A COMPREHENSIVE ANALYSIS OF MUTATIONS IN CANCERS AND NEURODEVELOPMENTAL DISORDERS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS OF
THE MIDDLE EAST TECHNICAL UNIVERSITY
BY

BENGİ RUKEN YAVUZ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
THE DEPARTMENT OF MEDICAL INFORMATICS

JUNE 2023

Approval of the thesis:

A COMPREHENSIVE ANALYSIS OF MUTATIONS IN CANCERS AND NEURODEVELOPMENTAL DISORDERS

Submitted by BENGİ RUKEN YAVUZ in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Health Informatics Department, Middle East Technical University** by,

Prof. Dr. Banu Günel-Kılıç
Dean, **Graduate School of Informatics**

Assoc. Prof. Dr. Yeşim Aydın-Son
Head of Department, **Health Informatics**

Assoc. Prof. Dr. Yeşim Aydın-Son
Supervisor, **Health Informatics Dept., METU**

Assoc. Prof. Dr. Nurcan Tunçbağ
Co-Supervisor, **Chemical and Biol. Eng. Dept., Koç University**

Examining Committee Members:

Assist. Prof. Dr. Burçak Otlu-Sarıtaş
Health Informatics Dept., METU

Assoc. Prof. Dr. Yeşim Aydın-Son
Health Informatics Dept., METU

Assist. Prof. Dr. Aybar Can Acar
Health Informatics Dept., METU

Assoc. Prof. Dr. Tunca Doğan
Computer Engineering Dept., Hacettepe University

Assoc. Prof. Dr. Ceren Sucularlı
Bioinformatics Dept., Hacettepe University

Date: *05.06.2023*

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : **Bengi Ruken Yavuz**

Signature : ____________________

ABSTRACT

A COMPREHENSIVE ANALYSIS OF MUTATIONS IN CANCERS AND NEURODEVELOPMENTAL DISORDERS

Yavuz, Bengi Ruken
Ph.D, Department of Health Informatics
Supervisor: Assoc. Prof. Dr. Yeşim Aydın-Son
Co-Supervisor: Assoc. Prof. Dr. Nurcan Tunçbağ

June 2023, 127 pages

A multitude of pathologies are driven by mutations, and even a single mutation can act as a prognostic marker and modify the global genome and protein expression, thus changing oncogenic signaling pathways. However, a single-driver mutation's contribution to the activation of oncogenic signaling pathways is minimal and necessitates additional mutations over time. This thesis aims to identify co-occurring mutations in genes that promote tumorigenesis and alter the response to treatments. To achieve this, we employed a statistical technique to identify significantly co-occurring mutations in pan-cancer cohorts, focusing on the identification of latent driver mutations with minor observable translational potential and low frequencies. We then discovered 4352 significant different gene double mutations that alter non-redundant pathways and interactions and promote cancer-specific tumorigenesis. Rare co-occurring in trans combinations can serve as metastasis markers, whereas excluded combinations may give rise to oncogene-induced senescence (OIS). Furthermore, this thesis investigates the shared and distinguishing features between neurodevelopmental disorders (NDDs) and cancer. Despite the differences in their clinical manifestations, individuals with NDDs are more likely to develop cancer. Cancer mutations are sporadic and arise during life, whereas NDD-associated mutations are germline. However, both NDDs and cancer share proteins, pathways, and mutations. Deep patterns that are rare but can prompt dramatic phenotypic alterations and serve as clinical signatures can be discovered through interrogating large genomic data and integrating it with small-molecule sensitivity data.

Keywords: mutation doublets, molecular signatures of cancer, latent drivers, cancer genome analysis, neurodevelopmental disorders

ÖZ

KANSERLERDE VE NÖROGELİŞİMSEL HASTALIKLARDA BULUNAN MUTASYONLARIN KAPSAMLI BİR ANALİZİ

Yavuz, Bengi Ruken
Doktora, Sağlık Bilişimi Bölümü
Tez Yöneticisi: Doç. Dr. Yeşim Aydın-Son
Eş-Danışman: Doç. Dr. Nurcan Tunçbağ

Haziran 2023, 127 sayfa

Çok sayıda patoloji mutasyonlar tarafından yönlendirilir ve tek bir mutasyon bile tanısal bir belirteç olabilir ve küresel genom ve protein ekspresyonunu değiştirerek onkojenik sinyal yollarını değiştirebilir. Bununla birlikte, tek sürücülü bir mutasyonun onkojenik sinyal yollarının aktivasyonuna katkısı çok düşüktür ve zaman içinde ek mutasyonlara ihtiyaç duyulur. Bu tez, tümör oluşumunu teşvik eden ve tedavilere yanıtı değiştiren aynı gen üzerinde bulunan mutasyonları tespit etmeyi amaçlamaktadır. Bu amaçla, pan-kanser gruplarında birlikte bulunan anlamlı mutasyonları belirlemek için istatistiksel bir yöntem kullandık, translasyon potansiyeli belirgin olmayan ve düşük frekanslara sahip gizli sürücü mutasyonlarının tanımlanmasına odaklandık. Daha sonra fazlalık olmayan yolları ve etkileşimleri değiştiren ve kansere özgü tümör oluşumunu destekleyen 4352 anlamlı farklı gen çift mutasyonu keşfettik. Trans kombinasyonlarında bulunan seyrek ikililer, metastaz belirteci olarak işlev görebilirken, ayrık kombinasyonlar, onkojen kaynaklı yaşlanmaya yol açabilir. Ayrıca, bu tez nörogelişimsel bozukluklar ve kanser arasındaki ortak ve ayırt edici özellikleri araştırmaktadır. Klinik belirtilerindeki farklılıklara rağmen, nörogelişimsel hastalıklara sahip bireylerin kansere yakalanma olasılığı daha yüksektir. Kanser mutasyonları seyrek ve somatikken, NDD ile ilişkili mutasyonları kalıtsaldır. Bununla birlikte hem nörogelişimsel hastalıklar hem de kanser, aynı proteinleri, yolakları ve mutasyonları paylaşır. Nadir olan ancak dramatik fenotipik değişikliklere yol açabilen ve klinik imza görevi görebilen derin motifler, büyük genomik verilerin sorgulanması ve küçük molekül hassasiyet verileriyle bütünleştirilmesi yoluyla keşfedilebilir.

Anahtar Sözcükler: mutasyon ikilileri, kanserin moleküler imzaları, gizli sürücüler, kanser genom analizi, nörogelişimsel hastalıklar

To my family…

ACKNOWLEDGEMENTS

I have several people to express my gratitude for their support; without their invaluable help, it would not be possible to conduct this research. First of all, I would like to express my sincere gratitude to my supervisors, Assoc. Prof. Nurcan Tunçbağ and Assoc. Prof. Yeşim Aydın-Son, whose encouragement, friendship, and unwavering support throughout the process enabled me to complete this dissertation.

I am grateful to Prof. Ruth Nussinov for her guidance, for sharing her vast experience, and for being supportive all the time. I am grateful to Ruth Nussinov's research group members for all of the fruitful scientific discussions, but especially to Dr. Chung-Jung Tsai for laying the foundation for this dissertation. I would like to thank my thesis committee members Assoc. Prof. Tunca Doğan, Asst. Prof. Burçak Otlu-Sarıtaş, Asst. Prof. Aybar Can Acar, Assoc. Prof. Ceren Sucularlı for their critical reading and useful comments.

I appreciate M. Kaan Arıcı and H. Cansu Demirel for their contribution for the preparation of Chapter 5. Besides, I am also thankful to members of the Network Modeling Research Group and to everyone I had the opportunity to work with at the METU Informatics Institute. I would especially like to thank to Okan Bilge Özdemir for the support he provided for thesis formatting issues.

I express my appreciation to my dear friends Elif Doğan-Dar and Fulya Başak Kumkumoğlu for all their support and friendship. Also, I am truly indebted to my family, my father Osman and my brother Erbil for their endless support in every aspect of life; in particular to my late mother Sevgi for everything she taught me during the time we had together.

This project has been funded in whole or in part with TÜBİTAK funds under the project number 121C292 and 121E245.

TABLE OF CONTENTS

ABSTRACT iv
ÖZ v
DEDICATION vi
ACKNOWLEDGEMENTS vii
TABLE OF CONTENTS viii
LIST OF FIGURES xi
LIST OF ABBREVIATIONS xii

CHAPTERS

1. INTRODUCTION 1
2. LITERATURE REVIEW 7
  2.1 Genome sequencing and variant classes 7
  2.2 Driver, passenger, and latent driver mutations 8
  2.3 Latent driver mutations from a biophysical perspective 10
  2.4 Oncogenes and tumor suppressor genes 12
  2.5 Double/multiple mutations and latent drivers 13
  2.6 Co-occurring and mutually exclusive mutations in different genes 15
  2.7 Connection between the neurodevelopmental disorders and cancer 17
3. DISCOVERY OF LATENT DRIVERS FROM DOUBLE MUTATIONS IN PAN-CANCER DATA REVEAL THEIR CLINICAL IMPACT 21
  3.1 Methods 21
    3.1.1 Data collection and processing 21
    3.1.2 Statistics and reproducibility 22
    3.1.3 Hyper-mutated samples and double mutations 23
    3.1.4 Allelic configuration of double mutations 24
    3.1.5 Mutational signature analysis 24
    3.1.6 Cell line network construction 24
    3.1.7 Patient-derived xenograft analysis 24
    3.1.8 Availability of data and materials 25

3.2 Results 25
3.2.1 Discovery of latent drivers through double mutations 25
3.2.2 Functional interpretation of double mutations by using the characteristics of mutation constituents and double mutant tumors 32
3.2.3 Doublets in the same gene are rare but are a signature for some cancer types 34
3.2.4 Linking double mutations to clinical data using cancer cell lines and xenografts 37
3.3 Concluding remarks 47
4. CO-OCCURRING MUTATIONS IN TRANS CAN BE METASTATIC MARKERS; EXCLUDED COMBINATIONS CAN ENCODE ONCOGENE-INDUCED SENESCENCE 49
4.1 Methods 49
4.1.1 Data collection and processing 49
4.1.2 Identification of significant double alterations 50
4.1.3 Survival analysis 50
4.1.4 Oncoprint maps 51
4.1.5 Transcriptome analysis 51
4.1.6 FPGrowth tree construction 52
4.1.7 Personalized page rank algorithm 52
4.2 Results 53
4.2.1 Functionally non-redundant mutations on different genes in tumorigenesis 53
4.2.2 Can double mutations co-occur in the same interface? 56
4.2.3 Double mutations are mostly present in primary tumors and some rare doubles are signature of metastatic tumors 61
4.2.4 Functionally equivalent alterations do not co-exist in tumors 62
4.3 Concluding remarks 65
5. NEURODEVELOPMENTAL DISORDERS AND CANCER NETWORKS SHARE PATHWAYS; BUT DIFFER IN MECHANISMS, SIGNALING STRENGTH, AND OUTCOME 67
5.1 Methods 67
5.1.1 Data collection and processing 67

5.1.2 TCGA 68

5.1.3 Cancer drivers 68

5.1.4 Visualization of mutations in protein sequences and 3D structures...........68

5.2 Results 68

5.2.1 NDD and cancer mutations can be shared, but their presentation and phenotypic damage differ 68

5.2.2 Mechanistic interpretation of NDD and cancer mutations 72

5.3 Concluding remarks 76

6. DISCUSSION 79

REFERENCES 87

APPENDICES 117

APPENDIX A 117

APPENDIX B 123

APPENDIX C 125

CURRICULUM VITAE 127

LIST OF FIGURES

Figure 1. Mutations affect the cellular network. 10
Figure 2. Free energy landscapes of conformational switches for different mutation types 12
Figure 3. Double mutations in the same gene. 26
Figure 4. Distribution of double mutant tumors across different tissues.. 26
Figure 5. Cancer specificity of the genes harboring double mutations. 28
Figure 6. Composition of the double mutations. 30
Figure 7. Box plot showing passenger mutation load in OGs and TSGs. 31
Figure 8. Tumor count distributions of known driver and potential latent driver mutations. 32
Figure 9. Same gene double mutations are specific to some tissues or cancer subtypes.. 36
Figure 10. Representation of mutations in genes to compose a doublet as a circular diagram. 37
Figure 11. Paired dot plot of the 23 double mutations in *PIK3CA*. 38
Figure 12. Presence of *PIK3CA* same gene double mutations across different cancer tissues.. 39
Figure 13. 3D structure of *PIK3CA*. 40
Figure 14. Response of *PIK3CA* breast cancer cell lines with double mutations in *cis* to drugs in network representation. 42
Figure 15. Tumor volume changes of single and double *PIK3CA* mutant xenografts without any treatment. 44
Figure 16. Structural and clinical aspects of EGFR double mutations.. 46
Figure 17. Overview of different gene double mutations. 54
Figure 18. Widespread different gene double mutations in different tissues and cancer subtypes 57
Figure 19. Transcriptome analysis of tumors with double mutations. 59
Figure 20. Different gene double mutations on the PIK3CA-PIK3R1 complex. 60
Figure 21. Some rare double mutations are specific to metastatic tumors. 61
Figure 22. Different gene double mutations in PAAD tumors.. 63
Figure 23. Detailed analysis of KEAP1-NFE2L2 alterations. 65
Figure 24. Overview of the data and workflow.. 69
Figure 25. Comparison of mutations between NDDs and cancer.. 72
Figure 26. Profiles of TCGA and NDDs mutations for PTEN and PI3Kα at the residue level on the sequence and structure. 75

LIST OF ABBREVIATIONS

TCGA	The Cancer Genome Atlas
AACR	American Association for Cancer Research
COSMIC	Catalogue of Somatic Mutations in Cancer
IntOGen	Integrative Onco Genomics
TFBS	Transcription Factor Binding Site
DEG	Differentially Expressed Genes
GENIE	Genomics Evidence Neoplasia Information Exchange
WebGestalt	WEB-based GEne SeT AnaLysis Toolkit
OIS	Oncogene Induced Senescence
CoMDP	**Co**-occurring **M**utated **D**river **P**athways
PIK3CA	Phosphatidylinositol-4,5-bisphosphate 3-kinase Catalytic Subunit Alpha
PTEN	Phosphatase and Tensin Homolog
EGFR	Epidermal Growth Factor Receptor
CDKN2A	Cyclin Dependent Kinase Inhibitor 2A
MYC	MYC Proto-oncogene, bHLH Transcription Factor
BRAF	B-Raf Proto-oncogene, Serine/threonine Kinase
KRAS	Kirsten Rat Sarcoma Virus
CDK6	Cyclin Dependent Kinase 6
TP53	Tumor Protein p53
RAS	Rat Sarcoma Virus
RAF	Rapidly Accelerated Fibrosarcoma
MAPK	Mitogen-activated Protein Kinase
mTOR	Mammalian Target of Rapamycin
AKT1	AKT serine/threonine kinase 1
PAAD	Pancreatic Adenocarcinoma
BRCA	Breast Cancer
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
NGS	Next Generation Sequencing
TERT	Telomerase Reverse Transcriptase

CHAPTER 1

INTRODUCTION

Cancer is a disease characterized by uncontrolled cell proliferation induced by genetic alterations. The impact of these alterations propagates into the molecular interaction network, changing signaling pathways and transcriptional regulation in the cell. Not all alterations equally contribute to a growth advantage of cancer cells. Some mutations are drivers, while others are passengers [1]. Whereas it is generally believed that passenger mutations do not bestow proliferative effects on the disease phenotype, their properties and possible roles are not fully understood [2]. Cancer genomics and evolution studies suggest that the accumulation of 'slightly' deleterious passenger mutations can slow cancer progression, and this could be exploited for therapeutic purposes [3]. Lately, another class of mutations was defined, dubbed "latent" or "mini-drivers" [4–6]. Even though not identified as drivers since the effect that latent drivers generate is marginal, when coupled with other activating mutations, latent mutations can additively intensify the signal. Their detection may help forecast cancer progression and improve personalized treatment strategies [5]. Curated driver genes and mutations have been deposited in multiple databases [7–9] and used to develop computational approaches to predict driver genes and driver mutations [10–15]. These methods, including frequency-based methods, subnetwork identification methods, and 3D mutation search methods, have been comprehensively compared [16–20]. One of the concerns with frequency-based approaches is that prohibitively large sample sizes are needed to identify infrequently mutated driver genes. Thus, in frequency-based approaches, there is a risk of generating biased results due to background mutation rates [21,22]. There are various resources and web servers that examine the effect of missense mutations on protein stability, protein-protein interactions, and the underlying molecular mechanisms [23,24]. However, frequency-based approaches fail in the identification of rare drivers which can be tissue-specific [25]. A recent multidimensional analysis of cancer driver genes in IntOGen showed that some drivers are cancer-wide whereas others are specific to a limited number of cancer types [12].

Even a single mutation in a gene can be considered as a prognostic marker and change the global genome and protein expression, eventually altering the signaling pathways [26]. However, it has been estimated that the contribution of a single driver mutation to cancer progression is very small and needs additional mutations over time [27]. Despite DNA repair, somatic mutations accumulate and different genotypes in individual tissues are generated. This mechanism, called ‘somatic mosaicism’, offers driver, or synergistic mutations an advantage in cancer cells [28]. Recently, the combination of single frequent mutations with a rare, or weak mutation in the *same* gene was shown to have a significant advantage in tumor progression and influence treatment response. These double mutations *in cis* in *PIK3CA* were shown to be more oncogenic, and more sensitive to inhibitors compared to a single mutation [29]. A recent work cataloged ‘composite mutations’ of *multiple* genes having more than one non-synonymous mutation in the same tumor [30]. Saito et al. demonstrated the functional implications of multiple driver mutations in the same oncogene with an emphasis on *PIK3CA* [31,32].

Tumorigenesis is a complex process that could be attributed to alterations in a group of genes [33]. Besides the mutations in the same gene, understanding synergistic or antagonistic interactions between the mutations in different genes is also vital to identify the relation between these gene groups, and tumor formation/progression. We inspect relationships among genes, including functional similarities (or differences), occurrence in the same or different pathways, and their location in the pathway, upstream (or downstream). Co-occurrence, or conversely, exclusivity, of two mutations in different genes (i.e., in trans) could be a random or a selected event in cancer evolution, providing a functional advantage. If two mutations in different genes occur less or more frequently than expected, they are mutually exclusive or co-occurring, respectively [34]. Mutual exclusivity was mostly observed on genes in the same pathway (or redundant pathways), and conversely, a co-occurrence trend was observed on genes in different (or parallel) pathways [35,36]. In our definition, if the pathways recruit the same downstream protein families, they are redundant; if evolutionary-independent, they are parallel. Exclusivity has been attributed to the expectation that acquiring a second mutation in the same pathway will have a negligible effect on tumor growth. It has been related to factors such as such as tumor subtype, synthetic lethality, and positive selection [37]. Alternatively, the additive effect of same- or different- pathway co-occurring driver hotspot mutations are likely to hyperactivate the proliferation signal. Potent and sustained hyperactivation can trigger an oncogene-induced senescence (OIS) cellular program [38–40].

Neurodevelopmental disorders (NDDs) arise from a dysfunctional nervous system during embryonic brain development. The origins of NDDs are still unclear. They may originate from dysregulation of neuron differentiation, during synapse formation and maturation, or other complex processes in the course of brain evolution, such as emergence from progenitor cells, neuron phenotypic specification, migration, and specific synaptic contacts. Flaws can result in faulty wired neuronal circuits [41,42]. Despite differing from processes associated with the emergence of cancer, data indicate that NDDs and cancer are related, with immunity likely common factor. The immune and nervous systems coevolve as the embryo develops [43]. The outcomes, cancer or NDDs, reflect the different cell cycle consequences, proliferation in cancer and differentiation in NDDs. Proliferation requires a stronger signal to promote the cell cycle than differentiation does. This further suggests that in addition to nodes in the major signaling pathways, transcription factors (TFs) and chromatin remodelers, which govern chromatin organization, are agents in NDDs. Gene accessibility influences the lineage of specific brain cell types at specific embryonic development stages [42].

The primary focus of this dissertation is identifying statistically rare driver mutations from double mutations in the same gene which can collectively promote tumorigenesis or induce drug resistance when they occur in *cis*. Then, cancer type-specific double mutations are identified and their effect on tumor growth or treatment response are evaluated computationally via the integration of pre-clinical models such as cell lines and patient derived xenografts. Next, epistatic relations between the mutations in different genes are examined. Mutually exclusive and co-occurring mutations are determined with a comprehensive statistical analysis; the results indicated that these epistatic relations largely depend on the redundant or synergistic roles of the different gene double mutation components. Besides, different gene double mutations that can be metastatic markers are proposed. Finally we employ mutation, transcriptomic and network reconstruction analyses to detect commonalities and differences between cancer and NDDs. Despite being very different phenotypic properties, cancers and NDDs converge on key mutations, proteins, and pathways.

The highlights addressed by these findings are as follows:

- Some statistically rare mutations can cooperate with other mutations in the same gene to promote cell proliferation or drug resistance.
- Oncogenic mutations in the long tails of the distributions are statistically rare and not yet labeled as oncogenic can be latent drivers.
- Latent drivers may be tissue, or cell specific, harbored in specific cancers.
- Mutations affecting the same pathway are usually mutually exclusive in a tumor to prevent functional redundancy, synthetic lethality and oncogene induced senescence, considering the additivity of the mutational impact.
- Co-occurring different gene double mutations can change transcriptional regulation and drug response of patient derived xenografts for some inhibitors.
- Rare doublets on PIK3CA and ESR1 mutations co-occur in metastatic tumors and might be metastatic markers.
- Our results provide a molecular explanation for PTEN and PIK3CA mutations frequently found in cancer and NDD samples, which may serve as the foundation for functional and in-depth structural research.
- Comparing expression scores of shared pathways by leveraging the transcriptomic profiles of NDDs and cancer samples revealed that NDD

samples have higher expression scores for genes functioning in differentiation than proliferation.

- Despite having common signaling pathways, their regulation and difference in signal levels enhance different cell states: Proliferation for cancer and differentiation for NDDs.

The dissertation consists of six main chapters, namely Introduction, “Literature Review”, “Discovery of Latent Driver Mutations from Double Mutations in Pan-cancer Data Reveal Their Clinical Impact”, “Co-occurring Mutations in Trans can be Metastatic Markers; Excluded Combinations can Encode Oncogene Induced Senescence”, “Neurodevelopmental disorders and cancer networks share pathways; but differ in mechanisms, signaling strength, and outcome”, and “Discussion”.

In Chapter 2, the rationale behind the double and latent driver mutations is presented; the studies that identify and investigate the function of such mutations are reviewed. The landscape of in *cis* mutations naturally extends to in *trans* mutations – the mutations in different protein molecules; various tools and databases investigating these in *trans* combinations are scrutinized. Then we investigate commonalities and differences between cancers and NDDs by utilizing mutations and network analysis.

In Chapter 3, latent driver mutations are identified from the mutation profiles of the tumors in pan-cancer cohorts of TCGA and AACR GENIE. A statistical approach is deployed to determine double mutations in the same gene and the components of double-mutations were labeled as latent mutations if their co-occurrence is significant and not yet labeled as a cancer driver. Co-occurring patterns that are predominantly present in specific tissues and tumor types are found. Tumor-type specific double mutations in the same gene which may promote tumorigenesis and alter the response to treatments are revealed. It is also revealed that tumors having at least one double mutation pair may lead to changes in response to drugs. Although the existence of a set of driver genes is considered cancer-wide, it is shown that having double mutations on those genes is cancer-specific. Same gene double mutations are relatively rare; however, their impact is elevated in tumor progression.

Chapter 4 focuses on finding the double mutations in different genes among the pan-cancer cohorts of TCGA and AACR GENIE with a statistical approach. The synergistic effect of different gene double mutations could elicit stronger oncogenic signal. Double mutations *in trans* that is in different genes on the protein-protein interface were rarely observed. However, some double mutations *in trans* were significantly higher in metastatic tumors. These could be evaluated as metastatic markers. *In trans* co-occurring double mutations can help in revealing alternative pathways in emerging drug resistance, and in identifying metastatic markers. These findings are supported with case studies for the cancer types they are enriched in.

In Chapter 5, we aim to uncover the shared features between neurodevelopmental disorders and cancer. We expect that these will help us understand the challenging question of how expression levels and mutations in the same pathways, and even the same proteins, including TFs and chromatin remodelers, can lead to NDDs versus cancer, with vastly different phenotypic presentations. Especially, we aim to discover what are the determining features deciding whether the major outcome is NDDs or cancer. We address this daunting goal by comprehensively leveraging mutations, transcriptomic data, and protein-protein interaction (PPI) networks.

In the final chapter of this dissertation, the main findings are summarized and discussed; in addition, some possible future perspectives are proposed.

CHAPTER 2

LITERATURE REVIEW

This chapter briefly discusses the background and literature related to this study. The literature review is addressed in six main sections: (1) genome sequencing and variant classes; (2) driver, passenger and latent driver mutations; (3) latent driver mutations from biophysical perspective; (4) oncogenes and tumor suppressor genes (5) double/multiple mutations and latent drivers and (6) co-occurring and mutually exclusive mutations in different genes.

2.1 Genome sequencing and variant classes

Cancer is a disease of uncontrolled cell proliferation driven by molecular alterations. The impact of these alterations diffuses into the molecular interaction network and changes signaling pathways and transcriptional regulation in the cell. Precision medicine efforts to identify actionable mutations that cause cancer in patients have incited big-data initiatives. They have encouraged the development of experimental methods for data collection using next-generation sequencing methods, and they have provided thousands of genomic profiles of primary and metastatic tumors [44–49].

Human genome is made up of ~3 billion base pairs, of which ~1% are translated into functional proteins [50,51]. These are the proteins where mutations are most likely to have an immediate phenotypic impact. Next-generation sequencing (NGS) techniques have allowed for the rapid sequencing of large amounts of DNA and the identification of single nucleotide, structural, or copy number variants that affect DNA sequence. The DNA regions covered by these techniques differ. Whole-genome sequencing (WGS) can reveal variation in any region of the human genome, including coding, noncoding, and mitochondrial DNA sections. It determines the order of each nucleotide in an individual's DNA. Whole-exome sequencing (WES) scans for variants in protein-coding regions (exons). Exons are the sections of the genome that make up exomes, which account for ~1% of the entire genome. WES captures low frequency alterations better than WGS because sequencing a nucleotide multiple times provides greater sequencing depth. Targeted panel sequencing is an alternative technique that focuses on a subset of genes or coding regions of genes known to contain mutations that contribute to disease pathogenesis, and it provides more depth than WGS/WES [52,53]. Due to the advancements in NGS techniques, several pan-cancer data sets have been released. The Cancer Genome Atlas (TCGA) contains molecular and clinical data from over 10,000 tumors from 33 cancer types that were obtained using WES [54,55]. The Pan-Cancer Project, also known as the ICGC/TCGA Pan-Cancer Analysis of Whole Genomes project, examined 38 different tumor types from over 2658 donors [49]. These datasets contain a wide range of genomic data obtained through whole exome/genome or targeted/panel sequencing methods. As a result, a massive amount of omics data, including but not limited to mutation, transcriptomic, phospho-proteomic profiles, and copy number variations, as well as clinical information, has been accumulated.

