Biomedical Informatics at MU Institute for Data Science & Informatics. She is a Graduate Research Assistant at the Interdisciplinary Data Analytics and Search (iDAS) Laboratory, supervised by Dr. Chi-Ren Shyu. Her research interests include functional genomics, cancer informatics and data mining. Specifically, Olha is working on deciphering immune-related gene signatures to evaluate the efficacy of combination treatments with immunotherapy for cancer patients using explanatory subgroup discovery. In the past, she was a recipient of prestigious Fulbright Graduate Scholarship from Ukraine, that allows her to gain knowledge and expertise in whole-transcriptome profiling of cancer patients. Olha is also an alumna of Edmund Muskie Internship Program from US Department of States, that allows her to conduct research in analyzing differential gene expression patterns using single-cell sequencing technology at the Center of Public Health Genomics in University of Virginia, Charlottesville. During her PhD studies, Olha have presented her research on multiple local and national conferences, including AMIA Translational Bioinformatics Symposium, ISMB and ICIBM conferences. She is also a reviewer for conference manuscripts and podium presentations for AMIA and IEEE BIBM conferences. Olha will complete her Ph.D. degree in December 2022 and will continue as a Postdoctoral Research Associate in Dr. Britt Ann Goods laboratory at Dartmouth College.