Several open-access servers or portals provide access to this diverse set of omics data [56–58]. These resources allow downloading pan-cancer, cancer tissue- or subtype-specific mutation data in *mutation annotation format* (MAF) which is a tab-delimited text file with aggregated mutation information from variant call files (VCF) upon availability. Missense single nucleotide variations, which change one amino acid residue at the protein level, cause a significant portion of all currently recognized human genetic diseases [59,60]. A nonsense mutation causes a protein to be prematurely terminated and produces a shorter isoform. As a result, this truncated form may not preserve the original function of the protein and may even completely inactivate it [61]. An insertion or deletion that involves base pairs that are not multiples of three results in a frameshift mutation; this type of mutation may also result in the creation of shortened proteins due to an early stop codon [61]. Duplication and deletion variations are genetic changes that result in the duplication or deletion of a portion of DNA. This may result in the creation of additional copies of a gene or the complete loss of a gene, which may affect the function of the protein encoded by the gene.

Genetic variants are variations in an individual's DNA sequence that can affect the structure and function of proteins. These variants can be inherited from a person's parents or acquired during an individual's lifetime. Genetic variants can be categorized into a wide range of categories, including single nucleotide polymorphisms (SNPs), insertions and deletions, and structural variations. Each category has a unique impact on the structure and operation of proteins, and these variations can influence the development and progression of various diseases. For example, genetic variants have been implicated in the onset of cancer, as well as other diseases such as cardiovascular disease, diabetes, and neurodevelopmental disorders [62]. Understanding the genetic basis of these diseases is important for developing effective treatments and preventive measures.

2.2 Driver, passenger, and latent driver mutations

Millions of somatic mutations in tumor cells have been identified using next-generation sequencing. Identifying the molecular changes that lead to tumor formation is a major challenge in the interpretation of cancer genomes. Not all mutations contribute equally to cancer cell growth advantage. Some mutations are the drivers, while others are the passengers. To identify potential drivers, frequency-based and function-based approaches have been developed [63–68]. Conventionally, frequent mutations are referred to as "driver mutations", and are thought to contribute to the development of cancer; in contrast, rare alterations, known as "passenger mutations," do not [5,25]. Many of the driver mutations identified by frequency-based methods also were experimentally validated as functionally important [69]. For the majority of cancer types, only a small number of genes are altered in a big proportion of tumors, whereas a far larger number of genes are only rarely modified. These investigations have so far uncovered about 140 genes that, when changed, have the potential to promote or "drive" cancer. A tumor normally contains two to eight of these "driver gene" mutations; the remaining mutations are "passengers" that do not confer a selective advantage for growth [1].

Although it is widely believed that passenger mutations do not provide any proliferative advantage to the tumor cells, little is known about their characteristics and potential functions. [2]. As per cancer genomics and evolutionary simulation studies, the accumulation of "slightly" damaging passenger mutations can slow the growth of cancer, which could be utilized as a therapeutic advantage [3]. But in the long-tail of mutation frequency distributions, there may also be low-frequency driver mutations in addition to high-frequency ones. Experimental studies suggest that a number of long tail mutations have functional significance; however, to prioritize long tail mutations for functional research and to quantify their functional relevance, complementary methodologies are required [69]. It is crucial to be able to spot rare drivers amid passenger mutations because both recurrent and rare drivers provide cancer cells a selective growth advantage [25].

Recently, a new category of mutations known as "latent" or "mini-drivers" was defined [4–6]. Latent mutations can cooperate with other activating mutations to enhance the signal, even though they are not recognized as drivers because of the minimal effect they have. Latent driver mutations behave like passengers and do not impart a cancer signature, however, they contribute to the development of cancer and drug resistance when paired with other new mutations. In addition to the conventional trend of labeling frequent mutations as drivers, the detection of latent drivers, which are rare drivers indeed, could enhance personalized treatment strategies and forecast the progression of cancer [5,25].

According to the "latent driver" hypothesis, mutations may eventually lead to protein ensembles populating a constitutively active state from the structural standpoint of protein molecules (or an inactive state for repressor proteins). Even if the extent of the shift caused by a pre-existing mutation is minor, this could explain why it was unreported. The emergence of a cooperative second mutation may drive the protein and cell to a constitutively active (or inactive) state with a cancer hallmark [5].

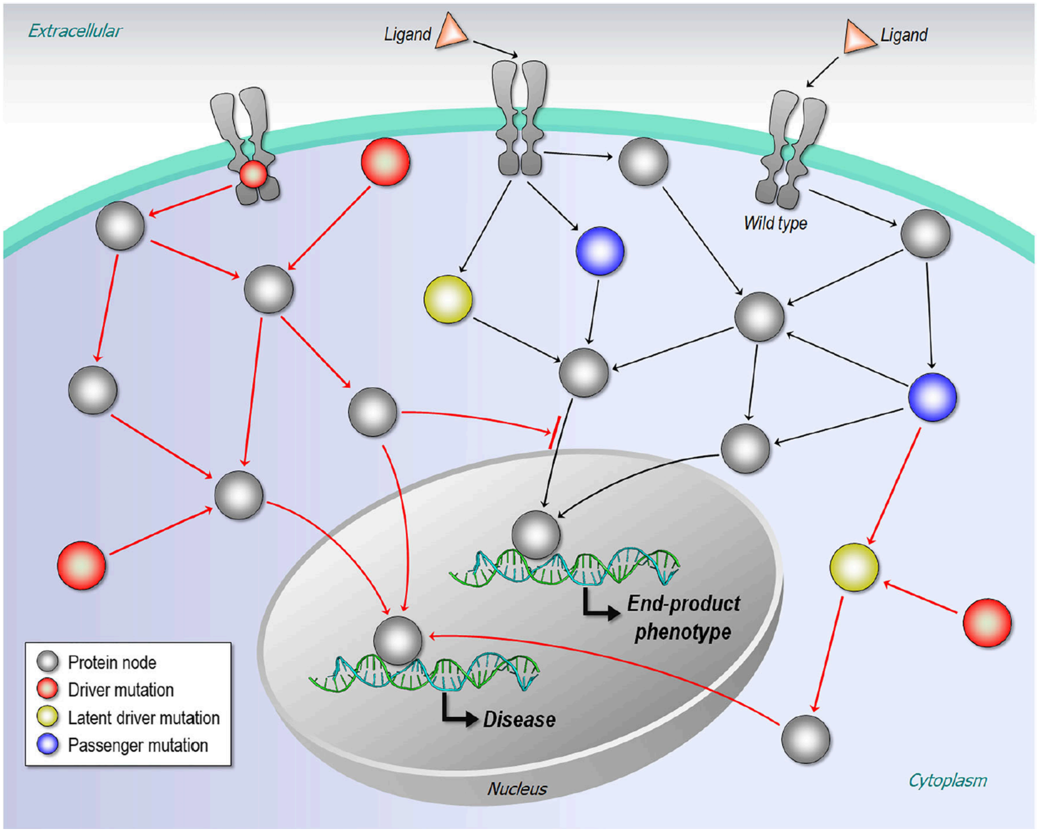

Extracellular
Ligand
Ligand
Wild type
End-product phenotype
Disease
Nucleus
Cytoplasm
Protein node
Driver mutation
Latent driver mutation
Passenger mutation

Figure 1. Mutations affect the cellular network. This figure is obtained from the article [70] which is distributed under the terms of the Attribution (CC BY) 4.0 International (http://creativecommons.org/licenses/by/4.0/).

Figure 1 depicts mutations in the cellular network that affect drivers, passengers, and latent drivers. Driver mutations (red nodes) can change the phenotypic outcome from normal to disease state on their own. A passenger mutation (blue nodes) does not result in cancer cells. Latent driver mutations (yellow nodes) can express a cancer cell phenotype when combined with a newly evolved mutation [70].

2.3 Latent driver mutations from a biophysical perspective

Proteins exist in conformational ensembles in which the inactive state is more populated, and they switch between conformations to perform their functions. Proteins can transition from inactive to active (or vice versa) states using three different mechanisms. If there are two allosteric stimuli, one of them may be sufficient to drive the transition between states. The combined effect of the two stimuli is not greater than the effect of the single stimulus. However, in some cases, both stimuli are required for full activation. All stimuli contribute gradually to full activation in a graded mechanism [5].

A pathogenic allosteric driver mutation can start (or stop) oncogenic signaling constitutively. When it comes to latent driver mutations, a graded activation/suppression mechanism is usually applied. Latent driver mutations boost signaling and increase the number of active (or inactive) conformations in the cell. Each new mutation raises the level of activity (or repression). A set of latent driver mutations or a subsequent concomitant external allosteric action is required for full activation. Since they are located at established functional loci, orthosteric driver mutations are easier to identify. The majority of driver mutations, however, occur at unknown allosteric sites. Allosteric rare/frequent driver mutations, like orthosteric driver mutations, can cause functional change [4].

Tissue-specific high- or low-frequency drivers switch the ensemble from inactive to active state. Rare drivers, on the other hand, are uncommon because, in the absence of further mutations, the protein activation they stimulate is insufficient to provide cancer cells with a selection advantage. When combined with other mutations, they can shift the protein from inactive to active state. This system guards against mutation-causing events.

In Figure 2 [25], Nussinov et al. provides a theoretical background for free energy landscapes of proteins as compared with the wild-type protein, frequent allosteric drivers, latent drivers, and multiple rare driver mutations. The wild-type protein populates the inactive state more than the active state (Figure 2A). The emergence of an allosteric driver mutation increases the number of active conformations in the population (Figure 2B). A single latent driver mutation is not capable of shifting populations into the active state (Figure 2C). However, when there are multiple rare driver mutations, the active state is more densely populated than the inactive state, which is very similar to the case of an allosteric driver mutation (Figure 2D).

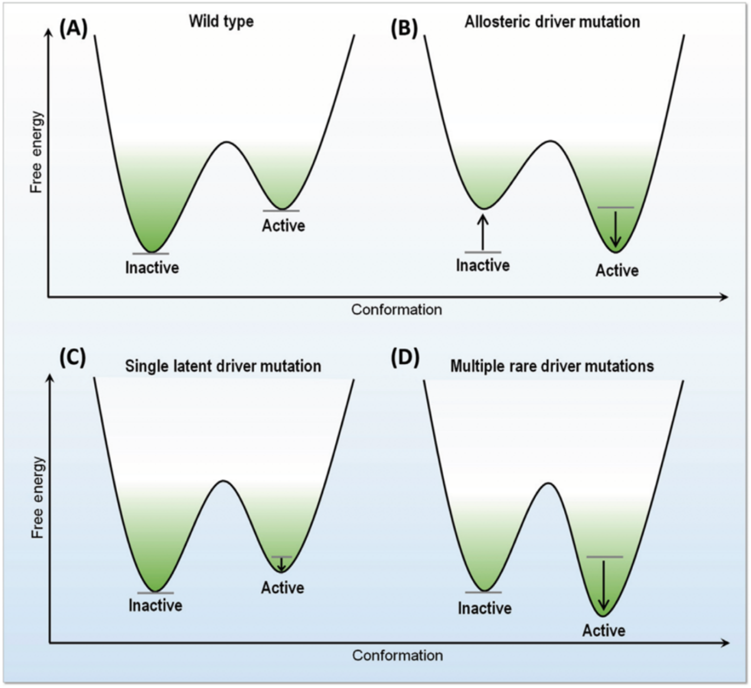

(A)
Wild type
(B)
Allosteric driver mutation
Free energy
Inactive
Active
Inactive
Active
Conformation
(C)
Single latent driver mutation
(D)
Multiple rare driver mutations
Free energy
Inactive
Active
Inactive
Active
Conformation

Figure 2. Free energy landscapes of conformational switches for different mutation types. This figure is obtained from the article [25] which is distributed under the terms of the Attribution (CC BY) 4.0 International.

2.4 Oncogenes and tumor suppressor genes

Human neoplastic disorders arise and spread through a multi-step process involving the accumulation of genetic alterations in somatic cells. Tumorigenesis is characterized by genetic changes that affect important cellular processes involved in growth and development. These genetic changes frequently involve the inactivation of tumor suppressor genes (TSGs) and the activation of collaborating oncogenes (OGs), both of which appear to be required for the overall neoplastic phenotype [71,72]. Suppressed TSGs and overactive OGs contribute significantly to cell proliferation and apoptosis during cancer formation via genetic variations such as somatic mutations and deletions [73]. Oncogene activation and tumor-suppressor gene (TSG) inactivation are two key factors that cause cancer [74,75].

Proto-oncogenes, which are oncogenes before mutations appear, regulate normal cell division. When mutated, an oncogene has the potential to cause cancer. Based on protein sequences, heuristic approaches, such as the 20/20 rule, classify a gene as an oncogene if more than 20% of its mutations occur at a specific residue [76,77]. Oncogenes are normally involved in the regulation of cell growth and division, and when they are mutated or overexpressed, they can cause cells to grow and divide uncontrollably, leading to the formation of a tumor [78]. Oncogene products can be made up of a wide range of molecules, including transcription factors, chromatin remodelers, growth factors, growth factor receptors, signal transducers, and apoptosis regulators, all of which are crucial for the development of tumors. Well-known examples of oncogenes include PIK3CA, RAS, WNT, MYC, ERK, and EGFR [79].

A tumor suppressor gene is a type of gene that inhibits cell growth and division, thereby aiding in cancer prevention. Because they are enriched in protein-truncating mutations, Vogelstein et al. proposed that genes with more than 20% truncating events can be considered as TSGs [1,80]. TSGs typically function by encoding proteins that inhibit cell division or cause cell death in response to DNA damage or other cellular stress. When a tumor suppressor gene is mutated or inactivated, cells can continue to divide and grow uncontrollably, leading to the development of a tumor.

The functions of TSGs are broadly distributed across various cellular processes, including signaling pathways, chromatin remodeling, and DNA damage, suppressing metastasis and repair processes among others [81]. The loss of function for a tumor suppressor may lead to cancer due to uncontrolled cell division. Because of their importance, extensive studies have been undertaken to understand the different functional mechanisms of tumor suppressors [82]. The two-hit model has been commonly applied to study the roles of TSGs in cancer [83,84]. Unlike oncogenes, most TSG loss-of-function mutations act recessively in nature. In contrast to haploinsufficient TSGs, which function dosage-dependently, a TSG must presumably have mutations in both copies of the gene in order to completely lose its activity [75]. When one copy of a gene is deleted or mutated, the amount of normal product produced by the single wild-type gene is insufficient for full function, resulting in haploinsufficiency, a condition in which two wild-type copies of a gene are required for a normal phenotype [85]. TSGs that are well known include TP53, PTEN, STAG1, NOTCH1, and CDKN1A.

Tumor suppressor genes (which prevent tumorigenesis and can be inactivated by mutations or deletions) and proto-oncogenes (which promote proliferation and can be activated by mutations such as amplifications or chromosomal translocations) are the two primary types of cancer driver genes [86]. As of September 2022, OnkoKB knowledge base deposited 1081 cancer genes of 275 are OGs, 275 are TSGs, and 40 are both OG and TSG [87]. Yet another knowledge base, TSGene version 2.0, contains 1217 human TSGs (1018 protein-coding and 199 non-coding genes) curated from over 9000 articles and pan-cancer genomic data collection [81].

2.5 Double/multiple mutations and latent drivers

Recent developments in cancer genomics have made it possible to identify and characterize numerous cancer genes and the associated mutations that are related to cancer initiation, sustain the disease once it has started and potentially alleviate it [1,14,32,88]. Scientists have been putting a tremendous effort for identification and functional characterization of specific driver mutations, which have revealed a variety of phenotypes and therapeutic vulnerabilities. These "driver" mutations, which are characterized as frequent mutations that give a cell a selective growth advantage, promote clonal expansion and the development of disease [30,89,90].

Even a single mutation in a gene can be considered as a prognostic marker and change the global genome and protein expression, eventually altering the signaling pathways [26]. However, it has been estimated that the contribution of a single driver mutation to cancer progression is very small and needs additional mutations over time[27]. Despite DNA repair, somatic mutations accumulate and different genotypes in individual tissues are generated. This mechanism, called ‘somatic mosaicism’, offers driver, or synergistic mutations an advantage in cancer cells[28]. Recently, the combination of single frequent mutations with a rare, or weak mutation in the *same* gene was shown to have a significant advantage in tumor progression and influence treatment response. These double mutations *in cis* in PIK3CA were shown to be more oncogenic, and more sensitive to inhibitors compared to a single mutation [29]. A recent work cataloged ‘composite mutations’ of *multiple* genes having more than one non-synonymous mutation in the same tumor [30]. Saito et al. demonstrated the functional implications of multiple driver mutations in the same oncogene with an emphasis on PIK3CA [31,32].

However, to date, driver mutations whose frequencies are low and their observable translational potential are minor have escaped identification. Recently, several investigations have found evidence of multiple or double mutations on the same gene in the same allele (in *cis*). In these multiple or double mutations on the same gene, rare, functionally weak mutations are convergently selected, which may offer an explanation for why such mutations accumulate in tumor cells [29,31,32,91].

Several groups have recently investigated epistatic relationships for frequent and rare driver mutations. Two driver mutations can have synergistic, additive, or suppressive functional consequences. If two mutations are driving tumors independently, they should be additive. However, the functionality of two mutations can be higher (synergistic) or lower (suppressed) than the functionality of a single mutation, indicating a genetic interaction between mutations. The term "epistasis" best describes these genetic interactions; in fact, several recurrent multiple mutations have synergistic functions, indicating positive epistasis [32]. Synergistic epistasis can occur when a conformation change caused by one mutation is required for a second mutation to realize its effect on protein function (conformational epistasis). In such cases, the secondary mutated residue may contact a novel substrate only when the first mutation causes a conformational change. Another explanation includes the stability thresholds in which the mutation has a detrimental effect only when it reduces the protein stability below a critical threshold. In this scenario, single mutations have little effect, but two mutations reaching the critical threshold have a very deleterious effect [92–94].

Recently, the combination of single frequent mutations with a rare, or weak mutation in the same gene was shown to have a significant advantage in tumor progression and influence treatment response. These double mutations in *cis* in PIK3CA were shown to be more oncogenic, and more sensitive to inhibitors compared to a single mutation.

A recent work cataloged 'composite mutations' of multiple genes having more than one non-synonymous mutation in the same tumor. Saito et al. demonstrated the functional implications of multiple driver mutations in the same oncogene with an emphasis on PIK3CA. Multiple mutations in a single gene rarely co-occur in patient tumors. However, when they are together, they may cause dramatic phenotypic differences and can be signatures of specific tumor tissues or cancer types [29–31]. For example, double mutations in PIK3CA increase the sensitivity to PI3K inhibitors in breast cancer [29], while double mutations in EGFR predominantly exist in lung cancer [31].

The comprehensive investigation of pan-cancer genomes by Saito et al. revealed multiple driver mutations in the same oncogene in cancer patients. They discovered 14 oncogenes that are shared by all cancer types and 6 that are unique to specific cancer types. Most of these in *cis* multiple mutations were functionally weak, infrequent mutations that conferred enhanced oncogenicity when combined. These mutations also influence carcinogenesis, have positive epistatic relationships, and are linked to increased drug sensitivity. Researchers discovered enhanced downstream signaling activation and/or greater sensitivity to inhibitory treatments in related cells with multiple mutations compared to those with single mutations. They proposed that the primary mechanism for oncogenic multiple mutations, which constitute a frequently reported driving event, is a clonal selection of suboptimal mutations, which are uncommon individually but account for a significant percentage of oncogenic mutations collectively [31,32].

2.6 Co-occurring and mutually exclusive mutations in different genes

Tumorigenesis is a complex process that could be attributed to alterations in a group of genes [33]. Multiple activating mutations are needed to initiate cancer or promote tumor progression [1,95]. For example in colorectal cancer, loss of function mutations in APC needs TP53 or KRAS mutations to complete malignant transformation [86,96]. Random mutations are sequentially selected during tumor evolution if they provide a fitness advantage over the underlying genomic landscape [37]. As we reviewed in Section 2.5, these activating double/multiple mutations can be in the same protein molecule. Recent studies have shown that *cis* mutations are also positively selected in cancer genomes and can potentiate oncogenic signaling by the additive or synergistic effect of constituent mutations [29–31].

There are also scenarios where multiple mutations can co-exist in different genes that participate in the same or different pathways. Co-occurrence, or conversely, exclusivity, of two mutations in different genes (i.e. *in trans*) could be a random or a selected event in cancer evolution, providing a functional advantage. If two mutations on different genes occur less or more frequently than expected, they are mutually exclusive or co-occurring, respectively [34]. Mutual exclusivity was mostly observed on genes in the same pathway (or redundant pathways), and conversely, a co-occurrence trend was observed on genes in different (or parallel) pathways [35,36]. Nussinov et al. define redundant and parallel pathways as follows: if the pathways recruit the same downstream protein families, they are redundant; if evolutionary-independent, they are parallel [35]. Exclusivity has been attributed to the expectation that acquiring a second mutation in the same pathway will have a negligible effect on tumor growth. It has been related to factors such as tumor subtype, synthetic lethality, and positive selection [37]. Alternatively, the additive effect of same- or different-pathway co-occurring driver hotspot mutations are likely to hyperactivate the proliferation signal. Potent and sustained hyperactivation can trigger an oncogene-induced senescence (OIS) cellular program [38–40]. Moreover, there are some exclusive gene pairs when one of them mutated, the other needs to be in wild-type form to boost oncogenic signaling ("synthetic essentiality") [86].

OIS suppresses cell proliferation caused by aberrant activation of oncoproteins in normal cells. In the literature it is often discussed as caused by activated RAS and RAF proteins, especially the $BRAF^{V600E}$ hotspot [97]. RTK-RAS pathway members NRAS and BRAF are mutually exclusive in the TCGA skin cutaneous melanoma cohort. Harboring a mutation on one of the genes can promote tumor progression by amplifying signaling output. Similarly, *BRAF* mutation and the *RAS* mutation are mutually exclusive among metastatic colorectal cancer tumors [97,98]. OIS [99–101], a tumor-suppressive mechanism arresting cell cycle progression, can be a main reason for the exclusive co-occurrence of driver mutations in the same or different pathways. Senescence phenotypes depend on strong oncogenic stimulus [102]. Overexpression of oncogenes such as RAS, BRAF, and MYC induces OIS, as does increased PI3K/AKT signaling. Loss of PTEN, which negatively regulates the PI3K/AKT pathway, can trigger OIS through a p53-dependent pathway [103]. Sustained hyperactivation of the PI3K/AKT/mTOR pathway results in cellular senescence [104,105]. Cisowski et al. observed that co-expression of $\mathrm{KRAS}^{\mathrm{G12D}}$ and $\mathrm{BRAF}^{\mathrm{V600E}}$, two strong drivers, led to oncogene-induced senescence in their lung cancer study on mouse models [106]. Examples of mutually exclusive relations between gene pairs include BRCA2-TP53, BRCA1-PARP1, PTEN-PIK3CA in breast cancer; in ovary cancer BRCA1-CCNE1, BRAF-KRAS, ERBB2-KRAS [107].

There are several tools for identifying biologically meaningful, mutually exclusive gene sets. Liu et al. introduced a statistical framework MEScan, that takes into account the background mutation rate of highly mutated genes in the genomic data, hence accurately identifies tissue specific pan-cancer mutually exclusive genes [108]. MEGSA (mutually exclusive gene set analysis) is another tool relying on the likelihood ratio test and a model selection procedure [109]. The statistical tool Mutual Exclusivity Modules (MEMo) integrates multiple data types to identify driver modules in cancer pathways [110]. Co-occurring Mutated Driver Pathways (CoMDP) is a mathematical method identifying significantly co-occurring driver pathways from mutation data [111]. Mina et al. developed an algorithm SELECT, that captures evolutionarily dependent alterations at the pan-cancer level that can affect drug resistance and sensitivity [112]. Mutual Exclusivity Module Cover (MEMCover) identifies mutual exclusivity hubs and pan-cancer dysregulated networks across tissues, avowedly correlating to cancer drivers with significant growth advantages [113].

Complex biological functions are mostly managed through protein-protein interactions (PPIs), and mutations occurring in protein-protein interfaces cause changes on the interactome, resulting in different phenotypes [114]. Evaluating the effects of single or coexisting mutations in PPIs is crucial to understand the molecular mechanisms that lead to tumor formation and/or progression. It is also vital for understanding the mechanisms of network rewiring and its contribution to drug resistance/sensitivity. Alterations perturbing PPIs can be highly related to patient survival and drug resistance/sensitivity [4,68,115]. Li et al. identified ~260 PPIs related to cancer and OncoPPI based STK11-CDK4 interaction demonstrated to be sensitive to the palbociclib, a CDK4 inhibitor. Similarly, Cheng et al. identified 470 oncoPPIs affecting patient survival and drug resistance/sensitivity [115]. Detection of druggable PPIs is promising since some are cancer-specific and could be therapeutic targets [116,117]. Several mutations are classified as metastatic markers in gastric carcinomas of Fujian cohort. One of these is PTPRT mutations [118]. Chondrosarcoma TERT promoter mutation is a metastasis marker [119]. A database, My Personal Mutanome (MPM) deposits ~500,000 mutations that are mapped to ~300,00 functional sites and ~100,000 structure-resolved/predicted PPIs, including ligand-protein binding sites and eight different posttranslational modifications (PTMs). For these mapped interactions, 8884 survival results and 1,271,132 drug treatment responses are also available for precision oncology efforts [120].

2.7 Connection between the neurodevelopmental disorders and cancer

Recent epidemiological studies on large cohorts with neurodevelopmental disorders (NDDs) demonstrated an increased risk for cancer compared to the general population. In one study, a standardized incidence ratio model was applied to a cohort of 8438 patients with autism retrieved from the Taiwan National Health Insurance database during 1997-2011 [121]. Increased cancer risk was also observed in a population-based study among 2.3 million individuals with autism spectrum disorder (ASD) from Nordic countries during 1987-2013 with co-occurring birth defects, including intellectual disability [122]. A correlation between autism and cancer with shared risk factors was also pointed out [123]. Another cohort study proposed that patients with bipolar disorder and their unaffected siblings have an especially higher risk of breast cancer compared to normal control groups [124]. Association between brain, hepatocellular, and lung cancer among people with epilepsy was manifested by animal experiments, genotoxicity studies, and epidemiological observations. Possible underlying mechanisms have also been proposed [125,126].

Comparisons of the time windows of NDDs and cancer frequently concluded that while cancer is predominantly caused by somatic mutations and alterations in signaling and transcriptional programs, NDDs are primarily linked to mutations that occur during embryonic development. A recent study has similarly suggested that mutations in cancer susceptibility genes are not necessarily inherited or somatic; they can also arise throughout embryogenesis as a result of errors occurring during cell division

[127]. These *mosaic mutations*, occurring in early embryogenesis, were suspected to be associated with some rare cancers. Genetic changes associated with RASopathies are believed to be often sporadic, not inherited. Along these lines, according to the NCI page [61], this means that typically multiple family members do not share the same syndrome.

Different than this view, here our thesis is that inherited and *de novo* mutations (missense or truncation) are major causes of NDDs such as intellectual disability, ASD, epilepsy [128–131], and cancer [5,70,132]. As in cancer [132], more than one mutation is required for observable symptomatic NDDs. Our premise is that family members can harbor these embryonic mutations; however, they are not diagnosed as having the disorder. Their offsprings are however already susceptible to it. The presence of multiple mutations may also explain the phenotypic overlaps of the disorders. Further, individuals with NDDs have somewhat higher probabilities of eventually coming down with cancer, likely due to the preexistence of the mutations in the shared proteins, making them more susceptible. We hypothesized that strong driver mutations in cell growth/division pathways are chiefly responsible for uncontrolled cell proliferation in cancer. NDDs' weak/moderate strength mutations may be a reason why an inherited NDD has not been identified in a parent while predisposing an offspring to it. An additional mutation promotes NDD clinical manifestation. It may be inherited from the other parent or emerge during embryogenesis.

Tumor suppressor phosphatase and tensin homolog (PTEN), which carries inherited (germline) and *de novo* mutations in NDD patients, is related to cancers and several NDDs, collectively named PTEN hamartoma tumor syndrome (PHTS). The NDDs include phenotypes, such as Cowden syndrome (CS), Bannayan-Riley-Ruvalcaba syndrome (BRRS), Proteus syndrome (PS), Proteus-like syndrome (PSL), macrocephaly, and ASD. Besides, NDDs often overlap mutation-wise and genome-wise [133–137]. Among these, deletions and duplications of the 16p11.2 region are common. About 48% of deletion carriers and 63% of duplication carriers have at least one psychiatric diagnosis [138,139]. The region is associated with a variety of psychiatric conditions. RASopathies, which include Noonan syndrome (NS), cardiofaciocutaneous (CFC) syndrome, neurofibromatosis type 1 (NF1), and Legius syndrome (LS), are NDDs that result from the overactivation of the Ras/mitogen-activated protein kinase (MAPK) pathway due to the germline mutations and (or) overexpression in embryogenesis [62,140–143]. Their phenotypic overlaps may emerge due to shared proteins/pathways as in the case of *PIK3CA*-related overgrowth spectrum (PROS), PS, and CS which share phenotypic characteristics with RASopathies [141]. The commonality of proteins/pathways playing roles in cancer and RASopathies prompted consideration of mitogen-activated protein kinase (MEK) inhibitors and Ras-targeted therapies for some of the RASopathies like selumetinib for the treatment of some patients with NF1 [144–146]. Early-onset cancers, including breast cancer, and mutations occurring during embryonic development, such as in *BRCA1/2*, were also observed to be related [147,148]. Another mutated gene associated with these conditions and White-Sutton syndrome is *POGZ* gene that encodes pogo transposable element derived with the zinc finger domain, which disrupts brain development [149,150].

Cell proliferation and differentiation take place in both cancer and NDDs. Since NDDs are mostly related to dysregulated differentiation, mutations in genes regulating chromatin organization rank high. MAPK is primarily responsible for division and is the major pathway in proliferation. The phosphoinositide 3-kinase/protein kinase B/mammalian target of rapamycin (PI3K/AKT/mTOR) pathway primarily acts in cell growth, the primary pathway in differentiation [151]. Embryonic mutations in both pathways give rise to several NDDs [152]. Hundreds of genes are implicated in NDDs; however, they are involved in a few conserved pathways regulating transcription, chromatin accessibility, and synaptic signaling [153]. In particular, PI3K/mTOR and Ras/MAPK are frequently linked with NDDs and synaptic dysregulation [41]. Proteins in the Wnt, BMP/TGF-β (bone morphogenetic protein/transforming growth factor β), SHH (sonic hedgehog), FGF (fibroblast growth factor), and RA (retinoic acid) pathways, are also involved in autistic brain development. To identify similarities in genes and pathways, Forés-Martos et al. compared the gene expression profiles of 22 cancer types with the frontal cortical tissues from ASD patients with a correlation analysis [154].

NDD data has expanded recently, particularly *de novo* mutation data obtained by trio-sequencing and publicly available databases, however, it is still not as prevalent as the whole exome/genome sequencing data for cancer [155,156]. 32,991 *de novo* variants obtained from 23,098 trios are deposited in denovo-db. According to the database definition, *de novo* mutations are germline *de novo* variants present in children but not in their parents. The Deciphering Developmental Disorders (DDD) Study provides detailed genotype-phenotype information for 14,000 children with developmental disorders, and their parents from the UK and Ireland. Additionally, there are some knowledge databases with curated sets of genes and variants associated with one/multiple developmental diseases or cancer [157,158].

CHAPTER 3

DISCOVERY OF LATENT DRIVERS FROM DOUBLE MUTATIONS IN PAN-CANCER DATA REVEAL THEIR CLINICAL IMPACT

In this chapter, aided by informatics techniques, somatic mutations in pan-cancer cohorts of TCGA and AACR GENIE across ~60,000 patient tumors are screened and co-occurring patterns that are predominantly present in specific tissues and tumor types are identified. As a result, tumor-type specific double mutations in the same gene which may promote tumorigenesis and alter the response to treatments were identified. It also reveals that tumors having at least one double mutation pair may lead to changes in response to drugs. The components of double mutations were cataloged as latent mutations if their co-occurrence is significant and not yet labeled as a cancer driver and this led to uncover 140 latent driver mutations. The oncogenic activation of the protein may be through a single, or multiple additive contributions of the mutations. Although the existence of a set of driver genes is considered cancer-wide, it is shown that having double mutations on those genes is cancer-specific. Same gene double mutations are relatively rare; however, their impact is elevated in tumor progression. The results of this chapter are published under the title “Pan-cancer clinical impact of latent drivers from double mutations” in 2023 [159].

3.1 Methods

3.1.1 Data collection and processing

All available somatic missense mutation profiles are downloaded from The Cancer Genome Atlas (TCGA) and the AACR launched Project GENIE (Genomics Evidence Neoplasia Information Exchange) [45,56,108]. The TCGA mutation annotation file contains more than 10,000 tumors across 33 different cancer types. We used the merged MC3 file to get TCGA pan-cancer data. The somatic variants without sufficient normal depth coverage and variants found in the panel of normal samples were evaluated as possible germline variants and were removed from the file before merging.

The GENIE mutation file (Release 6.2-public) contains 65,401 patients and 68,897 tumor samples across 648 cancer subtypes under the Oncotree classification. Within the GENIE cohort 2930 patients match with multiple tumor barcodes. For those cases, only one primary tumor barcode is kept when available; if not, only one metastatic or unspecified tumor barcode is kept for further analysis without any other constraint. Among these patients, 2019 has sequenced primary tumors, 757 have sequenced metastatic tumors and the remaining 154 have tumors of the type not specified.

Next, we selected non-synonymous mutations including missense, nonsense, nonstop and frameshift mutations (altering only one position on a protein). We also excluded the mutations where the wild type and/or mutant residue name is not specified. As a result of this filtering process, 9703 and 57,921 tumors remained with a total of 1,631,755 point mutations in the TCGA and GENIE cohorts, respectively.

We did a pre-filtering on the VAF (Variant Allele Frequency) value to control the heterogeneity of the samples to some extent given that variants were collected by bulk sequencing in both datasets. We calculated VAF by using the ratio of the values in the t_alt_count and t_depth columns of the MAF (Mutation Annotation File) file of the pan-cancer data sets. Then, we continued our analysis with the mutations that have VAF value more than 0.125, ensuring that the remaining mutations are present roughly in 25% of the sequenced cells. We continued the analyses with 62,567 samples (9588 and 52,979 samples from the TCGA and GENIE cohorts, respectively) from 619 cancer subtypes and 33 tissues (including OTHER category).

3.1.2 Statistics and reproducibility

We set pre-filtering criteria to find significant double mutations. This pre-filtering consists of total number of occurrences and VAF values of each individual mutation. We construct potential double mutations to be tested after prefiltering. Therefore, it is independent of the test statistic under the null hypothesis[71,72]. If an individual mutation is present in less than three tumors in the cohort and have a VAF less than 0.125, we filtered them out. We continued our calculations with the remaining 65,872 mutations on 12,724 genes, and for each gene and the mutations they are harboring in the final set we formed binary combinations. As a result, we obtained 2,230,203 potential double mutations to be tested in 62,567 tumor samples that have at least one point mutation with VAF>0.125 and assessed their statistical significance (Fisher's Exact Test). For each potential double mutation, we created a contingency table [[a,b],[c,d]] where a is the number of tumors having both alterations, b is the number of tumors having only the first alteration, c is the number of tumors having only the second alteration and d is the number of all tumors not having these two alterations together (d = 62,567-(a+b+c)).

We applied more filtering for the significant double mutations based on the nonsense mutation composition among double mutants and the VAF values of the constituents. Throughout our analyses, we assumed point mutations occur at the same position as same regardless of the mutant residue. We evaluated the VAF values of the components and the presence of a nonsense mutation in the upstream in a tumor-specific way for the significant double mutations and double mutant tumors. In the components of a doublet, despite having a mutation at the same position, the mutated amino acid may result in a missense, nonsense or frameshift mutation. Therefore, a double mutation can be one of the combinations of the following variant classes: missense+missense, missense+frameshift, missense+nonsense, frameshift+frameshift, nonsense+nonsense. Among the tumor barcodes having a double mutation, if at least half of the barcodes carries a nonsense mutation as a component of a doublet we filtered them from our dataset.

To inspect whether double mutation constituents are in the same set of sequenced cells in a tumor, we first calculated the total VAF value of double mutation components. If the total VAF value is greater than 0.5, mutation components encompass >100% of the sequenced cells, which is impossible unless there is an overlap. Therefore, we labeled the mutation constituents as highly likely overlapping for such records. We retained the double mutations where the constituents overlap for at least 20% of the records for further inspection and kept 7252 significant double mutations where 155 of them are present in at least three tumors.

We used the Catalog of Validated Oncogenic Mutations from the Cancer Genome Interpreter [9] to label double mutation components: if a mutation is among the 5601 driver mutations, we labeled it as known driver (D), otherwise potential latent driver (d). For each gene harboring at least one double mutation, we collected all the tumors with mutations present on at least 3 tumors as gene-mutant tumors. Then, we calculated the fraction (%) of tumors with double mutation components among the gene-mutant tumors. We classified a known driver mutation as a strong driver if it is present in more than 10% of the gene-mutant tumors; otherwise, it is a weak driver. Similarly, we dubbed a potential latent driver mutation as a strong latent driver if it is present in more than 1% of the gene-mutant tumors; otherwise, we classified it as a weak latent driver. Here, we considered mutations in each gene present in at least three tumors when generating gene-mutant tumor group.

Additionally, double mutations are annotated based on their functions, domains, chemical properties, and structural proximity (see Appendix B).

3.1.3 Hyper-mutated samples and double mutations

First, we listed all non-hyper-mutated tumors that have at least one mutation on the 54 genes carrying at least one double mutation. Then, for each double mutation, we noted the total number of non-synonymous mutations on these tumors and labeled the double mutant tumors as Double and the remaining gene-mutant tumors as Single.

To test the null hypothesis that the double mutant tumors (Double) have a lower or equal mutation burden compared to the remaining gene-mutant tumors (Single), we applied a permutation test ($p<0.01$) with 5000 iterations. We prepared a two-column table having the Double/Single group labels of the tumors in the first column and the total number of non-synonymous mutations in the second column for each double mutation. To compare the observed and expected mean mutation counts for the two tumor groups, we shuffled the group labels in the first column 5000 times by preserving the second column as is. Here, we set the test statistic for two groups as follows: Test Statistic= μ(Double) – μ(Single) where μ is the mean mutation count. We calculated the permuted test statistic at each iteration by shuffling the

Double/Single labels. At the end of 5000 iterations, we counted the number of iterations where the permuted test statistic is greater than the original test statistic (N) and found the p-value by N/5000.

3.1.4 Allelic configuration of double mutations

We exploited supplementary data files of the papers [29–31] to check *cis/trans* status of double mutations for the matching samples.

3.1.5 Mutational signature analysis

We used 96 mutation contexts deposited in COSMIC that the format of codons and putative substitutions is as follows: $C_1[C_2>C_2^{subs}]C_3$ where C_i is the nucleotide in the corresponding position for i=1,2,3 and $C_2>C_2^{subs}$ indicates the wild type nucleotide C_2 is substituted by C_2^{subs}. We assumed double mutations are of the same context either they have the same base pairs in $C_1[C_2>C_2^{subs}]C_3$ at the same position or C_1, C_2, C_2^{subs} and C_3 are mapped to the opposite strand with the same ordering[160].

3.1.6 Cell line network construction

We obtained a list of cell lines with the dual mutations from Cell Model Passports and their drug response information from CancerrxGene [161,162]. We also extracted information about drug targets and target pathways. We used 2 different approaches to select drugs for PTEN, APC, and PIK3CA double mutant cell lines: if a drug is in the gray zone (|z-score|<=2) in the single mutant cell lines but gives a significant drug response in a double mutant cell line (|z-score|>2). If there is a single mutant cell line that is sensitive (or resistant) to the drug but the dual mutant cell line gives an opposite response to the drug. (Drug response flips sensitive into resistant or resistant into sensitive between single and dual mutant cell lines). For EGFR we selected drugs that give significant drug response either in the single or double mutant cell line. Then we formed networks connecting mutations to cell lines, cell lines to drugs, and drugs to their target pathways.

3.1.7 Patient-derived xenograft analysis

We used the mutation profiles, transcriptomic data and drug responses of patient-derived xenografts in[163]. We determined xenografts harboring significant doublets. Then, we compared changes in tumor volumes of single and dual mutant xenografts for the untreated and drug-treated cases (single mutation is part of a significant dual mutation). We preferred to specify the time intervals in multiples of 5. When a given timepoint is not a multiple of 5, we used linear interpolation between two nearest numbers containing a multiple of 5.

$$Vol_i = Vol_{i-1} + \frac{t_i - t_{i-1}}{t_{i+1} - t_{i-1}}(Vol_{i+1} - Vol_{i-1}) \quad (2)$$

where t_i is a timepoint that is multiple of 5 between the given timepoints t_{i-1} and t_{i+1} and Vol_i is the volume (mm^3) at timepoint i.

3.1.8 Availability of data and materials

The results shown here are in whole or part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga. The authors would like to acknowledge the American Association for Cancer Research and its financial and material support in the development of the AACR Project GENIE registry, as well as members of the consortium for their commitment to data sharing. Interpretations are the responsibility of the study authors. The cell line data underlying the results presented in the study are available from GDSC in https://www.cancerrxgene.org/downloads, Cell Model Passports in https://cellmodelpassports.sanger.ac.uk/downloads, and The Cancer Dependency Map project in https://depmap.org/portal/download/. The PDX data underlying the results presented in the study are available in Gao et al[163]. Codes are available at https://github.com/bengiruken/LatentDriverDiscovery.

3.2 Results

3.2.1 Discovery of latent drivers through double mutations

Multiple mutations in a single gene rarely co-occur in patient tumors. Vasan et al. examined the PIK3CA-mutant cancer genomes and reported that 12 to 15% of breast cancers and other tumor types harbor multiple PIK3CA mutations, the majority of which (95%) are double mutations [29]. Similarly, Saito et al. performed a pan-cancer study to check the presence of multiple mutations in a subset of oncogenes among ~60,000 tumors. They discovered 20 oncogenes with a higher rate of multiple mutations than expected where 9% of samples with at least one mutation in these oncogenes had multiple mutations[31,32]. Despite their relative scarcity, when multiple mutations are together in the same gene, they may cause dramatic phenotypic differences and can be signatures of specific tumor tissues or cancer types[29–31]. For example, double mutations in *PIK3CA* increase the sensitivity to PI3K inhibitors in breast cancer[29], while double mutations in *EGFR* predominantly exist in lung cancer[164]. We defined latent driver mutations as mutations that due to their unobservable translational or structural effects, have not been associated with tumor development. However, when combined with other alterations, can contribute to cancer progression and drug resistance [5]. Some mutations cataloged as passengers may belong to this category. The collective action of latent driver mutations in oncogenes (OGs) can switch the protein ensemble to an active state; in tumor suppressor genes (TSGs) the inactive state. When the mutations are on the same allele (i.e., *in cis*), a latent driver mutation could couple with driver mutations; two or more latent driver mutations can also collaborate (Figure 3). In either case, the outcome can be a stronger effect. Along similar lines, a strong driver may couple with a weak driver or a latent driver, strengthening the pathological impact. Our definition of latent mutations applies only to mutations *in cis*. That is, in the same protein molecule (i.e., multiple same-allele driver mutations). Allosteric effects cannot be applied *in trans*, that is, to mutations in two different molecules, where one molecule has one mutation and the other has the second.

G1
Double mutation on the same gene

Figure 3. Double mutations in the same gene.

We exploited the mutation profiles from TCGA and GENIE pan-cancer cohorts to discover latent drivers (Figure 4). We included all non-synonymous mutations, including variant classifications such as missense, nonsense, nonstop, and frameshift mutations. We excluded frameshifts (insertions or deletions) that alter more than one position in a protein. We also excluded variants where the wild type and/or mutant residues are not specified. Finally, we filtered out the mutations that have VAF (Variant Allele Frequency) less than or equal to 0.125 to assure that the mutations are present approximately in 25% of the sequenced tumor cells.

We identified potential double mutations from proteins having two or more mutations at different positions in the pan-cancer data. Pairwise combinations of mutations in the same genes are pooled and evaluated as a potential double mutation. As a result, we obtained 2,230,203 potential double mutations to be tested among 62,567 tumors.

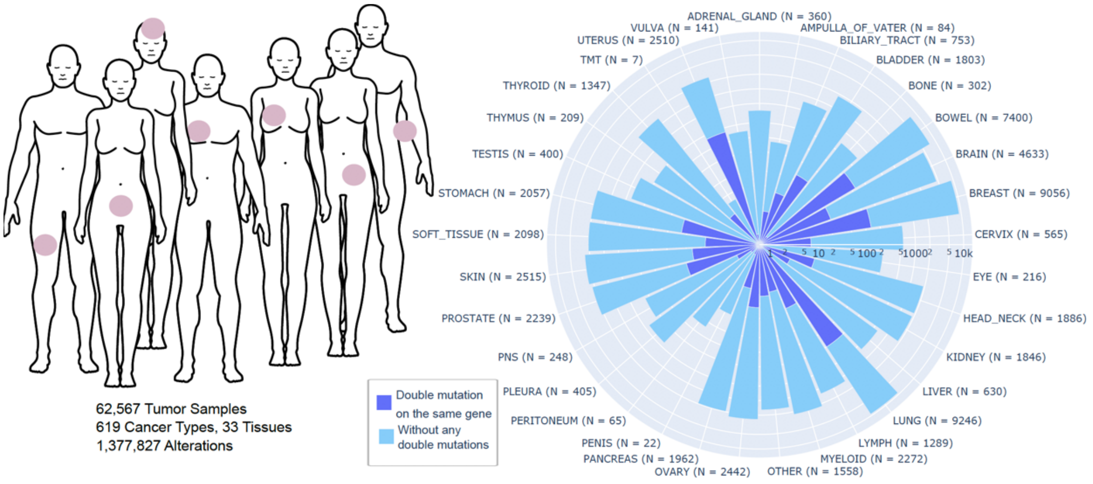


Figure 4. Distribution of double mutant tumors across different tissues. The data is filtered with variant allele frequency (VAF>0.125). Total number of tumors, alterations, cancer types in the union of TCGA and AACR GENIE studies (n = 62,567 tumor samples). Windrose plot shows the number of same gene double mutant (blue) tumors and without any significant double mutation (green) across 33 tissues on the log-scale axis.

To assess the significance of all potential double mutations (2,230,203 doublets), we constructed a 2x2 contingency table for each pair of mutations on each gene (12,724 genes). We built the tables according to the number of samples where constituents of a double mutation are present together, only one of the constituents is present, and none of them are present. (See Methods). Applying Fisher's Exact Test followed by multiple testing correction (Benjamini-Hochberg, q<0.1) resulted in 11,532 significant pairs. Then, we filtered out the doublets if both of the mutation constituents are nonsense (411 double mutations were filtered out of which 49 and 4 were on *APC* and *PTEN*, respectively; the rest scattered through 190 different proteins). A component in downstream of a nonsense mutation in a doublet is either a false positive (chance passenger with no functional consequence), or in *trans* (not a true double mutation affecting the same protein). Thus, we also filtered these significant double mutations out (1377 doublets where 80 and 15 are in *APC* and *PTEN*, respectively; the rest are on 552 different proteins). Then, we applied a stringent filtering to the rest to ensure that co-existing mutations are not erroneously identified. Given a mutation pair *(©, j)*, if mutation © is present in x_i% cells and mutation *j* is in x_j% cells and the total of x_i and x_j is greater than 100%, it is highly likely that double mutation components truly overlap in the cells. After filtering significant doublets based on the proportion of nonsense mutations in double-mutant tumors and total VAF value of each double mutation in the corresponding tumor, 7252 significant doublets were identified, 155 of which were present in three or more tumors. We labeled the constituents true co-existing mutations and retained 155 double mutations for further analysis (Methods, Appendix A-Figure 1, [159]).

We labeled double mutation components as known driver (D) if it is a validated oncogenic mutation in Cancer Genome Interpreter [9]; and otherwise potential latent driver (d). We classified a known driver mutation as a strong driver if it is present in more than 10% of the gene-mutant tumors; otherwise, it is a weak driver. Similarly, we dubbed a potential latent driver mutation as a strong latent driver if it is present in more than 1% of the gene-mutant tumors; otherwise, we classified it as a weak latent driver. Here, we propose that combinations of two strong latent driver mutations can act like a known driver, whereas weak latent drivers can only potentiate the effects of weak driver mutations. We classified mutation pairs as co-occurring based on the odds ratio (OR, log2(OR)>0), and the rest as mutually exclusive. As a result, we identified 148 co-occurring and 7 mutually exclusive double mutations. The mutually exclusive doublets are composed of known driver mutations, i.e., the constituents are either weak- or strong- known driver mutations.

We examined the cis/trans occurrences of the double mutation components. We used publicly available supplementary data from Saito et al., Gorelick et al., and Vasan et al. since raw data or allelic configuration information for the GENIE data, which constitutes around 90% of our dataset, is unavailable [29–31]. With the availability of raw data or allelic configuration information, it would be possible to enlarge the set of double mutations that are in *cis.* The analysis identified 36 tumor samples carrying double mutations matching our data. As a result, we could find *cis* and *trans* information in our double mutation dataset for only 19 doublets accumulated in six genes. For each of the 19 doublets, if the *cis* occurrence is higher among the double mutant group, we labeled it as *cis*, and *trans* otherwise. 8 (5 cis, 3 trans) of these doublets are in the TSGs *PTEN* and *TP53*, the remaining 11 doublets are in OGs *EGFR*, *ERRB2*, *KRAS*, and *PIK3CA* where 10 are in *cis*; and one of them is inconclusive due to the equal number of *cis* and *trans* occurrence.

Recently, the frequency of driver genes was analyzed together with the maximum prevalence of their mutations, distinguishing cancer-specific drivers versus cancer-wide drivers[13]. We applied a similar analysis to our dataset composed of double mutations in the same gene where we obtained the ratio of the number of tissues carrying double mutations (T_{double}) and single mutations (T_{single}). We also calculated the prevalence of double mutations compared to single mutations. For example, *KRAS* double mutations are observed on tumors in four tissues (bowel, pancreas, skin, lung), but single mutations of *KRAS* can be seen on tumors from 30 different tissues. Thus, the tissue specificity, T_{double}/T_{single}, of *KRAS* is ~0.13. Prevalence of *KRAS* is the ratio of the number of double mutant tumors (n = 8) to the number of all *KRAS*-mutant tumors (n~8000), which is ~0.001. Values closer to the origin on the x-axis indicate tissue specificity since for each gene the number of double mutations carrying tissues is smaller compared to the number of single mutations carrying tissues. Larger numbers on the y-axis represent the multitude of patients with double mutations on the gene. Hence, double mutations of *KRAS* can be considered as tissue specific with a low prevalence. As a result, although some genes and their single mutant states have been previously cataloged cancer-wide, we found sets of double mutations that are cancer tissue-specific. Examples include double mutations in *BRCA1*, *EGFR*, and *KRAS* (Figure 5).

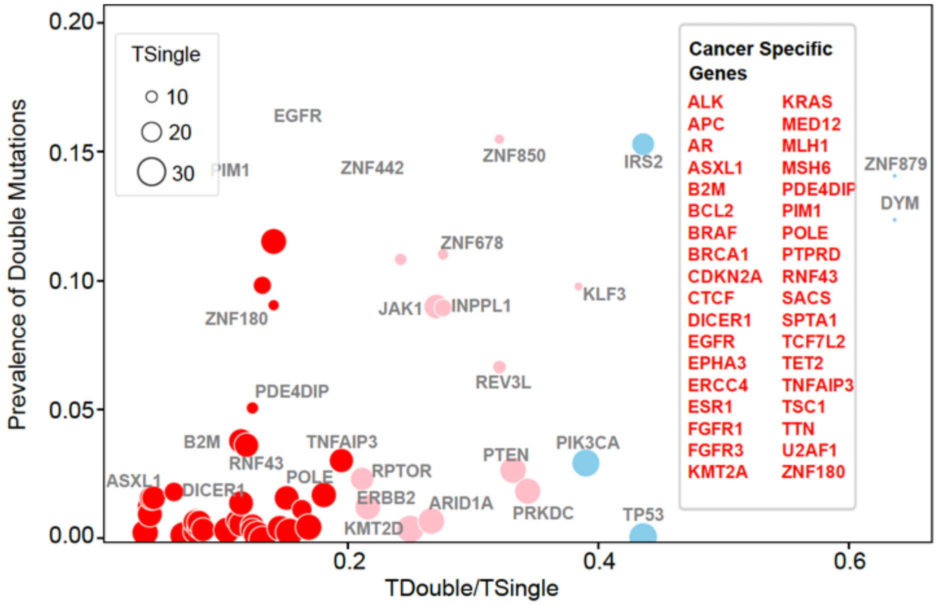


Figure 5. Cancer specificity of the genes harboring double mutations. Tissue specificity of same gene double mutations compared to their single mutant counterparts. Genes having cancer-specific double mutations are red and cancer-wide

double mutations are in blue (25 TSGs with 72 double mutations, 13 OGs with 55 double mutations, and the remaining 15 are labeled as both or neither).

Double mutation components that are not known drivers can be considered as ‘potential latent driver’ mutations. In a doublet, the components can be known drivers or potential latent drivers, so each doublet is cataloged as DD, Dd, and dd. That is, DD is a known driver-known driver doublet, Dd is a known driver-potential latent driver and dd is a doublet consisting of two potential latent drivers. Among the 155 same gene double mutations, there are 54 DD, 29 Dd, 72 dd where the mutations that are not cataloged as known driver (D) are potential latent drivers. The 155 same gene double mutations are composed of a pairwise combination of 213 mutations of which 73 are known drivers and 140 are potential latent drivers. Thus, our analysis can capture rare mutations that are potential latent driver candidates.

The 155 significant double mutations are composed of 213 mutations in 53 different genes. These 53 genes harbor 34,011 mutations that are observed in at least one tumor. Therefore, the fraction of double mutation components among all mutations (in 53 genes carrying at least one same gene double mutation) is ~0.6%. When we evaluate all mutations on 53 genes that are observed in at least three patients, the total number of such mutations is 6245 and the fraction is ~3.4% (Appendix A-Figure 2).

When we solely examine the double mutations in genes classified as OG or TSG, the number of doublets of type DD, Dd, dd is 37, 12, 6, and 17, 17, 38 on the 13 OGs and 25 TSGs, respectively. We observe that OGs have significantly more DD mutations than TSGs (p-value $< 10^{-7}$, two-sided Fisher’s Exact Test) and the fraction of double mutation components among all mutations on these 13 OGs (~1.2%) is almost two times higher than the fraction of double mutation components among all mutations on these 25 TSGs (~0.5%) (Figure 6). This result implies that becoming more oncogenic requires mostly co-occurrence of two frequent mutations while suspending tumor suppressor activities may involve rare mutations coming together.

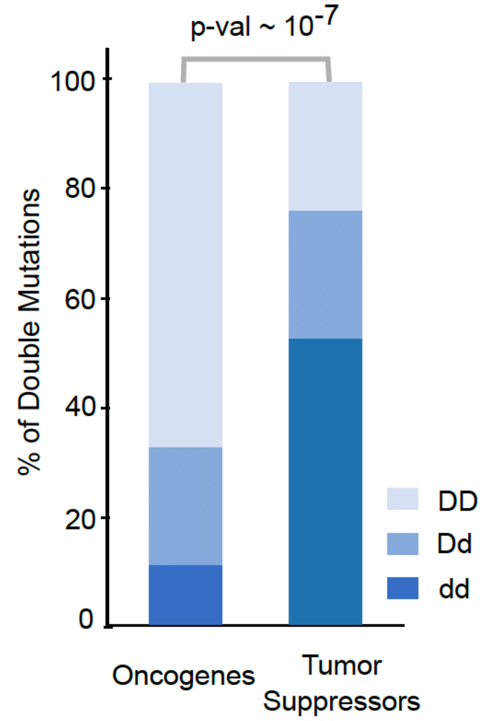


Figure 6. Composition of the double mutations based on known driver (D) and potential latent driver (d) labels in tumor suppressor genes and oncogenes where D is already known frequent driver mutations, d is relatively rare potential latent drivers (p-val $<10^{-7}$, two-sided Fisher's Exact Test).

In the pan-cancer dataset, same gene double mutations accumulate in 53 genes, of which 25 are TSGs, 13 are OGs, 2 are both OG and TSG, and the rest (13 genes) are in other functional categories. There are 821 double mutant tumors carrying at least one double mutation in these 53 genes. In total, the number of tumors having at least one double mutation in an OG and TSG is 468 and 307, respectively. Patient tumors that have at least one double mutation in any TSG have a significantly higher passenger mutation load compared to patient tumors having at least one double mutation in an OG (p-value $< 10^{-30}$, two-sided Mann Whitney-U Test, Figure 6). Given that only ~2% of the 41,734 tumors (having at least one mutation in the 53 genes) carry a double mutation, double mutations are comprising a very small portion of gene-mutant tumors. Especially, TSGs require a very high mutation load for two coexisting mutations in a single gene. Based on the mutation load, and in line with our previous result, loss of function through double mutations in TSGs requires considerably higher mutational load compared to gain of function in OGs. We further compared the mutation load of TCGA and GENIE cohorts separately, taking into account the differences in coverage between the sequencing technologies (Appendix A-Figure 3). There are 63 and 69 tumors with at least one double mutation in an OG and TSG, respectively, in the TCGA data set. Similarly, the GENIE data set has 405 and 238 tumors with at least one double mutation in an OG and a TSG, respectively. Our finding that tumors with at least one double mutation in any TSG have a significantly higher passenger mutation burden is preserved in both the TCGA and GENIE datasets (two-sided Mann Whitney-U Test, p-values 0.003 and 10^{-30}, respectively). In addition, comparing passenger mutation loads among all tumors from TCGA (9588 tumors) and GENIE (52,979 tumors) revealed that TCGA tumors have a larger passenger mutation load (Appendix A-Figure 3A-B). Among the sample group harboring at least one double mutation in a TSG, both passenger and passenger+driver mutation loads are higher than in OGs (Figure 7, Appendix A-Figure 4). Moreover, there are 43 known-driver and 22 latent driver mutations in OGs; and 30 known-driver and 74 latent driver mutations in TSGs when we compare the counts of the known-driver and potential latent driver mutations contributing to the formation of significant double mutations. Given these two facts, excess passenger load might have led to the discovery of more latent drivers in TSGs than OGs.

Double mutations in TSGs are more enriched in latent driver mutations compared to OGs. This abundance could be due to the higher passenger mutation load among tumors with double mutations in TSGs. Despite the small number of samples with *cis/trans* information, the double mutations in TSGs mainly occur in *trans*.
Here, the task is to decide whether latent driver mutations in TSGs are functional or they are false positives due to the passenger mutation burden. There is a need for pre-clinical models such as patient-derived xenografts or cell lines containing double mutations in *cis* in TSGs. To inspect the role of such latent drivers, it would be enlightening to perform a comparative analysis of tumor growth or drug response in wild type, single and double mutant (in *cis*) pre-clinical models.

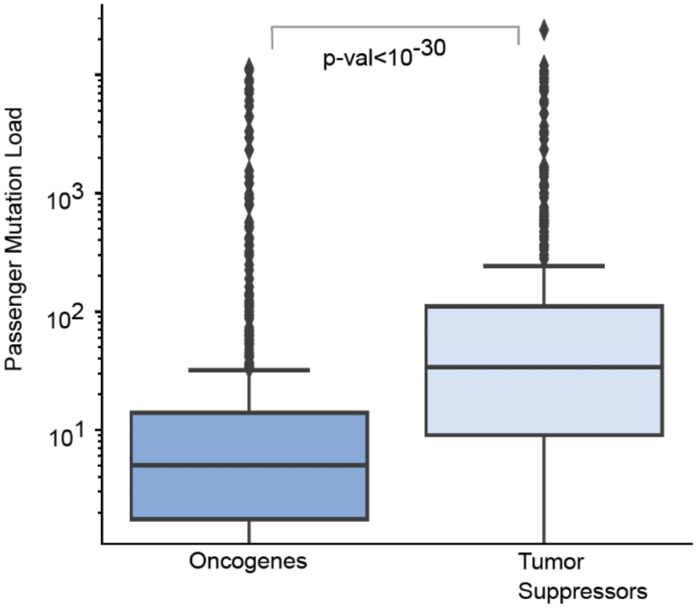


Figure 7. Box plot showing passenger mutation load in OGs and TSGs (p-val $<10^{-30}$, two-sided Mann Whitney-U Test).

3.2.2 Functional interpretation of double mutations by using the characteristics of mutation constituents and double mutant tumors

To interpret the functional consequences of double mutations, we elaborated on the frequencies of the mutations forming the significant pairs, the chemical properties of the wild-type and mutant residues, or the relationships of the double mutation components with mutational signatures. Known driver mutations have a higher frequency than potential latent driver mutations (Figure 8). The median values of tumor count for known driver and potential latent driver mutations are 70 and 9, respectively (p-val$<10^{-10}$, two-sided Mann Whitney-U Test). Potential driver mutations are relatively rare, and their pathological impact can be dramatic when they couple with another mutation. Therefore, we cataloged all potential latent driver mutations that contribute to a significant doublet in the same gene as strong or weak latent drivers. The list of 140 latent drivers can be found in the supplementary data of [159].

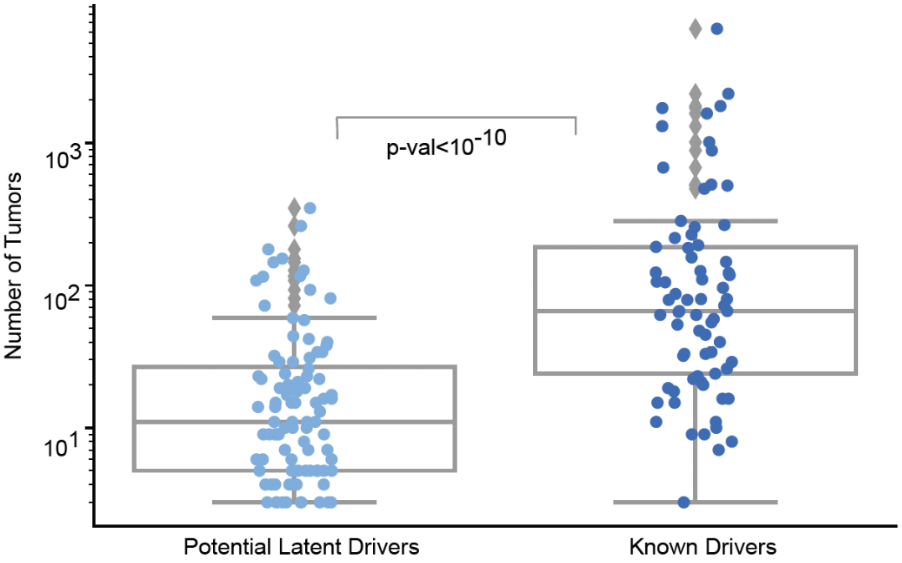


Figure 8. Tumor count distributions of known driver and potential latent driver mutations (p-val$<10^{-10}$, two-sided Mann Whitney-U Test).

Next, we followed a bottom-up approach to obtain the spatial, chemical, and pathway level organization of the double mutations. We used the pan-cancer mutation clusters deposited in 3Dhotspot where each cluster represents the set of mutations that are spatially close to each other [165]. We found that components of the doublets in the same gene are usually spatially distant from each other. The simultaneous presence of two strong spatially close driver mutations is rare in a patient tumor; there are only 15 doublets belong to the same cluster accumulated in *EGFR*, *KRAS*, *PIK3CA* and *TP53*. However, some weak drivers are proximal to either a strong driver or another weak driver, as in the cases of mutations at positions R130/R173 in *PTEN*. Spatially close residues may form potent allosteric couples, which may enhance proliferation.

There are four rare (significant double mutations observed in less than three tumors) BRAF doublets (see supplementary data of [159]). Here, the mechanisms of BRAF mutations were classified into those suggested to be activated as monomers (Class 1), acting as constitutive active dimers (Class 2), and those having impaired/dead kinase activity (Class 3) [166]. There are two doublets having a mutation from Class 2 (K601, L597). These rare double mutations are still kept when we apply a more stringent threshold for total VAF value (up to 40%).

During the formation of double mutations, we had assumed all mutations at a specific position in a protein as the same mutation. We traced back to the mutation positions and obtained wild type and mutated amino acid types to obtain the chemical class changes. A comparison of the fraction of chemical classes of the wild type and mutant residues revealed that Charged>Polar and Charged>Charged switches are more dominant among TSGs and OGs, respectively (p-values ~0.009, 0.04 respectively, two-sided Fisher's exact test, p=0.05) (Appendix A-Figure 5A, Appendix B). Similarly, for the double mutation components that are Known Driver [D] or possible latent driver [d], we compared the chemical class alterations of the mutations. Hydrophobic>Hydrophobic changes are more common among tumors carrying potential latent drivers. Charged>Polar and Charged>Charged changes are prominent among tumors carrying known drivers (Appendix A-Figure 5B).

In total, the number of tumors having at least one double mutation in an OG and TSG is 468 and 307, respectively. The distribution of variant classifications among the tumors carrying at least one double mutation in an OG is as follows: missense+missense (~99%), missense+frameshift (<1%), missense+nonsense (<1%) (Appendix A-Figure 6A). Among these tumors, doublets in which both mutation components are missense mutations are dominant. On the other hand, we see a more diffuse result when we analyze the tumors harboring at least one double mutation in a TSG (Appendix A-Figure 6B). These tumors have variant classifications as missense+missense (50%), frameshift+frameshift (~30%), missense+nonsense (~10%), missense+frameshift (~3%), frameshift+nonsense (~0.48%). The specific sample details related to variant classifications of double mutations in OGs and TSGs are provided in the supplementary data of [159].

During post-processing, we identified 3519 tumors as hyper-mutated out of 62,567 samples with at least one point-mutation with a Q3+8 x IQR threshold (see Methods). First, we used Fisher's exact test ($p < 0.05$) to test the robustness of the 155 double mutations against hyper-mutated samples. 19 doublets are present in hyper-mutated genes including *APC*, *KMT2D*, *ZNF442*, and *ZNF678*; therefore, we excluded these doublets from the subsequent analyses. Among the remaining 136 doublets, one is not significant according to Fisher's exact test ($p < 0.05$) evaluated in the non-hyper-mutated tumor group (see [159]).

Then, we performed a permutation test ($p < 0.01$) using the non-synonymous mutation burden of the double-mutant and single-mutant tumors (see Methods). For each double mutation, we tested the null hypothesis that the double mutant tumors (labeled "Double") have a mean mutation burden less than or equal to the mean mutation burden of the remaining gene-mutant tumors (labeled "Single"). We can reject the null hypothesis for 7 doublets since the p-values obtained with the permutation test are less than 0.01. For the remaining 129 significant doublets the evidence is not sufficient to conclude that the double mutant tumor samples have a lower or equal mean observed mutation load on the basis of failure to reject this as a null hypothesis [159].

We next conducted single base substitutions (SBS) signature analysis of double mutations to explore if components of doublets have common or different signatures (a.k.a. contexts). There are 96 single base substitutions (SBS) of the trinucleotide context where the mutated base is in the middle in square brackets expanded with 5' and 3' bases[48] (e.g. T[G>A]A). We only considered missense mutations in SBS analysis. As a result, we analyzed 711 records (tumor-specific information for each doublet) from 115 doublets in 649 tumors. Within this set, the majority of the double mutations are of different contexts (630 records), and all of these records match with one of the 96 channels (see Methods). There are 81 records (composed of 17 doublets in 77 tumors) where the double mutations are of the same context from 96 channels. The contexts T[G>A]A, C[G>A]A, C[A>G]T, and A[G>T]A are dominant and are present in 40, 13, 5, and 5 records, respectively. Doublets from the same context are mostly located in *PIK3CA* (Appendix B, [159]).

Taken together, double mutations are exceedingly rare phenomena and do not positively correlate with the tumors' mutation burden. The components of the doublets that have been classified as latent driver mutations also occur far less frequently than known driver mutations. Despite several genes with double mutations having a high rate of single mutations in different tissues, there are extremely few double-mutant genes that are tissue-specific. Additionally, in contrast to TSGs, the doublets in OGs are mainly comprised of known driver mutations. Double-mutant tumors with at least one TSG doublet have a higher passenger mutation load. These could be attributed to different mechanisms in elevating oncogenic signaling and lowering tumor-suppressive signaling through double mutations among OGs and TSGs, as well as their biological impact.

3.2.3 Doublets in the same gene are rare but are a signature for some cancer types

After identifying the doublets that are significant at the pan-cancer level, we also assembled tissue-specific sets of double mutations since tissues differ in sample size and are enriched in different genes and mutations. Identification of tissue-specific double mutations are particularly essential because they may point to the tissue of origin of the preclinical models to evaluate drug responses and tumor growth patterns. As shown in Figure 9A, co-occurring double mutations in the same gene are relatively rare, with varied frequencies across tissues. In some tissues, doublets are present in the same gene in up to 10% of the patient tumors (e.g., bowel and breast tissues). However, same gene doublets are either extremely rare or not present in other tissues, such as the pancreas, ovary, liver, kidney, biliary tract.

Since double mutations are significantly less common than single mutations (t-test, p-value~0.006), tissue-specific double mutations can have important roles to predict sensitivity/resistance to specific inhibitors. Here, we aimed to determine the fraction of tumors with at least one double mutation on the corresponding gene among all gene mutants in each tissue or cancer type. Figure 9A illustrates the tissue-specific prevalence of double mutations in the same gene. *TP53* and its double mutations are cancer wide. *PIK3CA* double mutations are quite common in breast and uterus tumors. Among lung tumors, *EGFR*, and among bowel tumors, *PIK3CA* double mutations are ahead by far. Bowel, breast, and lung tissues are enriched with double mutations on specific genes whereas brain tissue has significant but rare double mutations in multiple genes such as *FGFR1*, *IRS2*, *POLE*, *TP53*. LUAD (Lung Adenocarcinoma) is enriched with *EGFR* double mutations. COAD (Colon Adenocarcinoma) is enriched with B2M, PTEN and RNF43 double mutations. We note that *PIK3CA* double mutations are relatively more dominant in BRCA, IDC, ILC, COAD, and UEC subtypes (Figure 9B).

The most frequent mutation, G12D on *KRAS*, is rarely coupled with another mutation in *KRAS* [159]. The mutational mosaic of *KRAS* is distinguishable among different cancer types. G12D is predominantly present in pancreatic, lung and colorectal cancers. *KRAS* mutations are context specific, and a mutation may act differently in different cancer types.

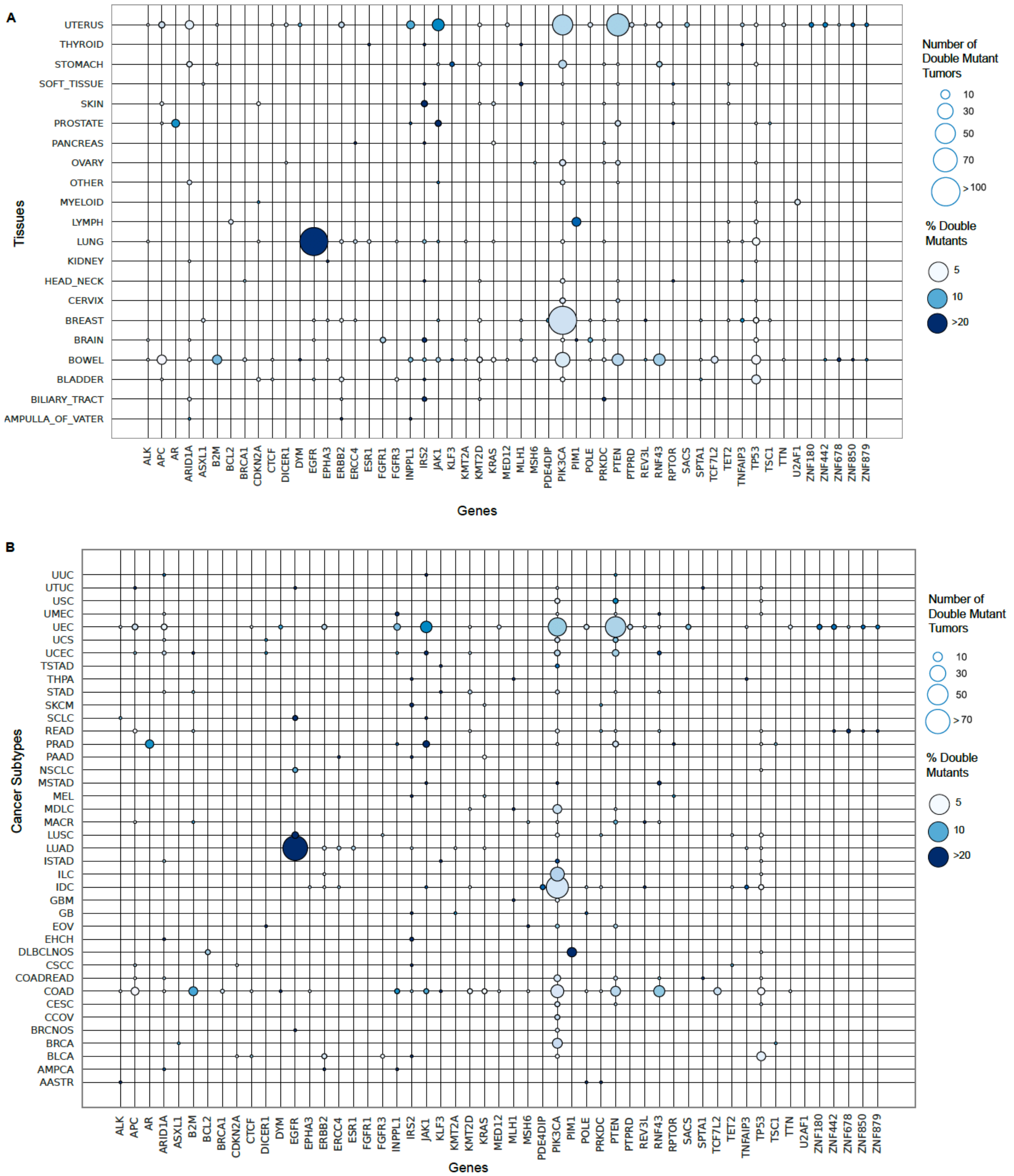


Figure 9. Same gene double mutations are specific to some tissues or cancer subtypes. Bubble plots show number (node size) and frequency (node color) of double-mutant tumors among gene-mutant tumors across different tissues and cancer subtypes (Oncotree). For the 53 genes with significant same gene double mutations, node size represents the number of patients carrying at least one doublet on a gene in a tissue or cancer type. **(A)** Presence of same gene double mutations across different cancer tissues where at least three tumors carry at least one same gene double mutation in one of the 53 genes. **(B)** Presence of same gene double mutations across different cancer subtypes. The cancer subtypes where at least five tumors carry at least one double mutation are listed on the y-axis.

PIK3CA has three driver mutations- H1047, E45, and E542- mostly accompanied by a group of rare mutations that are potential latent driver mutations. Along the same lines, the driver mutations L858, T790, G719 on EGFR; R130 and R173 in PTEN have rare potential latent driver mutation companions (Figure 10).

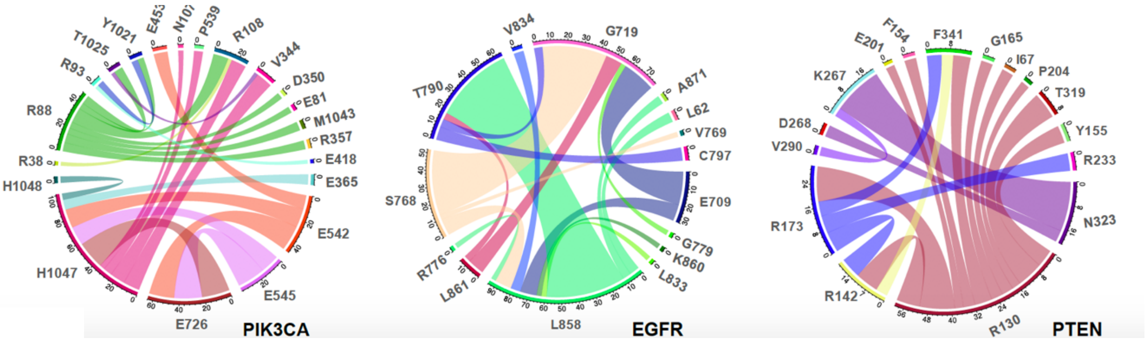


Figure 10. Representation of mutations in genes to compose a doublet as a circular diagram. The strips from one residue to another represents significant double mutations with size of strips indicating frequency of each double mutation.

Thus, even though rare, doublets on the same gene can be a signature for some cancer types, e.g., bowel, breast, and lung cancers.

3.2.4 Linking double mutations to clinical data using cancer cell lines and xenografts

We next explored the potential clinical association of these significant same gene double mutations. Since the patient-specific clinical and treatment data are sparse, we computationally screened differences in cell lines and patient derived xenografts (PDXs) from available experimental datasets. We used cancer cell lines from the DepMap project and PDX samples provided in Gao et al. [163]. In both datasets, mutation profiles and response to a panel of hundreds of drugs are available. Double mutations are rare in the patient tumor samples. We notice the same pattern: Despite scanning hundreds of cancer cell lines, double mutations on the same gene are rare. Among 155 same gene double mutations only 23 double mutations are present in at least one cell line in Cell Model Passports [161]. The intersection between the significant double mutations and their presence in the cell lines led us to pursue a detailed analysis on the genes *PIK3CA*, *EGFR*, *BRAF*, and *PTEN*.

PIK3CA has both strong drivers (e.g., H1047R, E545K), and weak drivers (e.g., R88Q, E453K, M1043I) which are components of 23 significant double mutations in the patient cohort. Despite *PTEN*, *TP53*, *EGFR* and the rest (53 genes in total) have a higher single mutation load compared to *PIK3CA*, their double mutation load is by far less (the paired dot plot for 53 genes is shown in Appendix A -Figure 7). Full activation of oncogenic *PIK3CA* is through at least two drivers acting in different, albeit complementary mechanisms, or enhancing each other. One well example of how the co-occurrence of in *cis* mutations might promote cancer is PI3Kα [29,31]. Moreover, crystal structures and experimental research have shown the activation mechanism at the atomic scale, and the role of frequent or rare driver mutations on this mechanism is widely discussed [36,167–170]. E542K, E545K, and H1047R are hotspot helical and kinase domain mutations that can activate PI3Kα, but they can also have additive effects when combined with the moderate mutations E453K/Q, E726K, and M1043V/I [25,36]. Sporadic and weak activating mutations in PI3Kα are also present. The weak mutations cause conformational changes that lead to PI3Kα activation. These weak mutations include E726K and M1043VI in the kinase domain, N345K, C420R, and E453K/Q in the C2 domain, and R38H/C, R88Q, R93Q, R108H, and G118D in the ABD domain [36] . Thus, the pathological impact of a single driver may be insufficient [40]. One well-known example is H1047 and E545 double mutation enhancing proliferation. However, E545 and E542 double mutations do not make *PIK3CA* reach the fully activated level. A combination of two strong latent driver mutations – but likely not two weak mutations – may act like a driver mutation. The frequency of double mutation components in *PIK3CA* is shown in Figure 11 where many doublets are composed of one frequent and one rare mutation [40,167].

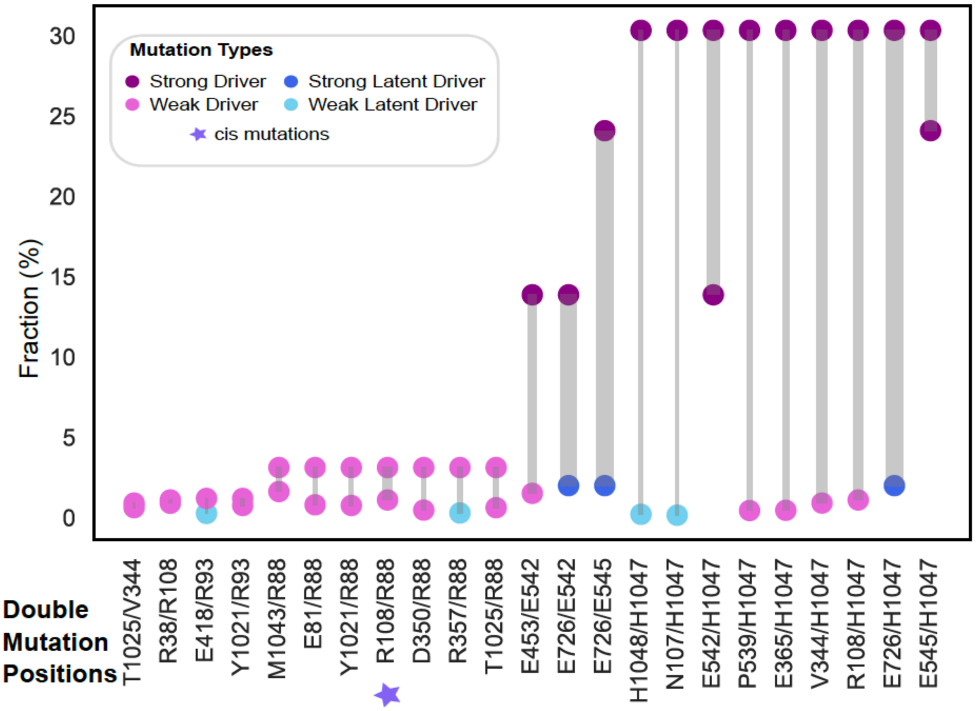


Figure 11. Paired dot plot of the 23 double mutations in *PIK3CA*, and the number of tumors carrying them. Colors indicate type of a mutation, strong driver (purple), weak driver (orchid), strong latent driver (blue), weak latent driver (light blue). Lines size connecting the dots is proportional to the number of tumors with double mutations.

Our frequency-based analysis revealed that E726 is a potential strong latent driver while N107, R357, E418 and H1048 might be weak latent drivers coupled with a weak or strong driver mutation. *PIK3CA* double mutations are also tissue- and context-specific as shown in Figure 12. Most are in breast tissue. An exception involves doublets consisting of R88Q which are depleted in breast but frequent in uterus tumors.

Their structural location is shown in Figure 13. Kinase mutations work by destabilizing the inactive or stabilizing the active state. These are better captured by their detailed conformational alterations. A detailed analysis of the folding free energy (ΔΔG) upon double or single mutation with DynaMut [171] shows the increased impact of several double mutations in the protein stability (Appendix A-Figure 8, Appendix B).

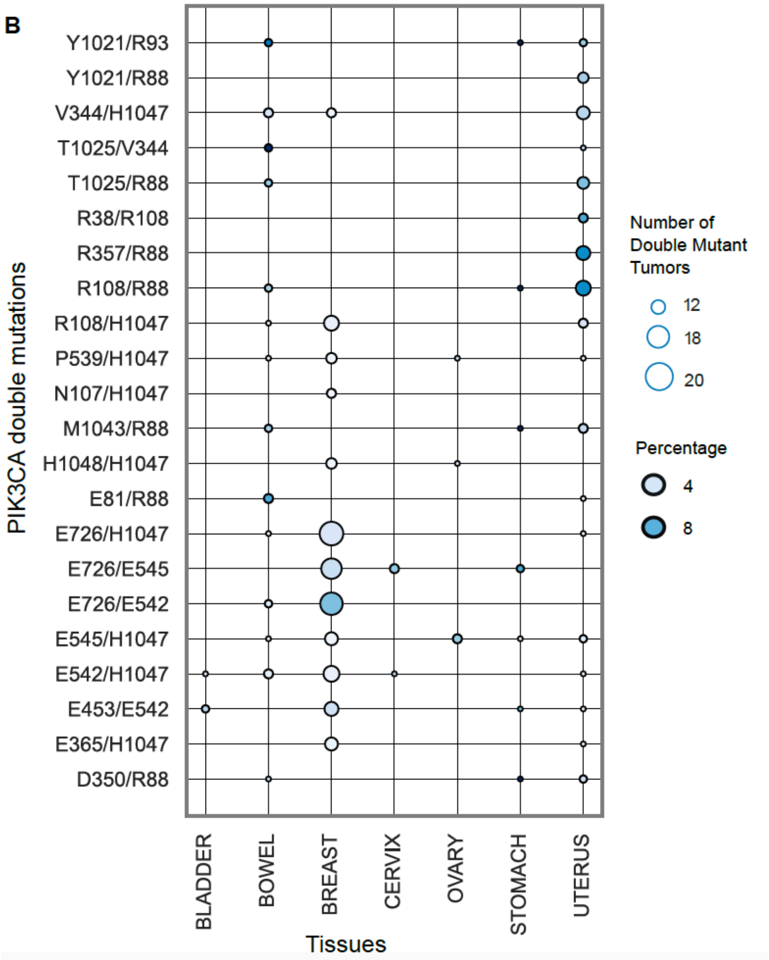


Figure 12. Presence of *PIK3CA* same gene double mutations across different cancer tissues. Dots are scaled based on the number of tumors having double mutations, and color corresponds to the percentage of double mutant tumors among single mutants.

The impact of co-occurring mutations in the same gene is mostly additive but can be also cooperative. When the double mutant phenotype incorporates traits from the single mutants, it can be regarded as additive. Additivity is considered to be a sign that there is no functional link between the driver mutations under evaluation. When the combined effect of two mutations on the phenotypes is greater than the total of each mutation's individual effects, they are referred to as cooperative (also known as synergistic, positive epistasis, or more-than-additive). But rather than just adding up the impacts of two mutations, it is possible to obtain a lesser effect (suppressed, negative epistasis) [32,92,172]. There are seven allosteric mutations at positions 83, 88, 365, 539, 542, 603, 629 in *PIK3CA* in BRCA as cataloged in Allosteric DB [173]. 13 out of 23 *PIK3CA* double mutations are harbored by at least one breast tumor and there are 215 double mutant tumors. The doublet P539/H1047 in *PIK3CA* is composed of one strong driver (at position 1047) and one weak driver mutation (at position 539) which is allosteric. Their effects are additive.

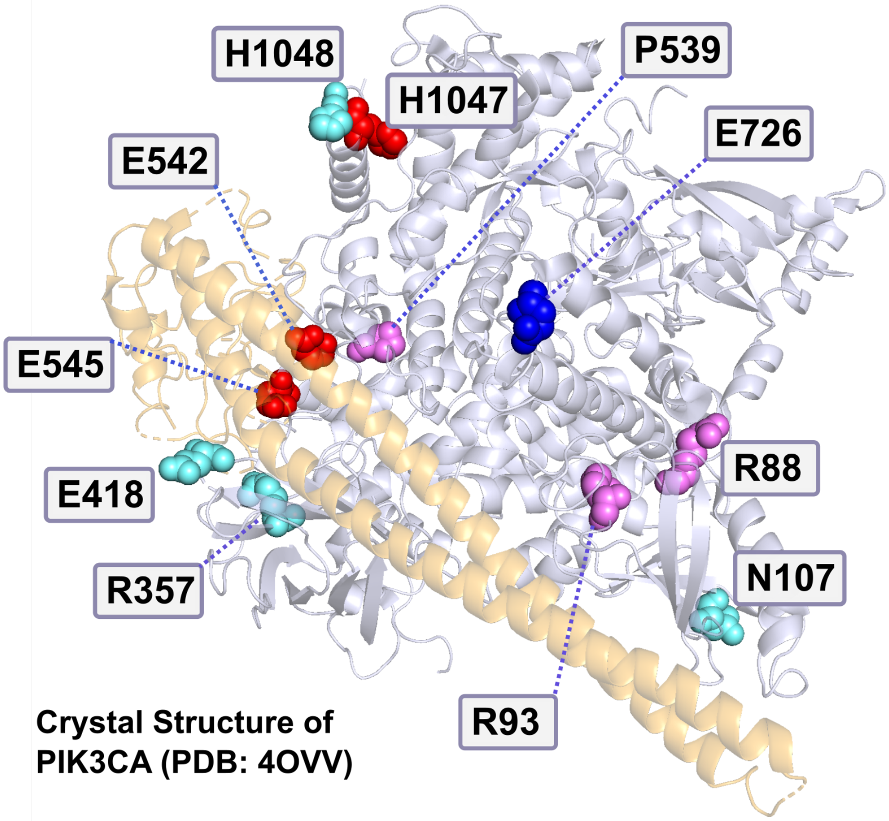


Figure 13. 3D structure of *PIK3CA* (PDB: 4OVV) with H1047, E726, E542, E545, R88, R93, P539, E418, R357, N107 mutations. Red, pink, blue and cyan residues are strong driver, weak driver, strong latent driver, and weak latent driver mutations, respectively.

We found a breast cancer cell lines with *cis* mutations [29] in *PIK3CA* belonging to the BRCA subtype: BT-20 has P539/H1047 double mutation. H1047R is a frequent driver. However, P539 is a rare mutation in the pan-cancer data, making it a potential weak driver. To illustrate the difference between the double mutations and single mutations in terms of drug response, a network of cell lines to drugs and target pathways is constructed (Figure 14A) where drugs are linked to each cell line which has altered response compared to their single mutation counterparts. Indeed, there is a difference in the response to PIK3α inhibitors in double-mutant cell line BT-20 compared to single mutant cell line counterparts (p-value = 0.015). Additionally, double mutant BT-20 cell line is significantly sensitive to the PIK3γ inhibitor CZC24832 while single mutant MFM-223 (H1047) cell line does not show significant response (Figure 14B). Despite factors contributing to the drug sensitivity including other single point mutations and gene copy numbers, double mutations in *PIK3CA* may be still an important contributor as evidenced by increasing its oncogenic activity described in the literature. Therefore, we further explored PDX data to compare double mutant and single mutant *PIK3CA* tumors in terms of the tumor volume changes and drug responses. We found two PDXs having double PIK3CA mutations (E726/H1047, R88/T1025). In PDX X-2524 with doublet H1047R/E726K, a strong known driver/strong latent driver combination, the volume change of the tumor between days 0 and 10 is more than 1700 mm^3, while single mutant tumors X-3077 and X-3078 (with mutation H1047R) have volume change of ~200 mm^3 in the first 10 days reaching ~400 mm^3 at around 35 days (Figure 15A). The double mutant PDX tumor has a dramatically higher growth rate. We compared the growth pattern of double mutant PDX with only single H1047 mutant PDX since there was not any single E726 mutant PDX within the data set.

Tumor growth rate data of these three PDX tumors are also available for drug treatment. BYL-719 (Alpelisib) treatment, a selective PI3Kα inhibitor, diminishes tumor volume by 88% (around 1600 mm^3) in the first 10 days in the double mutant in the xenograft (X-2524) which is dramatically higher than the single mutant xenografts (X-3077 and X-3078) implying increased drug sensitivity (Figure 15B). Drug combination of BYL-719 with LJM716, an anti-HER3 monoclonal antibody, is even more effective in reducing tumor volume than BYL-719 treatment alone because of the HER3 alteration in this PDX (Figure 15C). In cis E726K/H1047R doublet may be a potential strong driver of faster tumor growth rate and better response to PI3K inhibitor Alpelisib; however, no causal conclusions can be drawn without functional data for these cell lines and PDXs. Several factors may lead to this difference in tumor growth rate and response to PI3K inhibitor. Despite other alterations, these PDX models have only one known driver mutation (cataloged in Cancer Genome Interpreter) at position 1047 in PIK3CA and common in all three xenografts (X-2524: *PIK3CA* 1047/726 double mutant, X-3077 and X-3078: *PIK3CA* 1047 single mutant). Another factor is the copy number of the genes in PI3K/Akt/mTOR pathway which could affect *PIK3CA* activity, and drug response. The copy number values (the median values for individual exons called by ExomeCNV [163]) of *PIK3R1* and *AKT3* in the double mutant xenograft are two-fold higher than the single mutant samples (double mutant: 2.41, single mutants:1.34, 1.41). *PIK3R1* functions as a negative regulator of *PIK3CA*. Increased level of *PIK3R1* may negatively regulate the excessive activity of double mutant *PIK3CA*.

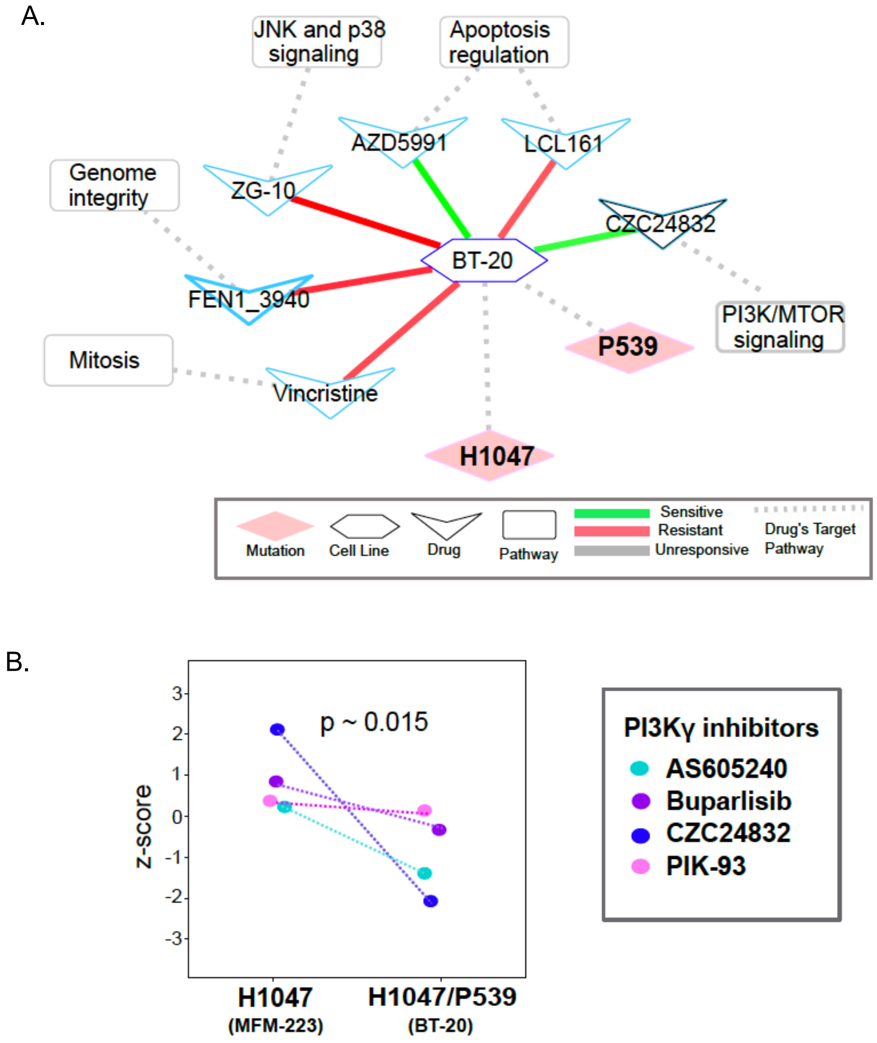


Figure 14. (A) Response of *PIK3CA* breast cancer cell lines with double mutations in *cis* to drugs in network representation. Dashed lines connecting cell line nodes (hexagon) to mutation nodes (diamond) indicates cell lines that harbor the corresponding mutations. Green and red lines connecting cell line nodes (hexagon) to drug nodes (V-shaped) represent sensitivity and resistance of the cell lines to the corresponding drugs, respectively. Dashed lines between the drug nodes (V-shaped) and pathway nodes (rectangle) indicate that the drugs have target(s) in the corresponding pathways. **(B)** *PIK3CA* mutation doublets in breast cancer and the associated violin plot illustrating response to PI3Kγ inhibitors.

On the other hand, not all tumors having double mutation in *PIK3CA* show a similar pattern. For example, growth rate of the tumor (X-3093) with R88/T1025 is slower than of the tumor having a single mutation (at position R88), because both mutations are potential weak drivers and mutations in *PTEN* (E7 and R130*) in addition to other alterations. A tumor with only the R88 mutation is more responsive to PI3K inhibitors compared to that with R88/T1025 (Appendix A-Figure 9A-H). *PTEN* is a tumor suppressor and *PIK3CA* is an oncogene. Active PI3K phosphorylates signaling lipid PIP_2 to PIP_3.This activates a cascade of protein kinases leading to the cell cycle. *PTEN* suppresses cancer by dephosphorylating PIP_3 back to PIP_2. Loss of function at *PTEN* and gain of function at *PIK3CA* ascends PIP_3 levels in the cells [174]. *PTEN* is a negative regulator of the PI3K/Akt/mTOR signaling pathway. Overactivation of *PIK3CA* and loss of activity of *PTEN* due to the double mutations can lead to hyperactivation of PI3K/Akt/mTOR signaling which may result in oncogene induced senescence (OIS), potentially explaining the blockage of tumor growth in the double mutant X-3093 xenograft.

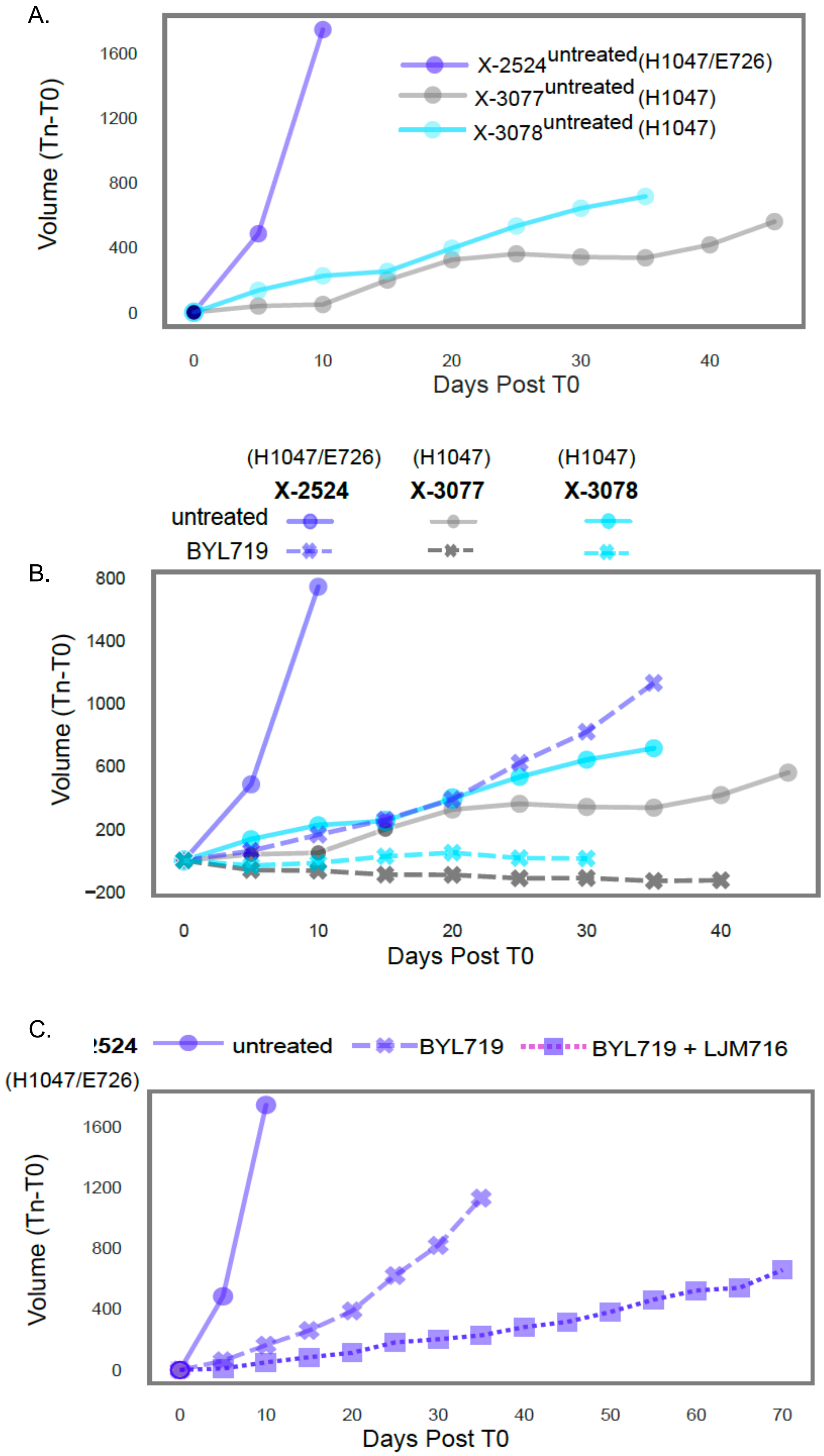


Figure 15. (A) Tumor volume changes of single and double *PIK3CA* mutant xenografts without any treatment. **(B)** Tumor volume comparison of the single and

double mutant xenografts without any treatment and with BYL719 (Alpelisib) treatment. **(C)** Comparing tumor volume changes of the double *PIK3CA* mutant xenografts without any treatment and with BYL719 and BYL719+LJM716 treatment.

Another oncogene with latent driver mutations is *EGFR*; the mutations L62, G779, K860, A871 are weak latent accompanied by weak/strong driver mutations (Figure 16A). There are 17 double mutations on *EGFR*; these doublets are mostly composed of weak drivers (7 doublets) and weak+strong driver combinations (6 doublets). A combination of a weak driver and a strong driver mutation T790/ L858 double mutation in *EGFR* is present in one cell line (NCI-H1975) of lung cancer. H3255 cell line has only one mutation at position L858 in *EGFR* (Figure 16B). Both mutations are in the kinase domain to which the RTK inhibitors bind (PDB: 4I23, Figure 16C). However, response to the inhibitors is significantly different in the cell line with double mutant *EGFR*. It is more resistant compared to the single mutant cell line (p-value=0.01, two-sided Mann Whitney-U Test, Figure 16D).

L858R in *EGFR* is sensitive to *EGFR*-targeted tyrosine-kinase inhibitors (TKIs). After treatment with TKIs, T790M, has been observed. It decreases TKIs' binding [32,175]. L858R lies in the A-loop of the drug binding pocket and destabilizes the inactive conformation. The "gatekeeper" residue T790M is in the hinge region of the binding pocket. L858/T790 increases the protein stability and changes the conformation of the binding pocket which generates resistance to the *EGFR* inhibitors[176,177]. Another double mutation is T790/C797. The sensitivity of the T790M mutant in lung cancer tumors to the third generation TKIs vanishes with the emergence of C797S [32].

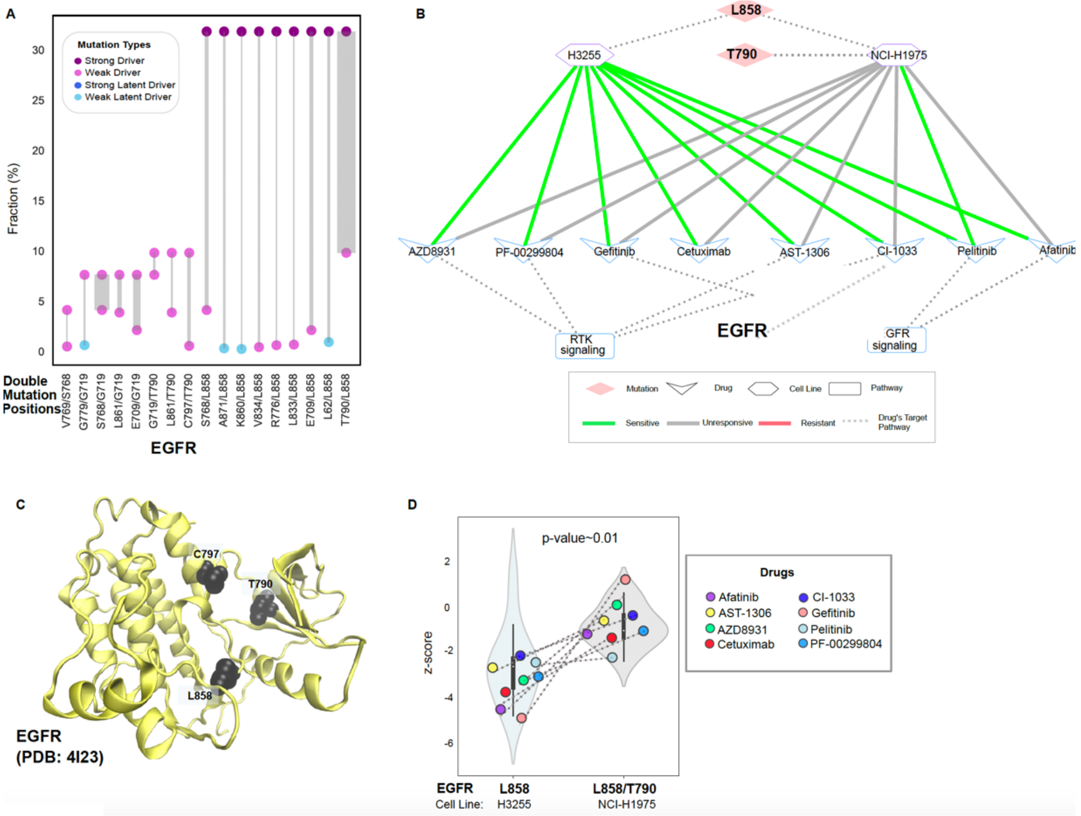


Figure 16. Structural and clinical aspects of EGFR double mutations. (A) Paired dot plot of *EGFR* double mutations. Each paired dot represents one double mutation. Dots are colored according to their type, driver (purple), weak driver (orchid), strong latent driver (blue), weak latent driver (sky blue). Line size connecting the dots is proportional to the number of tumors with double mutations. **(B)** *EGFR* mutation doublets in lung cancer cell lines and their response to drugs in network representation. Dashed lines connecting cell line nodes (hexagon) to mutation nodes (diamond) indicate cell lines that harbor corresponding mutations. Green and red lines connecting cell line nodes (hexagon) to drug nodes (V-shaped) represent sensitivity and resistance of the cell lines to the corresponding drugs, respectively. Dashed lines between the drug nodes (V-shaped) and pathway nodes (rectangle) indicate that, the drugs have target(s) in the corresponding pathways. **(C)** Representation of double mutations in *EGFR* structure. **(D)** EGFR mutation doublets in lung cancer together with the violin plot that shows the response to RTK inhibitor in double mutant and single mutant cell lines. More negative z-score means more sensitivity and more positive z-score means more resistance to the drug molecule**.**

Collectively, pre-clinical models -PDXs and cell lines- bearing double mutations show different growth and drug response patterns. The *PIK3CA* double-mutant PDX grows faster, and its growth trend differs from the single-mutant PDXs. Better response to the PI3K inhibitors both in double mutant cell lines and PDX gives clues to their clinical behavior, despite the necessity of functional data. On the contrary, *EGFR* double mutation may lead to increased resistance by altering the inhibitor binding pocket. Overall, the double mutations and single mutation counterparts are not the only genetic difference between pairs of single mutant and double mutant cell lines or PDXs. However, the significant difference between double and single mutants in terms of drug response and tumor growth make them good candidates for further exploration of their clinical association.

3.3 Concluding remarks

A major observation from our comprehensive analyses in this chapter is that double mutations are infrequent event. We attributed the relative rarity of strong doublet hot spots to oncogene induced senescence (OIS). Another highly plausible explanation is that our doublets count identical mutations. However, the doublets can consist of mutations of similar chemical character. Our results, supported by drug response data of cell lines and patient-derived xenografts, and transcriptomic profiles of single and double mutant tumors, provide a strong background for therapeutic potentials of double mutations. Our results may form a basis for further experimental evaluation of molecular alterations to be exploited for therapeutics across different cancer types. The cancer-specific doublet that we observed here can be useful in clinical identification. Mechanistically, the actions of *same gene* double mutations are more straightforward to interpret as compared to double mutations in different proteins in independent pathways. How double mutations in independent pathways work is highly challenging to understand.

CHAPTER 4

CO-OCCURRING MUTATIONS IN TRANS CAN BE METASTATIC MARKERS; EXCLUDED COMBINATIONS CAN ENCODE ONCOGENE-INDUCED SENESCENCE

Tumorigenesis is attributed to the activation of oncogenic signaling pathways via driver mutations. Tumor formation needs to be evaluated regarding to complex relations between driver mutations. Conventionally, the interactions between driver genes are mainly observed in two different ways: Co-occurring and mutually exclusive driver mutations in certain cancer types. Co-occurring driver mutations usually act through different/parallel pathways, while mutually exclusive driver mutations act through same/redundant pathways. This distinction basically relies on the fact that, the genes on the same/redundant path exert similar and different/parallel pathways exert same and different functions, respectively. The additivity of co-occurring driver mutations in different genes (in trans) can lead to powerful oncogenic signal, encoding aggressive proliferation. Rare co-occurring in trans combinations can serve as metastasis markers; excluded combinations may give rise to oncogene-induced senescence (OIS). In this chapter, a statistical approach was applied to identify significantly co-occurring mutations in the pan-cancer data of mutation profiles of ~80.000 tumor sequences from the TCGA and AACR GENIE databases. Co-occurring and mutually exclusive mutation pairs that additively can promote tumorigenesis through single or multiple pathways are cataloged. Although most of the different gene double mutations were present in primary tumors, rare occurrences can be a signature of metastatic tumors. 4352 statistically significant different gene double mutations that alter non-redundant pathways and interactions and promote cancer-specific tumorigenesis were identified. The relationships among genes, including functional similarities (differences), occurrence in the same or different pathways, and their location in the pathway, upstream (or downstream) were inspected throughout the chapter.

4.1 Methods

4.1.1 Data collection and processing

All available somatic missense mutation profiles are downloaded from two sources, The Cancer Genome Atlas (TCGA) and the AACR launched Project GENIE (Genomics Evidence Neoplasia Information Exchange) [45,56,108]. The TCGA mutation annotation file contains more than 11,000 human tumors across 33 different cancer types. The GENIE mutation file (Release 6.2-public) contains 70,679 samples across 671 cancer subtypes under Oncotree classification. The GENIE cohort contains multiple tumor barcodes belonging to the same tumor type. In such a case only one primary tumor barcode is kept for further analysis. We continued the analysis with 78,837 samples from 671 cancer subtypes and 34 tissues (including UNKNOWN and OTHER categories). There are 50,138 primary and 22,703 metastatic tumors within the final dataset. We used pathways obtained in[178] to associate with dual mutant genes. The dataset contains ten canonical pathways: cell cycle, Hippo, Myc, Notch, Nrf2, PI-3- Kinase/Akt, RTK-RAS, TGF-β signaling, p53 and β-catenin/Wnt. In addition to these, we included categories Cohesin Complex (contains cohesin complex subunits SMC3, STAG2 and RAD21), PKMTs methylate histone lysines and Generic Transcription Pathway (from Reactome).

4.1.2 Identification of significant double alterations

The total number of mutations and copy number variations is 1638191 in 19443 genes. We only evaluated dual combinations of 23715 (on 5215 genes) of these alterations observed on at least 5 tumors. We constructed binary combinations of the remaining 21983 mutations. Then we created a contingency table for each combination of tumor numbers having both mutations, only the first or second alteration and none of those two alterations. Based on the contingency table, we calculated the p-value by using Fisher Exact Test with the formula below:

$$\frac{\left(\frac{a+b}{a}\right)\left(\frac{c+d}{c}\right)}{\left(\frac{a+b+c+d}{a+d}\right)} \quad \mathit{(1)}$$

where a is the number of tumors having both alterations, b is the number of tumors having only the first alteration, c is the number of tumors having only the second alteration and d is the number of tumors not having these two alterations.

~4352 significant pairs on different genes were decided using the Fisher Exact Test for p=0.05 and Benjamini-Hochberg FDR correction was applied (q<0.05). We used the Catalog of Validated Oncogenic Mutations from the Cancer Genome Interpreter [9] to label double mutation components: if a mutation is among the 5601 driver mutations we label it as known driver (D), otherwise driver (d) (based on what is the driver identified/determined).

4.1.3 Survival analysis

For survival analysis, 10336 patients in MSK impact 2017 and 11160 patients in TCGA and their overall survival status are used [108]. We compared survival times of tumor groups with significant same/different gene double mutations and single mutations in a specific cancer subtype. The first group is the union of patients with significant doublets whereas the second is the union of patients that carry only one component of these significant double mutations. Then we gathered overall survival times (time in months) and vital status (1: Deceased, 0: Alive) of these patients for survival analysis. We utilized the "survival" library of R to do Kaplan Meier Survival Analysis of double and single mutant groups. The survival probability at any particular time is calculated by the formula given below [179]:

$$S_t = \frac{(Number\ of\ subjects\ living\ at\ the\ start)-(Number\ of\ subjects\ )}{Number\ of\ subjects\ living\ at\ the\ start} \quad (2)$$

4.1.4 Oncoprint maps

To reveal mutual exclusivity and co-occurrence patterns between double mutations we plotted oncoprint maps by using ComplexHeatmap package of R.

4.1.5 Transcriptome analysis

To identify differentially expressed genes in the group of patients with double mutations compared to the single mutant patient group we downloaded RNA-seq transcriptome data of the TCGA project from the cBioPortal database (https://www.cbioportal.org). We used median Transcripts Per Kilobase Million (TPM) values of RNA-seq data of PAAD and BRCA cohort of TCGA. For the PAAD cohort, 177 patients with TPM values, we constructed two groups where Group 1 is tumors having at least one significant double mutation of type $KRAS^{G12D/V/C}$ $+TP53^{mutation}$ and group 2 is tumors having either single mutant $KRAS^{G12D/V/C}$ or single mutant TP53. For the BRCA cohort, 1082 patients with TPM values, we constructed two groups where the $PIK3CA^{Double\ mutant}$ group has tumors having at least one significant PIK3CA same/different gene double mutation and the $PIK3CA^{Single\ mutant}$ group has tumors having at least one PIK3CA mutation but without any significant same/different gene double mutation of PIK3CA. We calculated the $\log_2FC$ value of each gene between the double mutant and single mutant groups by using the formula:

$$log2FC = \frac{Mean(Gene's\ TPM\ values\ among\ Double\ Group\ Patients)}{Mean(Gene's\ TPM\ values\ among\ Single\ Group\ Patients)} \quad (3)$$

We identified differentially expressed genes between the double mutant tumors group and the single mutant tumors group (comparison of means of TPM values by Mann-Whitney U-Test). If $|log2FC|$>0.5 and p-value < 0.01 we considered the corresponding gene as differentially upregulated or downregulated in the double mutant group. We continued our analysis with these differentially expressed genes. Then we calculated z-scores of each gene as follows:

$$z = \frac{TPM\ value\ of\ Gene - \mu}{\sigma} \quad (4)$$

where is μ is the mean of TPM values and σ is the standard deviation across all samples in the double mutant and the single mutant groups. After obtaining z-scores for each gene, we sorted the genes as downregulated and upregulated and represented these values as a heatmap (https://seaborn.pydata.org). We used Webgestalt (http://www.webgestalt.org) for the gene set enrichment analysis where the functional database is selected from Reactome and significantly up- or down-regulated pathways are found. FDR threshold is selected 0.05 and the list of genes ranked by their logFC values are given as input.

4.1.6 FPGrowth tree construction

We are inspired by the prediction of association rules in database transaction systems which find the most frequent associations between items in each transaction. In our setup, each tumor is considered as a transaction and each alteration is considered as an item in the transaction. We used the pyspark library together with the modules pyspark.ml and pyspark.sql to find the association rules. The FP growth algorithm is selected for tree construction where each node represents one alteration and each edge in the tree represents the association of the nodes [180]. In the constructed tree, all nodes in the path from root to the distant node are associated with each other and strongly present together in the tumors. The tendency of the alterations to be specific to metastatic tumor is calculated by

$$propensity(i) = \frac{xi/Ni}{X/N} \quad (5)$$

where x_i is the number of metastatic tumors having double mutation i, N_i is the number of tumors having double mutation i, X is the number of metastatic tumors and N is the number tumors in our dataset.

4.1.7 Personalized page rank algorithm

We used a network diffusion-based algorithm to find the most affected region of the interactome given a set of nodes. Given a directed or undirected graph G(v,e) where v∈V and e∈E and a set of seed nodes S⊆V, the personalized PageRank algorithm solves the seed set expansion problem, where it finds which additional nodes may exist in the community besides the nodes in S and ranks them according to their importance. We used the PageRank [181] function implemented in Python networkx library [182]. The damping parameter, alpha, is selected as 0.85 and the number of iterations are 100. In NFE2L2-KEAP1 example, the seed nodes are selected as S={NFE2L2, KEAP1, KRAS}.

4.2 Results

4.2.1 *Functionally non-redundant mutations on different genes in tumorigenesis*

Collecting missense mutations on each gene and counting their pairwise combinations result in 4352 significant double mutations on different genes composed of co-existing mutations from 282 genes. 9276 non-significant double mutations are identified with Fisher Exact Test ($p \geq 0.05$). To examine the double mutations at the pathway level, we merged pathway information from KEGG and Reactome for each gene. In Figure 17A, we see that, both 3655 double mutation constituents belong to at least one pathway (~84%). Among these 2999 of the double mutations do not share a common pathway (~69%), and the remaining 656 mutation doublets have at least one common pathway (~15%). The set size of all doublets where at least one component does not belong to any pathway is 697 (~16%). When we looked at the potential of oncogenes (OG) and tumor suppressor genes (TSG) to form significant pairs among the 3958 annotated doublets, we identified 1998 OG+TSG, 1204 TSG+TSG, 532 OG+OG, 128 TSG+OG/TSG, 93 OG+OG/TSG, 3 OG/TSG+OG/TSG doublets (for 394 doublets, at least one component is not annotated) (Figure 17B). Different gene double mutations (~45%) are dominated by OG+TSG type doublets.

We examined whether double mutations formed by various types of combinations of oncogene or tumor suppressor gene mutations were known drivers (D) or drivers (d). Figure 17C shows that for OG+OG, OG+OG/TSG, OG+TSG, TSG+OG/TSG and TSG+TSG type double mutations fraction of known driver-known driver combinations are 58%, 3%, 42%, 2%, 20; fraction of known driver-driver combinations is 21%, 45%, 34%, 17%, 28%; fraction of driver-driver combinations is 20%, 51%, 23%, 79%, 51%. OG + OG accumulates more DD type mutations than TSG+TSG, which is consistent with our findings on the same gene double mutations.

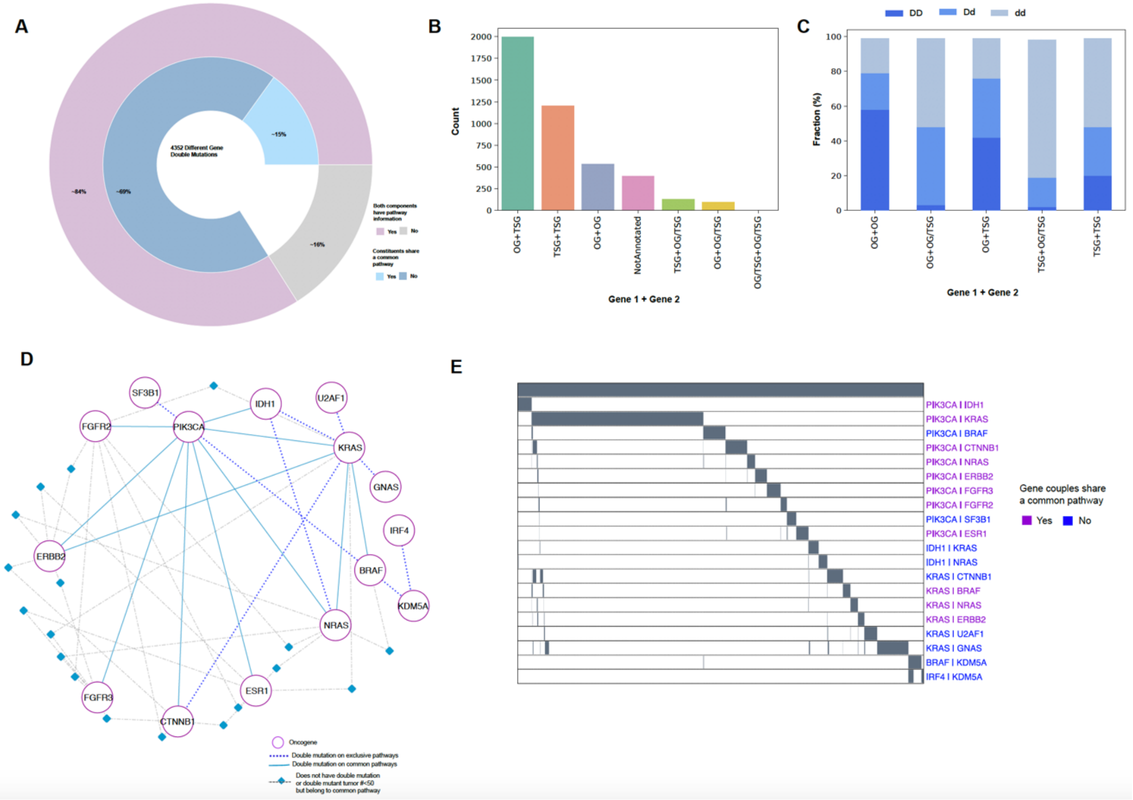


Figure 17. Overview of different gene double mutations. (A) The nested pie chart displays details about the pathways of 4352 different gene double mutations. In the outer layer, light plum region represents the fraction of double mutations where both components belong to at least one pathway in KEGG or Reactome (~84%); for the doublets in the gray region with fraction ~14%, at least one component does not belong to any pathway. In the inner layer, ~69% of the doublets are formed by components belonging to exclusive pathways (royal blue), ~15% of the doublets are composed of constituents sharing a common pathway (sky blue). **(B)** Count plot of different gene double mutations. Among 3958 annotated doublets Gene1+Gene2, there are 532 OG+OG, 93 OG+OG/TSG, 1998 OG+TSG, 3 OG/TSG+OG/TSG, 128 TSG+OG/TSG, 1204 TSG+TSG type combinations of genes harboring a double mutation component (for 394 doublets, at least one component is not annotated). **(C)** Stacked bar plot of 3958 annotated doublets (x-axis, Gene1+Gene2) representing fractions (y-axis) of known driver (D) and driver (d) combinations. Blue, royal blue, light blue regions represents the fraction of known driver-known driver (DD), known driver-driver (Dd), driver-driver (dd) type doublets respectively. **(D)** We call a gene couple A|B if there is a different gene double mutation where the component mutations are on the genes A and B, and the gene couple A|B is mutated on the union of tumors with double mutations on these genes. Network representation of oncogene couples seen together in 50 or more patients: Nodes are oncogenes forming a gene couple, blue solid lines between nodes indicates the genes connected by this edge are constituents of a gene couple and they are members of at least

one common pathway; dashed blue line indicates the genes connected by this edge are constituents of a gene couple but they do not belong to any common pathway. Gray dashed lines with a blue diamond mark, on the other hand, show that on the nodes they connect, there is no double mutation of any different gene, but that these two genes coexist on at least one pathway. **(E)** Oncoprint map of the oncogene couples. Tumor samples with double mutations on the oncogene couples are usually mutually exclusive.

We call a gene couple A|B if there is a different gene double mutation where the component mutations are on genes A and B, and the gene couple A|B is mutated on the union of tumors with double mutations on these genes. Figure 17D depicts oncogene couples seen together in 50 or more patients. Nodes are oncogenes forming a gene couple, blue solid line between nodes indicates the genes connected by this edge are constituents of a gene couple and they are members of at least one common pathway; dashed blue line indicates the genes connected by this edge are constituents of a gene couple but they do not belong to any common pathway. Gray dashed lines with a blue diamond mark, on the other hand, show that on the nodes they connect, there is no double mutation of any different gene, but that these two genes coexist on at least one pathway. PIK3CA mutations co-occur with mutations on KRAS, FGFR2, IDH1, ERBB2, FGFR3, CTNNB, ESR1, NRAS genes. These PIK3CA partner genes share at least one common pathway with PIK3CA. KRAS forms gene couples with the genes ERBB2, NRAS and BRAF sharing a common pathway with KRAS. The density of gray dashed lines suggests that oncogenes are less likely to co-mutate with genes of at least one common pathway, hence, from this perspective the genes on the same pathway are mutually exclusive. The oncoprint map of oncogene couples in Figure 17E reveals that tumor samples harboring double mutations are mostly mutually exclusive.

Post translational modification sites (PTMs), controls protein functions and protein-protein interactions. If one or both of the different gene double mutation components hit a PTM site, this may increase or decrease the oncogenic signal emerging from a different gene double mutation.
To this end, we checked whether the constituents of double mutations match with any PTM site deposited in the PhosphoSitePlus database (https://www.phosphosite.org). Among the 1136 doublets, one constituent mutation of the 174 doublets (~15%) matches with a PTM site; for two doublets on CTNNB1S37/U2AF1S34 and CTNNB1S33/U2AF1S34 both mutations are on phosphorylation sites (for 3216 doublets both components do not match with a PTM site). The low frequency of mutations matching a PTM site might be related to the possibility that physical and chemical changes in one of the mutations reduce the oncogenic signal potential of the double mutation. (Why checking this? It needs explaining) [183,184].

The co-occurrence/mutual exclusivity pattern at the gene level can be examined in more detail with the help of double mutations. The majority of genes with double mutations do not share a common pathway (~69% of all doublets), might be related to the fact that these genes do not have a common function and therefore may increase oncogenic signal via activating different oncogenic pathways.

4.2.2 Can double mutations co-occur in the same interface?

We collected protein-protein interfaces from the Interactome Insider [185] which consists of PDB and Interactome3D interfaces [18]. One constituent of 29 out of 4352 different gene double mutations matches with an interface mutation. Appendix C-Figure 1A shows the mutation prevalence (node size) and tendencies to contribute to double mutations in different genes and tissues.

We ask whether genes contributing to a significant double mutation are in the same or different pathways. The list of genes contributing to dual mutations is piped to the EnrichR tool [186] and Reactome pathway annotations are collected for each gene. Our results show that a large portion of the genes act in different pathways (corresponding to 3180 double mutation pairs in different genes). To show these results in an abstract way, we used ten prominent pathways from [178] (including Cell Cycle, HIPPO, MYC, NOTCH, NRF2, PI3K, TGF-Beta, RTK RAS, TP53, WNT) and Cohesin Complex subunits (Appendix C-Figure 1B) and calculated the prevalence of significant dual mutations on different genes in these pathways. Double mutations from 83 genes out of 282 genes are in these ten pathways and the cohesin complex. For example, co-occurring mutations – one in a gene in the RTK/RAS pathway and one in a gene in PI3K or TP53 pathways – can significantly accelerate tumorigenesis. The results show that co-occurring driver mutations can be in different pathways.

We assessed the prevalence of gene double mutations in different tissues. Figure 18A and 18B show gene pairs having significantly co-occurring mutations. For example, mutations in PIK3CA and FGFR3 co-occur significantly in bladder cancer tissue while BRAF and RNF43 co-occurring mutations are specific to colon cancer. Dual mutations in PIK3CA and ESR1 are significantly associated with BRCA. We analyzed KRAS, APC, and their associated co-occurring mutations separately where KRAS and CDKN2A mutations are in pancreatic cancers (Appendix C-Figure 2A-B).

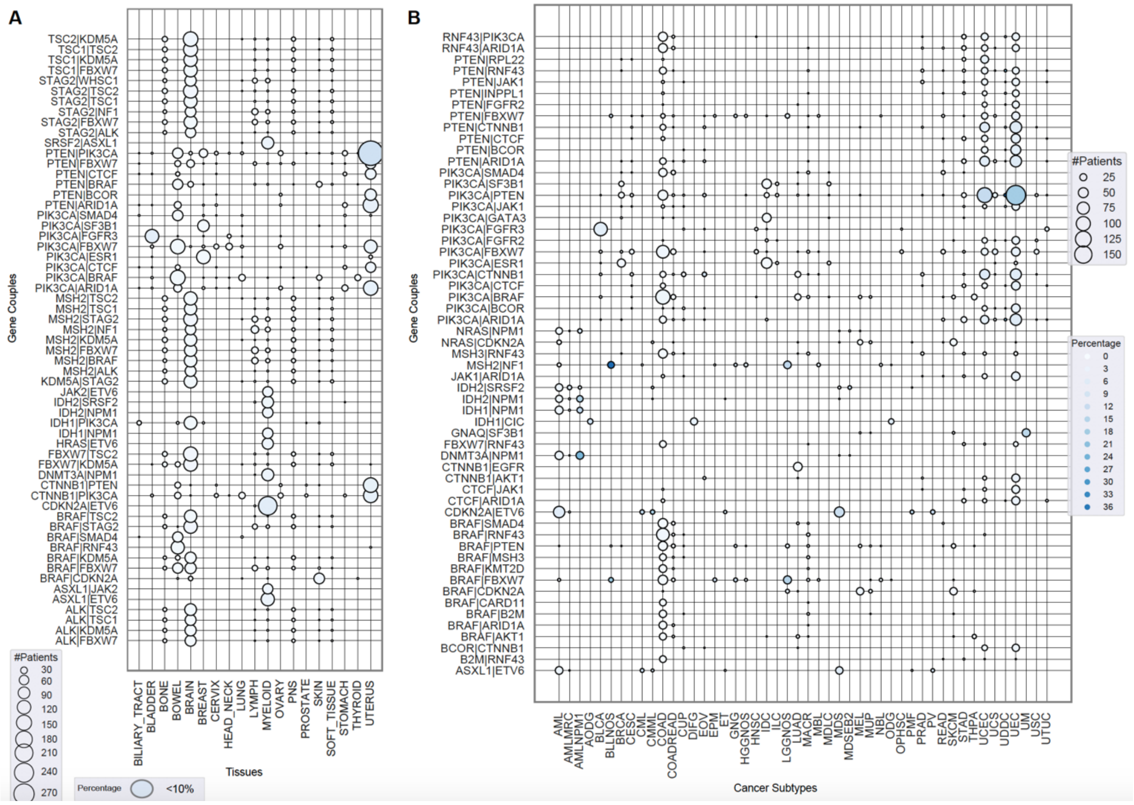


Figure 18. Widespread different gene double mutations in different tissues and cancer subtypes (The double mutations where one of the constituent mutations is on TP53, APC or KRAS are excluded from the list). We call a gene couple A|B if there is a double mutation where the component mutations are on the genes A and B, and the gene couple A|B is mutated on the union of tumors with double mutations on these genes. Various gene couples are prominent in different tissues and cancer subtypes. Node size is the overall number of patients with different gene double mutations in each tissue/subtype, and node color is a fraction of all double mutant tumors to the all tumors in tissues/subtypes. **(A)** Different gene couples (mutated on at least 50 patients in a tissue) and their tissue prevalence (have mutation at least 3 patients on a gene couple). **(B)** Different gene couples (mutated on at least 20 patients in a cancer subtype) and their cancer subtype prevalence (have mutation at least 5 patients on a gene couple)

A driver alteration may affect the signaling pathways and the transcriptional profile of genes. In a transcriptional regulation network, these altered genes can contribute to upstream activities through feedback loops [187]. We searched for transcriptional differences between tumors having double mutations in different gene pairs and their single mutation counterparts.

One of the most frequent mutations in pancreatic cancer patients is KRAS^{G12D}. It pairs with a mutation in TP53, which impairs DNA binding. In the pan-cancer dataset, there are 2232 PAAD tumors, among these 1720 have KRAS^{G12D} mutations. There are 131 significant mutation doublets composed of KRAS^{G12} mutations and at least one mutation in TP53 that exists in 1068 patient tumors. Among them, mutations at positions 248 and 273 are directly in contact with DNA. Position 175 is far from the DNA binding region however it is in contact with a Zinc ion and a mutation at that position destabilizes p53 and eventually prevents its binding to DNA. Some tumors have significant double mutations composed of TP53 and KRAS^{G12D} mutations. Because this coupling may have an impact on transcriptional regulation, we compared the transcriptome profiles of Group 1 (defined as PAAD tumor having at least one significant doublet composed of *KRAS*^{G12D} and TP53 mutations) to Group 2 (defined as the PAAD tumors having mutation either in KRAS or in TP53 or mutation in any of these genes that do not contribute to a significant doublet). This analysis is used to distinguish the transcriptional alterations when a driver mutation in TP53 co-occurs with a KRAS^{G12D} mutation. Group 1 has 67 patient tumors. Group 2 has 35 of which TP53 is wild type in 25 samples and the rest have TP53 mutations. None contributes to co-occurring mutations. Using a conventional transcriptomic analysis, we obtained 412 differentially expressed genes between Group 1 and Group 2 in PAAD samples (Figure 19A, 169 upregulated and 243 downregulated genes where p-value < 0.01 (Mann Whitney U Test) and $|\log_2 FC| > 0.5$). This allowed us to identify TP53 mutations contributing to doublets in transcriptional regulation. A set of genes in immune response, positive regulation of cell proliferation and cell-cell signaling, are enriched in Group 1 compared to Group 2. Next, we found the significant upstream transcription factors that regulate differentially expressed genes using the TFBSCons dataset in the DAVID tool. As a result, BACH1 (268/412 DEGs), PPARG (165/412 DEGs), NFKB1 (76/412 DEGs) are retrieved as the main regulators (Figure 19B). BACH1 is an epithelial-mesenchymal transition enhancer promoting pancreatic metastasis [188]. We constructed the network of the transcription factors and KRAS with CancerGeneNet and found that the proliferation phenotype is upregulated (Figure 19C). This result also agrees with driver mutations in TP53 significantly affecting the transcriptional output compared to the single mutant counterparts [30].

Similarly, PIK3CA double mutations predominantly exist in breast tissue. We compared the transcriptomic profiles of tumors with single mutant PIK3CA to double mutant PIK3CA. The first tumor group has at least one mutation doublet in PIK3CA and the second tumor group has PIK3CA mutations that are not a component of any significant doublet. We obtained 46 tumors in the first group and 296 tumors in the second which has transcriptomic data and satisfies these conditions. The transcriptome analysis of double and single mutant samples revealed that among 117 significantly expressed genes 90 are upregulated and 27 downregulated in double mutant PIK3CA (Figure 19D). In this comparison, olfactory receptors are upregulated and the GPCR signaling pathway is significantly more active in dual mutant tumors (Figure 3E). We also constructed a subnetwork of these downregulated genes using a diffusion-based approach (see Methods). The resulting network recovers receptor enhancers and transporters (Figure 19F).

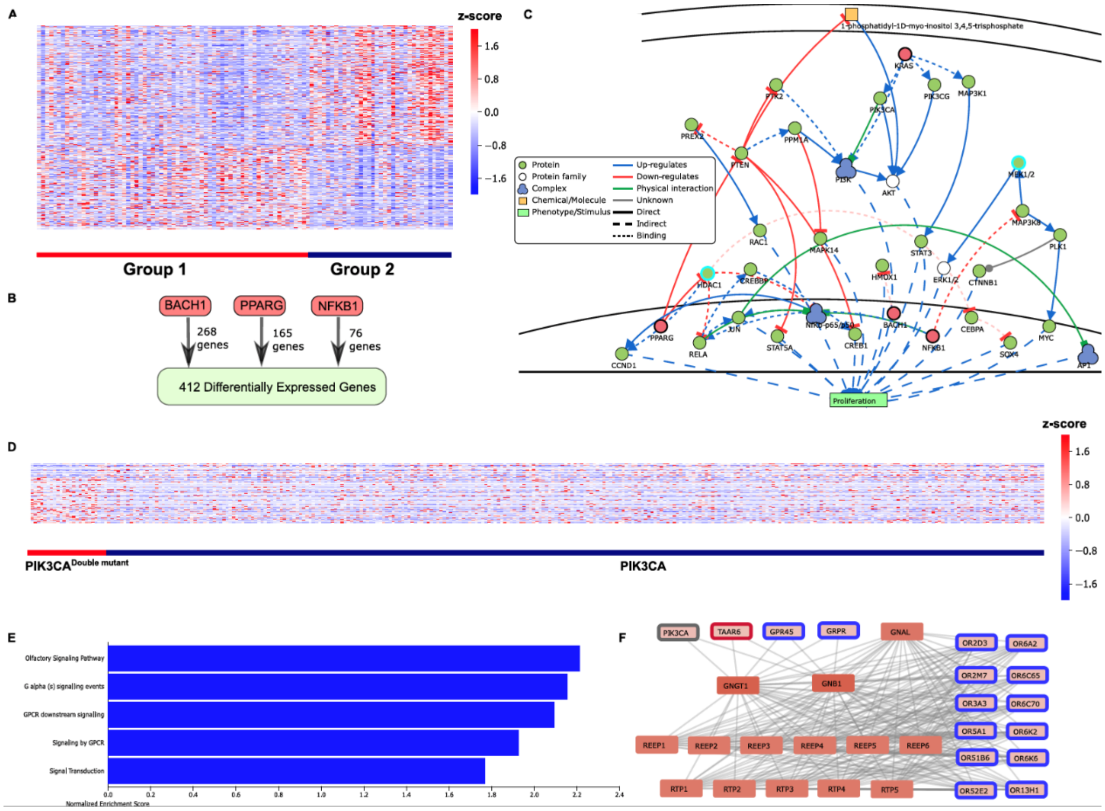


Figure 19. Transcriptome analysis of tumors with double mutation (A) Comparison of gene expression profiles of double mutant and single mutant TP53 cases in KRAS^{G12} mutant PAAD tumors revealed 412 differentially expressed genes (169 upregulated, 243 downregulated) between Group 1 and Group 2. The heatmap shows the z-scores of differentially expressed genes (DEGs) across patient tumors. **(B)** Main regulator transcription factors of differentially expressed genes are BACH1, PPARG, NFKB1 obtained from TFBSCons dataset in the DAVID tool (https://david.ncifcrf.gov). **(C)** The network of KRAS and important transcription factors linked to proliferation phenotype generated by CancerGeneNet [65]. **(D)** The heatmap of differentially expressed genes between double mutant and single mutant PIK3CA cases in breast cancer tumors. Among 117 significantly expressed genes in double mutant tumors, 90 genes are upregulated, and 27 genes are downregulated. (E) Enriched Reactome pathways in differentially expressed

genes obtained from Webgestalt web tool. Among the double mutant tumors, olfactory receptors are highly upregulated and the GPCR signaling pathway is more active compared to the group without any significant doublet on PIK3CA. **(F)** Subnetwork of DEGs in the GPCR signaling pathway where the red node border represents upregulation and blue represents downregulation in dual mutant cases.

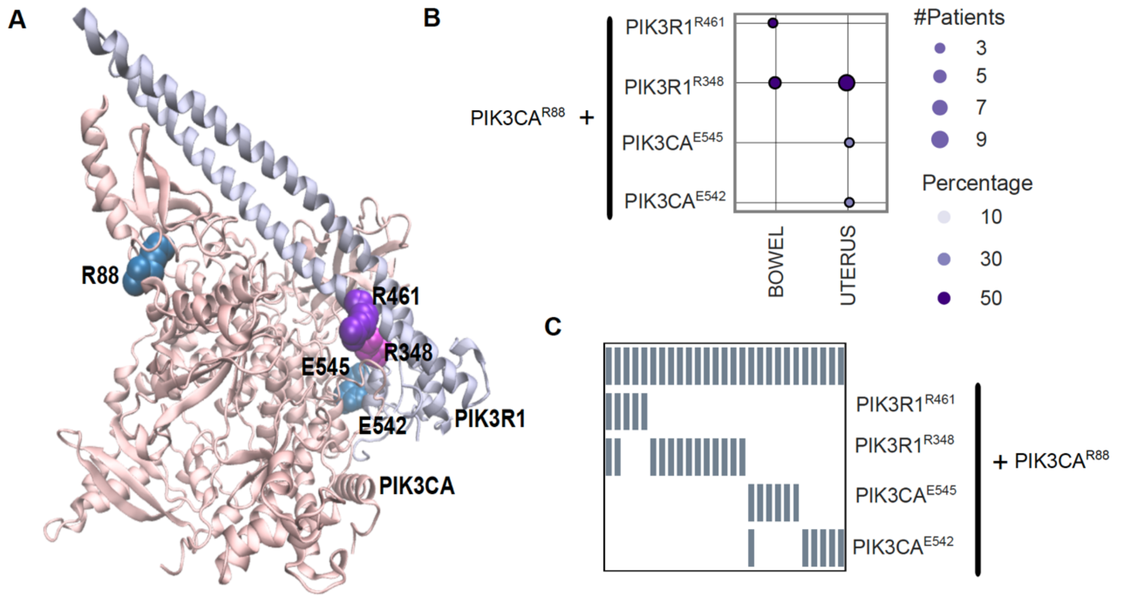


Figure 20. Different gene double mutations on the PIK3CA-PIK3R1 complex. (A) PIK3CA^{E542}, PIK3CA^{E545}, PIK3R1^{R348}, PIK3R1^{R461} mutations on PIK3CA-PIK3R1 interface. **(B)** Tissues (x-axis) harboring different/same gene double mutations (y-axis) in at least 3 tumors. **(C)** Same gene double mutations on PIK3CA and different gene double mutations on PIK3CA-PIK3R1 complex are mutually exclusive.

Another interesting case occurs on the PIK3CA-PIK3R1 complex (PDB: 4OVV), (Figure 20A). We investigated the two different gene double mutations of PIK3CA^{R88} (located at the ABD–kinase domain interface) pairing with two interface mutations PIK3R1^{R348} and PIK3R1^{R461}. In the pan-cancer dataset, 14 tumors harbor PIK3CA^{R88}/PIK3R1^{R348} and 5 tumors harbor PIK3CA^{R88}/PIK3R1^{R461} double mutation. PIK3CA^{R88}, PIK3R1^{R348} (number of patients = 58) and PIK3R1^{R461} (number of patients = 22) are weak driver mutations (labelled as Known Driver, D, number of patients < 500). These double mutations are prevalent among bowel and uterus tissue tumors (Figure 20B). In Figure 20A, both interface mutations PIK3R1^{R348} and PIK3R1^{R461} are in close proximity of the interface mutations PIK3CA^{E542} and PIK3CA^{E545}. In our previous study about same gene double mutations [189], 6 patients with the double mutations PIK3CA$^{R88/E542}$ and PIK3CA$^{R88/E545}$ (one patient carries both doublets). We also observe that PIK3CA^{R88}/PIK3R1^{R348} and PIK3CA^{R88}/PIK3R1^{R461} double mutant tumors do not carry any PIK3CA^{E542} or PIK3CA^{E545} mutations. Also, these two PIK3CA-PIK3R1 double mutations are mutually exclusive with same gene double mutations of PIK3CA in the pan-cancer dataset (Figure 20C). This suggests that in the absence of helical domain mutations PIK3CA^{E542} or PIK3CA^{E545} on the interface, the interface mutations PIK3R1^{R348} and PIK3R1^{R461} can cooperate with PIK3CA^{R88} to over activate PIK3CA.

4.2.3 Double mutations are mostly present in primary tumors and some rare doubles are signature of metastatic tumors

GENIE has samples from primary and metastatic tumors and the mutational profile may have different patterns in these two groups. We found that more than 70% of all double mutations exist in primary tumors. However, there is a small set of double mutations that are specific to metastatic tumor samples. We applied the frequency pattern growth tree approach (see Methods) to obtain the association rules of molecular alterations in metastatic tumors (Figure 21A).

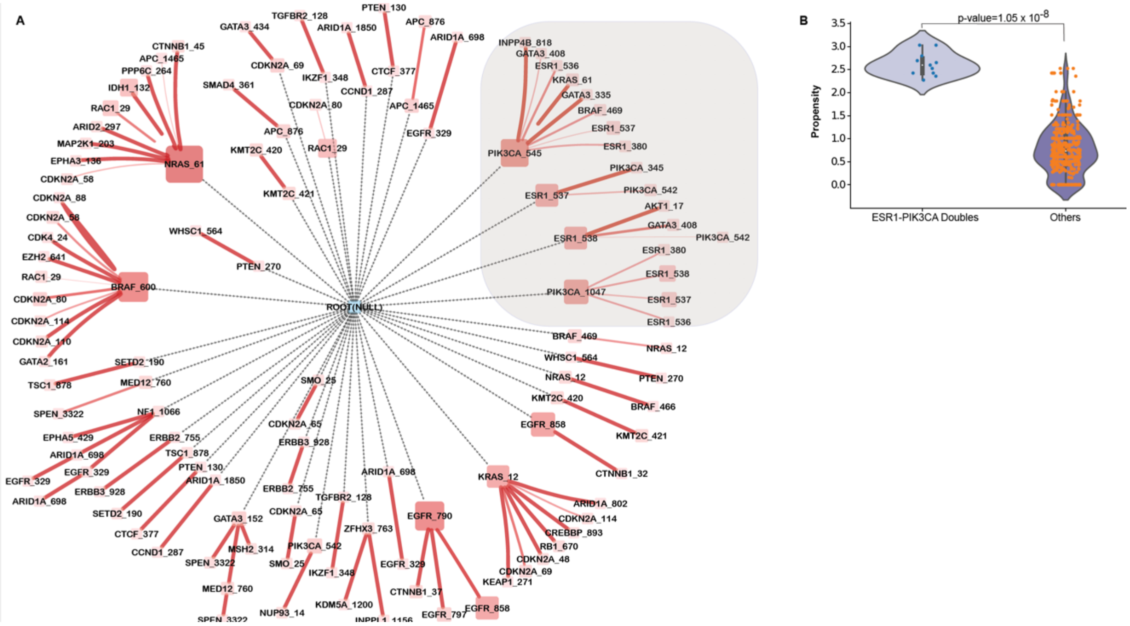


Figure 21. Some rare double mutations are specific to metastatic tumors. **(A)** Frequency pattern growth tree of missense mutations in metastatic tumors where links represent association between mutations. Different gene double mutations in ESR1, GATA3, and PIK3CA genes are enriched in metastatic tumors. **(B)** Propensity of double mutations from ESR1-PIK3CA pair and from other genes where one side is from ESR1 or PIK3CA in metastatic tumors.

We identified a strong association between mutations in ESR1, GATA3, and PIK3CA genes in metastatic tumors. ESR1 mutations at positions 536, 537, and 538, which are sequence neighbors, are exclusively paired with the major drivers of PIK3CA at positions 1047, 542, and 545 (Figure 21B). ESR1 (Estrogen Receptor 1) mutations coupled with PIK3CA mutations mostly exist in metastatic tumors. We also calculated the propensity of dual mutations in ESR1 and PIK3CA to be in metastatic tumors and doublets of each mutation with mutations in other genes. The result strongly shows that mutation pairs, one from ESR1 and one from PIK3CA are a marker of metastatic tumors (Figure 21A, p-value $< 10^{-13}$). GATA3 mutation at position 408 is coupled with ESR1 or PIK3CA mutations. GATA3 is also functionally related and cooperates with ESR1 in transcriptional regulation [190]. Additionally, PAK1 amplification co-exists with ESR1 mutations in metastatic tumors which is either absent or very rare in primary tumors. Another interesting result is the tendency of EGFR dual mutations to be in metastatic samples. Positions 790 and 797 are gatekeeper residues and mutations at these positions promote resistance to RTK inhibitors in lung cancer. We noted that EGFR790 and EGFR797 dual mutations are significantly accumulated in metastatic samples (14 out of 17 samples). Additionally, more than half of the samples having EGFR R790/R858 double mutations are from metastatic origin which is also shown in the FP-Growth Tree in Figure 21A.

In supplementary data of the study [191], there is a list of actionable mutations for 1106 metastatic tumors from the Hartwig Medical Foundation (HMF) data set. Among 393 metastatic breast tumors, 22 harbor one PIK3CA and one ESR1 mutation. The mutations at the positions 542, 545, H1047 of PIK3CA confirm a double mutation with one of the mutations at positions 380, 536, 537 and 538 on ESR1, and determined to be metastatic markers by the FP Growth algorithm.

4.2.4 Functionally equivalent alterations do not co-exist in tumors

Alterations may have similar phenotypic outcomes. We dub them ‘functionally equivalent’. According to our analysis, these alterations either rarely or never coexist in tumor tissues and appear mutually exclusive. Exclusive alterations have similar signaling outputs. We provide two examples below. $KRAS^{G12D}$ is coupled with alterations in a context-specific way. $KRAS^{G12D}$ and the deep deletion of CDKN2A predominantly co-exist in pancreatic cancers. Mutually exclusive alterations may exist in different pathways but result in similar phenotypic effects. Although dual mutations are extremely rare in KRAS, they promote a downstream alteration that leads to proliferation phenotype. Among these alterations, mutations in SMAD4, CDKN2A, U2AF1, and GNAS are mutually exclusive and rarely coexist. In Figures 22A and 22B, we show the presence of these mutations in $KRAS^{G12D}$ mutated PAAD tumors.

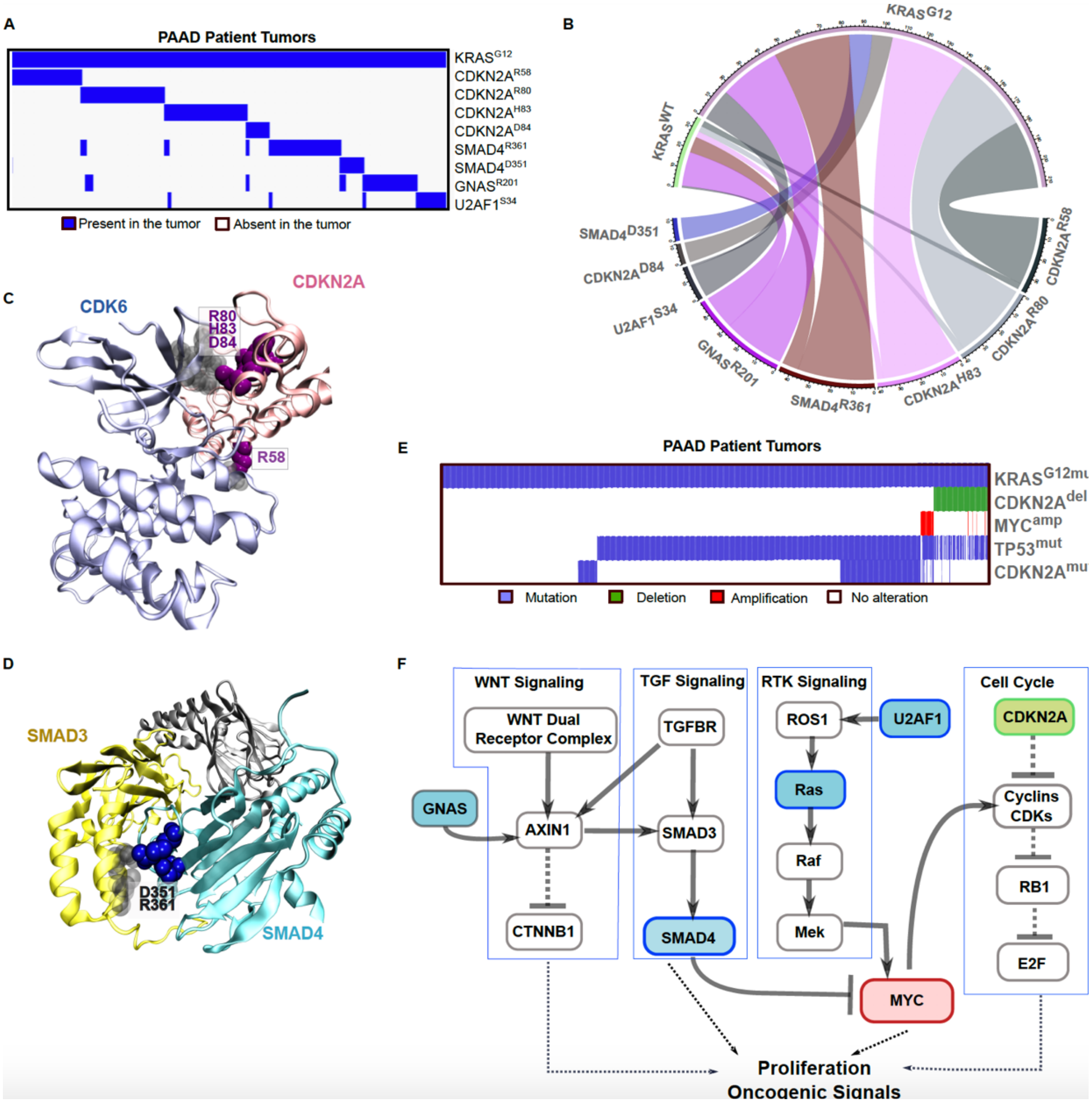

A
PAAD Patient Tumors
KRAS G12
CDKN2A R58
CDKN2A R80
CDKN2A H83
CDKN2A D84
SMAD4 R361
SMAD4 D351
GNAS R201
U2AF1 S34
Present in the tumor
Absent in the tumor
B
KRAS G12
KRAS WT
SMAD4 D351
CDKN2A D84
U2AF1 S34
GNAS R201
SMAD4 R361
CDKN2A H83
CDKN2A R80
CDKN2A R58
C
CDK6
CDKN2A
R80
H83
D84
R58
D
SMAD3
SMAD4
D351
R361
E
PAAD Patient Tumors
KRAS G12mu
CDKN2A del
MYC amp
TP53 mut
CDKN2A mu
Mutation
Deletion
Amplification
No alteration
F
WNT Signaling
WNT Dual Receptor Complex
GNAS
AXIN1
CTNNB1
TGF Signaling
TGFBR
SMAD3
SMAD4
RTK Signaling
ROS1
U2AF1
Ras
Raf
Mek
MYC
Cell Cycle
CDKN2A
Cyclins CDKs
RB1
E2F
Proliferation
Oncogenic Signals

Figure 22. Different gene double mutations in PAAD tumors. **(A)** Oncoprint of the mutations that coexist with KRAS^{G12}. KRAS^{G12} partners on CDKN2A, SMAD4, GNAS, U2AF1 shows a mutually exclusive pattern. **(B)** Circular plot of the mutations in KRAS^{wt} and KRAS^{G12} showing tumor frequencies. **(C)** CDKN2A/CDK6 complex. The mutations R58, R80, H83, and D84 are on the interface region between CDKN2A and CDK6**. (D)** SMAD4 mutations D351 and R361 are in the interface between SMAD4/SMAD3 complex. **(E)** Oncoprint of the most frequent alterations that coexist with KRAS^{G12}. A TP53 mutation on the DNA binding domain accompany to KRAS^{G12} and this couple is followed by either MYC amplification or CDKN2A deletion. **(F)** A subnetwork of highly frequent alterations coexisting with KRAS^{G12} leading to proliferation and oncogenic signal.

CDKN2A and SMAD4 mutations are in binding regions and rarely co-occur in the proteins in tumors. CDKN2A mutations (R58*/Q, R80*, H83N/R/Y, and D84G) are in the interface with CDK6 (PDB: 1BI7) (Figure 22C). All, except R58*/Q, are spatially clustered. Disruption of CDKN2A-CDK6 interaction inhibits their ability to interact with cyclins D and phosphorylate the retinoblastoma protein thus proliferation is positively regulated. SMAD4 mutations at positions D351 and R361 are located in the interface between SMAD4/SMAD3 heterodimer [192] (PDB: 1ygs (dimer), 1U7F (trimer)) (Figure 22D). These mutations disrupt both homo- and hetero-oligomerization of SMADs. All mutually exclusive mutations shown in Figure 22A and 22E are on non-overlapping paths linked to MYC and their functional outcome may be similar. U2AF1 and GNAS paths are linked to KRAS and downstream proliferation phenotype (Figure 22F).

Co-occurrence of MYC amplification and CDKN2A deletion is extremely rare. MYC amplification promotes proliferation while tumor suppressor gene CDKN2A functions as a negative regulator of the proliferation of normal cells. They cooperate as shown by at least one TP53 mutation in the DNA-binding domain. This result also supports that multiple alterations on similar paths resulting in similar phenotype or on similar functional sites do not coexist as potent alterations can promote oncogene-induced senescence.

Another example of exclusive mutations relates to NFE2L2 and KEAP1 (Figure 23A-B). These mutations significantly co-occur with KRAS$^{G12D/V/C}$ in lung tissue. NFE2L2 is a transcription factor and a missense mutation at position 79 abolishes the binding of KEAP1 to NFE2L2. NFE2L2 level is tightly regulated by KEAP1 which is a substrate recruiter in ubiquitin-ligase complex [193]. Mutations on NFE2L2 and KEAP1 may disrupt their interaction increasing NFE2L2 transcriptional activity, eliciting senescence. Lung tissue-specific mutations in KEAP1 exist at positions 271, 272, 320, 333, and 417. Among them, 333 and 417 are located on the Kelch motif. NFE2L2 mutations at positions 24, 29, and 79 are also lung tissue specific. KEAP1 mutations are annotated as loss of function and NFE2L2 mutations are annotated as the gain of function in OncoKB. They are present both in primary and metastatic lung tumors. Mutations in KEAP1 are predominantly in lung tissue and are not simultaneously present in the same tumor sample. Transcriptional and siRNA experiments also show that KRAS^{G12D} up-regulates NRF2 transcription by Myc [194]. The oncogenic activity of NRF2 is context-specific [193] and appears linked to enhanced oncogenicity and drug resistance [195]. In Figure 23C, we applied a personalized PageRank algorithm to obtain the key nodes in the KRAS-NFE2L2-KEAP1 axis and constructed a network. This network clearly shows exclusive subnetworks and shared proteins between KRAS and NFE2L2-KEAP1.

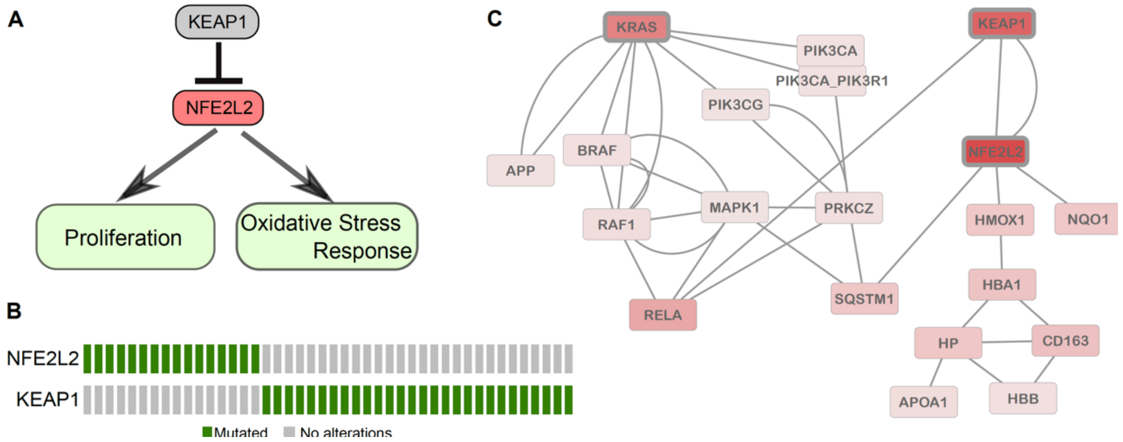


Figure 23. Detailed analysis of KEAP1-NFE2L2 alterations (A) KEAP1 is a ubiquitin ligase which functions in degradation of NFE2L2. Dysregulation of the interaction between NFE2L2 and KEAP1 results in proliferation and oxidative stress response. **(B)** Oncoprint of the mutation presence (green) in KRAS^{G12} mutant patient tumors having at least one mutation in KEAP1 or NFE2L2 which clearly shows their exclusiveness. **(C)** KRAS-KEAP1-NFE2L2 originated subnetwork constructed by pagerank algorithm.

4.3 Concluding remarks

In this chapter, we carry out a statistical approach to identify co-occurring mutations *in trans*. Mutations affecting the same pathway are usually mutually exclusive in a tumor to prevent functional redundancy, synthetic lethality [37,178,196] and OIS [106], considering the additivity of the mutational impact. We also observe this pattern in protein interfaces that double hotspot mutations in the same interface are exclusive i.e., CDKN2A R58, R80, H83 and R84. Some of the *in trans* double mutations that are prominent in specific tissues or cancer subtypes are associated with poor overall survival. For example, co-occurring mutations in IDH2/NPM1, IDH1/NPM1, DNMT3A/NPM1 pairs are significantly prevalent in AML tumors which have been recently associated with poorer overall survival [197]. Another combination is mutated RNF43 together with BRAF^{V600E} which significantly co-occur in our dataset. This cooperativity was defined as a marker of aggressive right-sided colorectal cancer (RCRC) subtype [198]. Co-occurring mutations in PIK3CA and ARID1A in uterine endometrioid carcinoma (UEC) subtype raise in sensitivity to PI3K inhibitors [199].

We also adopted the fp-growth tree method, which has been developed for database systems, to find the association rules between single point mutations, deletions, and amplifications events. Our analysis revealed association with metastatic tumors. We also discover that PIK3CA and ESR1 mutations co-occur in metastatic tumors and double mutations in EGFR are more frequent in metastatic tumors compared to primary tumors.

As noted earlier, and reemphasized by our work, advanced computational approaches are crucial to reveal and forecast such important patterns in cancer progression. Vast amounts of data are being generated. Interpreting it and extracting patterns is vital for precision oncology. Co-occurring mutational signatures for metastatic tumors, coupled with identification of the respective proteins and pathways, can create interactive maps. Linking them with drugs can create immensely useful tool for the attending physicians [120,200].

CHAPTER 5

NEURODEVELOPMENTAL DISORDERS AND CANCER NETWORKS SHARE PATHWAYS; BUT DIFFER IN MECHANISMS, SIGNALING STRENGTH, AND OUTCOME

In this chapter, we aim to uncover the similarities and differences between neurodevelopmental disorders (NDDs) and cancer. We expect that these will help us understand the challenging question of how mutations in the same pathways, and even the same proteins can lead to NDDs and cancer, with vastly different phenotypic presentations. Especially, we aim to discover what are the determining features deciding whether the major outcome is NDD or cancer. We address this inspiring and daunting goal by comprehensively leveraging mutations, transcriptomic data, and protein-protein interaction (PPI) networks. We use *de novo* mutations in ~10,000 samples with NDDs from denovo-db and somatic mutations of ~10,000 tumor samples from The Cancer Genome Atlas (TCGA). We compare the effects of mutations (at the same position and the same/different mutant amino acid) on the pathogenicity of commonly mutated genes in NDDs and cancer. We observe that mutations in NDDs tend to be weak. To evaluate the pathway-level properties of NDDs and cancer, we reconstructed the disease-specific networks for ASD and breast cancer and identified 23 common transcription factors (TFs). Most of the targets of these common TFs are mutated in both ASD and breast cancer and involved in MAPK, cell cycle, and PI3K/AKT pathways. By using transcriptomic profiles of ASD and breast/brain/kidney cancers we show that in breast cancer samples there is an increase in signaling strength in shared pathways involved in proliferation and a decrease in differentiation. This however is not the case among ASD samples where the signaling level is high in shared pathways involved in differentiation and low in proliferation. This chapter is part of the preprint [42].

5.1 Methods

5.1.1 Data collection and processing

NDD mutations were obtained from denovo-db [155] which holds a collection of human germlines *de novo* variants of 20 phenotypes including but not limited to ASD, and intellectual disability NDDs. Variants from two ASD studies were collected by targeted sequencing of different patients coming from two different studies, while the remaining datasets come from either whole exome or whole genome studies. The phenotypes, the number of samples, unique mutated genes and unique mutations are given in Figure 1B. We mapped the genomic coordinates to the proteins to obtain the amino acid changes on the protein level using VarMap [201]. We only kept the point mutations that map to the canonical protein sequences. After these filtering steps, we obtained a total of 14,133 unique mutations on 7907 genes from 9737 samples.

5.1.2 TCGA

Somatic missense, nonsense and frameshift cancer mutations were downloaded from TCGA. There are 9703 tumor samples from 33 different cancer types in the annotation file where corresponding protein changes are also present. In total, we have 1,626,715 unique mutations on 19,438 genes. 7837 of these genes are also mutated in the NDD dataset. 11,601 of them are only mutated in TCGA, while there are only 70 genes that are mutated solely in NDDs.

5.1.3 Cancer drivers

A list of cancer driver mutations was downloaded from the Cancer Genome Interpreter (CGI) [9], which is available as the Catalog of Validated Oncogenic Mutations in their website. We only used the missense or nonsense mutations resulting in an analysis of 3688 driver mutations belonging to 237 genes.

5.1.4 Visualization of mutations in protein sequences and 3D structures

We utilized the ProteinPaint tool [202] to show NDD and cancer mutations on PTEN and PI3Kα. To map the mutations to the 3D structures of PTEN (PDB: 1D5R [203]) and PI3Kα (PDB: 4OVV [204]) we used PyMol [205].

5.2 Results

5.2.1 NDD and cancer mutations can be shared, but their presentation and phenotypic damage differ

NDDs and cancer are highly complex diseases caused by impairments in cellular processes such as cell growth, proliferation, and differentiation. This challenging complexity has led to the community's desire to understand how their genetics, cellular environment, and the signaling pathways are converging to express their distinct phenotypic outcomes [62,153,206,207]. Cancer results from gene alterations that provide cells with a growth advantage. Whereas numerous studies focused on the connection between the mutations -germline, *de novo*, or somatic- and cancer [208–213], the number of studies related to NDDs increased, though still lagging behind, far from reaching the same level. Qi et al. observed that among patients with NDDs, germline damaging *de novo* variants are more enriched in cancer driver genes than non-drivers [206]. Bioinformatics analyses conducted on 219 cancer-related genes from Online Mendelian Inheritance in Man (OMIM, https://www.omim.org/about) and *de novo* mutations from 16,498 patients with NDDs, including ASD, congenital heart disease, and intellectual disability, had significantly more *de novo* mutations in cancer-related genes than in the 3391 controls [214]. In another study focusing on ASD, an evolutionary action method identified missense *de novo* variants that are most likely to contribute to the etiology of the disorder [215].

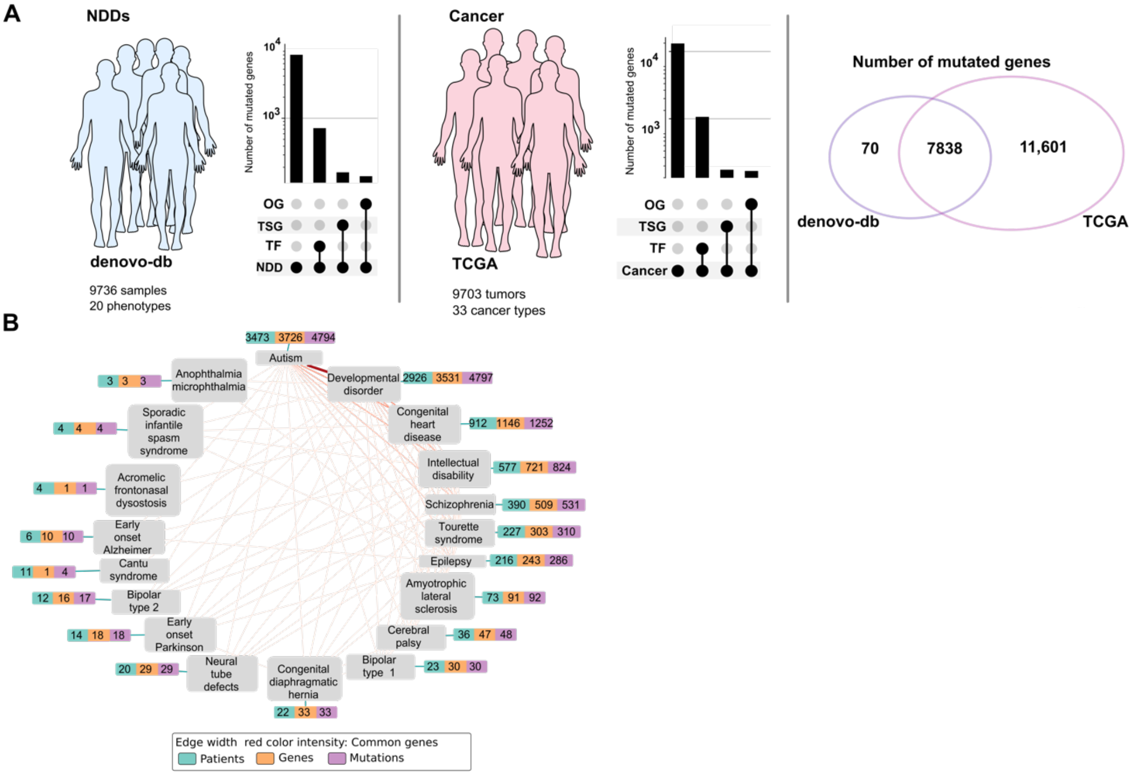


Figure 24. Overview of the data and workflow. (**A**) Statistics from NDDs and cancer datasets. Denovo-db deposits mutation profiles of 9736 samples with NDDs across 20 phenotypes (*left panel*). TCGA provides mutation profiles of 9703 tumors across 33 cancer types (*middle panel*). The length of each bar (*y*-axis in a logarithmic scale) in the upset plots shows the number of all mutated genes and the number of TFs, TSGs, OGs among the mutated genes for NDDs (*left panel*) and cancer samples (*middle panel*). There are 712 TFs, 162 TSGs, and 147 OGs out of 7907 mutated genes among NDD samples. Similarly, there are 1579 TFs, 259 TSGs, and 249 OGs out of 19,438 mutated genes among the cancer samples. The Venn diagram (*right panel*) shows that there are 7837 common mutated genes between NDDs and cancer; the number of NDD- and cancer-

specific mutated genes are 70 and 11,601, respectively. Abbreviations: TSG, tumor suppressor gene; OG, oncogene. (**B**) Network of NDD phenotypes. Each node represents one phenotype in the network, and each edge represents the connection between two phenotypes if they share at least one commonly mutated gene. Each phenotype is represented with a vector of three numbers; the total number of patients having the phenotype (cyan), total number of genes carrying at least one mutation (orange), and total number of mutations associated with the phenotype (purple). The ticker edges represent the more commonly mutated genes. The most tightly connected pair among the phenotype pairs is autism and developmental disorder.

To identify genetic similarities and differences between NDDs and cancer, firstly we utilized publicly available mutation datasets. Public databases provide somatic mutation profiles of thousands of NDD and tumor samples, including denovo-db and TCGA, respectively. Denovo-db includes *de novo* mutation profiles for 20 different NDD phenotypes for 9736 samples [155]; TCGA covers 9703 samples with point mutations across 33 tissues (Figure 24A). Not all genes and variations in their protein products affect the phenotypic output in the same way. Oncogenes, tumor suppressors, TFs, and chromatin remodelers are well-known examples of specific genes whose defects can cause observable alterations in phenotypic outcomes. We compared mutations and mutated proteins between *de novo* mutations in NDD data deposited in the denovo-db and TCGA, focusing on point mutations that affect only one residue in a protein. We identified 7908 genes in NDDs and 19,439 genes in TCGA with point mutations, among which 7838 genes are common. There are 147 oncogenes, 167 tumor suppressor genes, and 712 TFs in the NDD data, while 248 oncogenes, 259 tumor suppressor genes, and 1579 TFs are in TCGA. ~40% of the mutated genes in TCGA also have mutations in NDD samples.

The network of NDD phenotypes in the denovo-db database covers 20 NDD phenotypes with a varying number of patients, mutated genes, and mutations (Figure 24B). Only two of these phenotypes –autism and developmental disorders– have more than 1000 samples. In autism, there are 3473 patients, 3726 mutated genes, and 4794 mutations; in the 2926 developmental disorder samples, there are 3531 mutated genes with 4797 mutations. In the network, the width of edges between the phenotypes is commensurate with the number of commonly mutated genes; autism and developmental disorder share the most. Congenital heart disease and intellectual disability have less than 1000 samples, 912 and 577, respectively. The remaining 15 phenotypes, including schizophrenia, epilepsy, and cerebral palsy, have less than 500 samples.

Our premise is that NDD mutations offer modest but prolonged signaling, whereas cancer mutations are associated with high signaling levels [41,62,216,217]. Driver mutations are frequent, which is why they are often identified as drivers unless there is experimental data for potent rare mutations [5,25]. Weaker or moderate mutations occur less frequently; otherwise, they are drivers. Similarly, the difference between passenger and driver mutations is also based on the statistics; their counts are low. As one indicator of mutation strength, we compared the frequency of the cancer driver mutations in TCGA and NDD mutations amongst TCGA samples. For cancer driver mutations, we used the Catalog of Validated Oncogenic Mutations from the Cancer Genome Interpreter (CGI) [9]. Only missense or nonsense mutations were included in the analyses, which comprised 3688 driver mutations in 237 genes. Among 14,133 unique NDD mutations, 1504 are in TCGA (Figure 25A). On the other hand, TCGA harbors 1060 unique driver mutations. Interestingly, only 23 mutations are shared across known cancer driver mutations and NDD (see inset Venn diagram of Figure 2A). This finding suggests that although there are shared mutations between the two pathologies, these mutations tend to be on the weaker side in terms of a driver effect. In addition, compared to driver mutations, the mutations present in both NDDs and TCGA are notably rare in the TCGA cohort, as demonstrated by the difference between the mutation frequency distribution in TCGA with a *t*-test ($p = 0.001$). Therefore, when we limit the mutations to those present in TCGA, only ~1% of NDD mutations are cancer drivers, and they have very low frequency among TCGA samples. Figure 25B depicts the number of mutated samples in commonly mutated genes among NDDs and cancer. Most commonly mutated genes have more mutation hits at different positions among all cancer samples. Our observations point to only relatively few common NDD and cancer driver mutations, making it crucial –even if difficult– to understand the mechanisms through which these common mutations impact gene function and disease phenotypes. We used pathogenicity scores from MutPred2 [218], which probabilistically predict the impact of variants on protein structure and function. We anticipate that variants may have impact on protein structure, which can either stabilize or destabilize the conformation of the protein depending on protein function and disease phenotypes. The more harmful a mutation is, the closer its pathogenicity score is to one. A comparison of the distribution of the pathogenicity scores of the NDD and driver mutations calculated using MutPred2 demonstrates that drivers have higher pathogenicity than NDD mutations (*t*-test, $p < 5 \times 10^{-30}$) (Figure 25C). We observe that most driver mutations accumulate in regions where the pathogenicity scores are larger than 0.8 on the *y*-axis. NDDs harbor mutations in key cancer genes such as *PTEN*, *PIK3CA*, *MTOR*, *KIT*, etc. These mutations have lower frequencies among tumor samples from TCGA, which is an indicator of the lower potency of these mutations. The number of residues hit by mutations among NDD samples is usually lower.

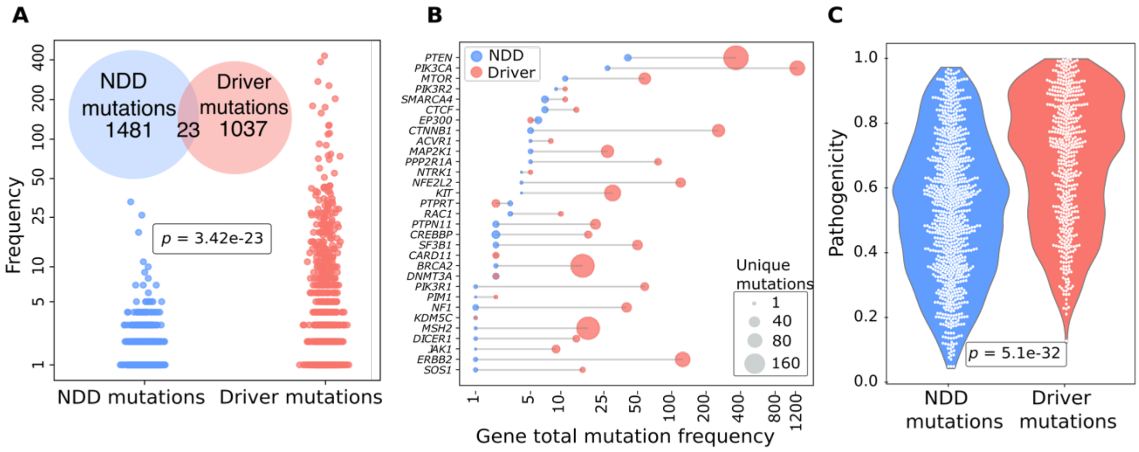


Figure 25. Comparison of mutations between NDDs and cancer. (**A**) Frequency-based analysis of mutations for NDDs and cancer. The cancer driver mutations in TCGA in comparison to the frequency of NDD mutations. The cancer driver mutations were selected amongst tumor samples only. Among the cancer mutations in TCGA, 23 mutations are shared between NDD and known cancer driver mutations, while 1481 are NDD-specific and 1037 are cancer-specific mutations (inset Venn diagram). Comparison of the frequency of these mutations in the TCGA cohort (*y*-axis in a logarithmic scale, where frequency=log10N+1, and *N* is the number of patients). The difference between mutation frequency distribution in TCGA with *t*-test shows that the mutations present in both NDDs and TCGA are significantly rare in the TCGA cohort when compared to driver mutations ($p < 0.001$). (**B**) Frequency of mutations on common genes in NDDs and known cancer drivers datasets. The dumbbell plot shows the mutation frequencies of common genes–the genes harboring at least one point mutation among NDDs and cancer samples–in cancer (TCGA) and NDDs (denovo-db) simultaneously. Cancer driver mutations (red) are more frequent than or equal to NDD mutations (blue) except *EP300* and *PTPRT*. The size of the circles represents the number of unique mutations each gene carries. The *x*-axis in a logarithmic scale represents the number of patients having at least one mutation in the corresponding gene in TCGA or NDD sets. (**C**) MutPred2 pathogenicity scores of NDDs and cancer driver mutations. Violin plots show the distribution of NDD and driver mutation pathogenicity scores. A comparison of the pathogenicity scores using a *t*-test shows that the pathogenicity of driver mutations is significantly higher ($p < 0.001$). Pathogenicity scores are between 0 and 1, where 1 is the most pathogenic.

5.2.2 Mechanistic interpretation of NDD and cancer mutations

PTEN phosphatase and PI3Kα lipid kinase are respectively negative and positive regulators in the PI3Kα/AKT/mTOR pathway. PTEN dephosphorylates phosphatidylinositol 3,4,5-trisphosphate (PIP_3) to phosphatidylinositol 4,5-bisphosphate (PIP_2) produced by PI3K. The signaling lipid PIP_3 recruits AKT and PDK1 (phosphoinositide-dependent kinase 1) protein kinases to the plasma membrane, thereby playing a vital role in cell growth, survival, and migration [167,219]. Loss of function of PTEN by germline or somatic mutations leads to increased PIP_3 concentrations at the membrane and promotes cell proliferation mediated by PI3Kα. Since the PI3Kα/AKT/mTOR pathway is one of the primary regulators of cell proliferation, the mechanistic hallmarks of the mutations are vital to understand. Analysis of mutations in PTEN (Figure 26A) and PI3Kα (Figure 26B) sequences reveals that NDD mutations on these proteins usually occur at less frequently mutated sites among tumors (see Methods). R130* mutation in NDD on PTEN is an exception, yet it is less frequent compared to the R130Q and R130G mutations at the same position in cancer.

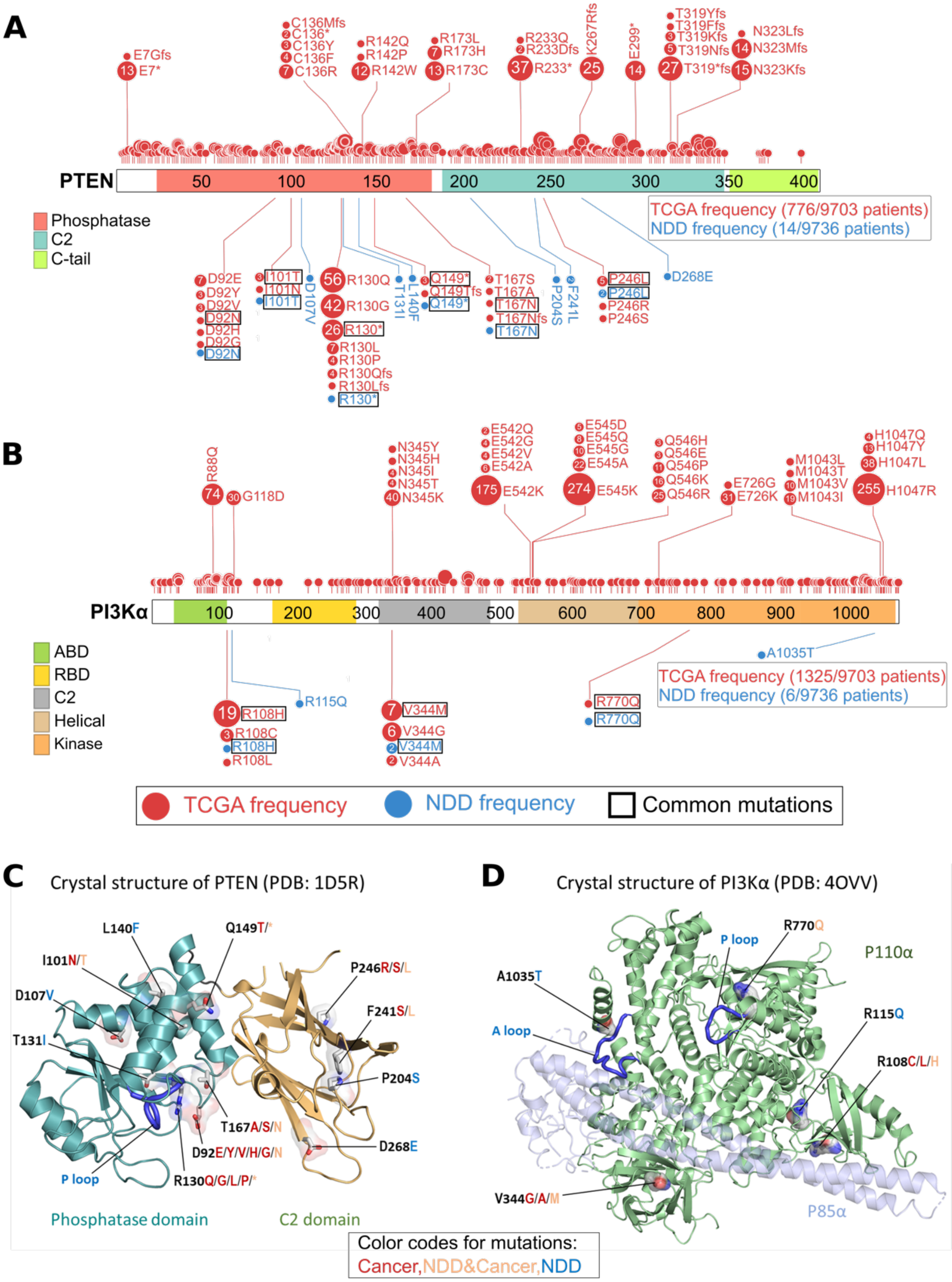

A
E7Gfs
13 E7*
C136Mfs
C136*
C136Y
C136F
C136R
R142Q
R142P
12 R142W
R173L
R173H
13 R173C
R233Q
R233Dfs
37 R233*
K267Rfs
25
E299*
14
T319Yfs
T319Ffs
T319Kfs
T319Nfs
27 T319*fs
N323Lfs
14 N323Mfs
15 N323Kfs
PTEN
50
100
150
200
250
300
350
400
Phosphatase
C2
C-tail
TCGA frequency (776/9703 patients)
NDD frequency (14/9736 patients)
D92E
D92Y
D92V
D92N
D92H
D92G
D92N
I101T
I101N
I101T
D107V
56 R130Q
42 R130G
26 R130*
R130L
R130P
R130Qfs
R130Lfs
R130*
T131I
L140F
Q149*
Q149Tfs
Q149*
T167S
T167A
T167N
T167Nfs
T167N
P204S
F241L
P246L
P246L
P246R
P246S
D268E
B
R88Q
74
30 G118D
N345Y
N345H
N345I
N345T
40 N345K
E542Q
E542G
E542V
E542A
175 E542K
E545D
E545Q
E545G
E545A
274 E545K
Q546H
Q546E
Q546P
Q546K
25 Q546R
E726G
31 E726K
M1043L
M1043T
M1043V
M1043I
H1047Q
H1047Y
38 H1047L
255 H1047R
PI3Kα
100
200
300
400
500
600
700
800
900
1000
ABD
RBD
C2
Helical
Kinase
19 R108H
R108C
R108H
R108L
R115Q
7 V344M
6 V344G
V344M
V344A
R770Q
R770Q
A1035T
TCGA frequency (1325/9703 patients)
NDD frequency (6/9736 patients)
TCGA frequency
NDD frequency
Common mutations
C
Crystal structure of PTEN (PDB: 1D5R)
L140F
Q149T/*
I101N/T
D107V
T131I
P246R/S/L
F241S/L
P204S
T167A/S/N
D92E/Y/V/H/G/N
D268E
P loop
R130Q/G/L/P/*
Phosphatase domain
C2 domain
D
Crystal structure of PI3Kα (PDB: 4OVV)
P loop
R770Q
P110α
A1035T
A loop
R115Q
R108C/L/H
V344G/A/M
P85α
Color codes for mutations:
Cancer,NDD&Cancer,NDD

Figure 26. Profiles of TCGA and NDDs mutations for PTEN and PI3Kα at the residue level on the sequence and structure. (**A**) Mutations of PTEN are shown as circles where phosphatase domain (red), C2 domain (dark green), and C-tail (light green) are represented as colored boxes along the sequence. The number and the size of the circle represents the frequency of each mutation in NDD (blue) or TCGA (red) datasets. Mutations shared by both datasets are highlighted with rectangular borders for emphasis. Total mutation frequencies and the total number of patients in each dataset are shown in the bottom right box. Nonsense mutations are abbreviated with star (*) sign. 6 of 12 PTEN mutations in the NDD set are present in TCGA. Only R130* has a high frequency relative to other shared mutations, yet it is much less frequent when compared to two other TCGA mutations on the same position, R130Q and R130G. (**B**) Mutations of PI3Kα (*PIK3CA*) are shown as circles where ABD (green), RBD (yellow), C2 domain (gray), helical domain (light orange), and kinase domain (orange) are represented as colored boxes along the sequence. The number and the size of the circle represents the frequency of each mutation in NDD (blue) or TCGA (red) datasets. Mutations shared by both datasets are highlighted with rectangular borders for emphasis. Total mutation frequencies and the total number of patients in each dataset are shown in the bottom right box. 3 of 5 PI3Kα mutations in the NDD set are present in TCGA. None of these TCGA mutations are on the most frequently mutated residues or among the most frequent mutations. Abbreviations: ABD, adaptor-binding domain; RBD, Ras-binding domain. The 3D structure of (**C**) PTEN (PDB: 1D5R) and (**D**) PI3Kα (PDB: 4OVV) with selected NDD and TCGA mutations. For each residue, mutated amino acids are colored in red, blue, or orange if they are present only among cancer, NDDs or both phenotypes, respectively. In PTEN, these mutations are known to affect the functions of protein including loss of phosphatase activity, reduced protein stability at the membrane, and failing to suppress AKT phosphorylation. In PI3Kα, these mutations may interrupt protein activation and reduce protein stability at the membrane.

While several residues of PTEN were mutated in both NDD and cancer, some mutations –such as T131I, L140F, and D268E– are NDD-specific (Figure 26A). As to the domain distribution, among the NDD samples, mutated residues D92, I101, R130, T131, L140, Q149, and T167 are on the phosphatase domain, and P204, F241, P246, and D268 are on the C2 domain (Figure 26C). PTEN's catalytic activity occurs in the phosphatase domain that contains the P loop (residues 123-130) with the catalytic signature motif, $_{123}$HCxxGxxR$_{130}$ (where x is any amino acid). PTEN mutations in the P loop, or nearby, such as at the residues R130 and T131, can directly constrain the P loop, leading to silencing PTEN catalytic activity. The mutation at the residue D92 in the WPD loop (residues 88-98) can disrupt the closed WPD loop conformation that can bring D92 to the active site. D92 is involved in the catalytic activity during the process of hydrolysis to release the phosphate group from Cys124 after transferring it from PIP$_3$. Other PTEN mutations, which are distant from the active site, can allosterically bias the P loop dynamics, reducing protein stability and its catalytic activity. A similar pattern is observed in PI3Kα; the rare mutations R108H, V344M, and R770Q are harbored in both NDD and cancer, while R115Q and A1035T are specific to NDD samples (Figure 26B). V344 is on the C2 domain; R770 and A1035 are on the N- and C-lobes of the kinase domain, respectively (Figure 26D). R770 is located near the P loop, and R108 is on the interface of the catalytic subunit p110α and the regulatory subunit p85α. The mutations at these positions in PI3Kα may promote protein activation and increase protein stability at the membrane, but their mutational effects appear to be weaker than the driver mutations.

Several studies investigated germline mutations in PTEN and their association with tumor susceptibility or developmental disorders. [220–223]. For example, the rare I101T mutation on PTEN is present in NDD and cancer samples. This mutation is identified as related to reduced lipid phosphatase activity and protein stability in a study conducted among 13 patients with PHTS who have autistic features, neurodevelopmental delays, and macrocephaly. The I101T mutant retained almost 30% of the lipid phosphatase activity of the wild-type protein; hence, it might be one of the major causes of tissue overgrowth and autistic appearance [223]. Although available data are limited, PTEN retains its tumor suppressive function in NDD while fully dysfunctional among cancer samples.

5.3 Concluding remarks

Our findings offer a mechanistic interpretation for *PTEN* and *PIK3CA* mutations frequently observed in cancer and NDD samples, which may form a basis for functional and detailed structural analysis, including molecular dynamics simulations. Comparing expression scores of shared pathways by leveraging the transcriptomic profiles of NDD and cancer samples revealed that NDD samples have higher expression scores for genes functioning in differentiation than proliferation. These findings provide an essential step toward understanding the etiology of the two different pathologies, NDD, and cancer. Despite having common signaling pathways, their regulation and differences in signal levels enhance different cell states: proliferation for cancer and differentiation for NDD.

Comparisons of the time windows of NDDs and cancer frequently concluded that while cancer is predominantly caused by somatic mutations and alterations in signaling and transcriptional programs, NDDs are primarily linked to germline mutations that express during embryonic development. A recent study has similarly suggested that mutations in cancer susceptibility genes are not necessarily inherited or somatic; they can also arise throughout embryogenesis as a result of errors occurring during cell division [127]. These *mosaic mutations*, occurring in early embryogenesis, were suspected to be associated with some rare cancers. Genetic changes associated with RASopathies are believed to be often sporadic, not inherited. Along these lines, according to the NCI page [61], this means that typically multiple family members do not share the same NDD.

Different than this view, here our thesis is that inherited and *de novo* mutations (missense or truncation) can be major causes of NDDs such as intellectual disability, ASD, epilepsy, and cancer. As in cancer, more than one mutation is required for observable symptomatic NDDs. Our premise is that family members can harbor these NDD germline mutations; however, they are not diagnosed as having the disorder. Their offspring are however already susceptible to it. Individuals with NDDs have higher probabilities of eventually coming down with cancer, likely due to the preexistence of the mutations in the shared proteins, making them more susceptible. Patients with autism have increased mutation load in genes that drive cancer. We hypothesize that strong driver mutations in cell growth/division pathways are chiefly responsible for uncontrolled cell proliferation in cancer. NDDs' weak/moderate strength mutations may be a reason why an inherited NDD has not been identified in a parent while predisposing an offspring to it. An additional mutation promotes NDD clinical manifestation. It may be inherited from the other parent or emerge during embryogenesis. It may also promote cancer by providing the companion mutations.

Here, we employed de novo mutations in ~10,000 samples with NDDs from denovo-db and somatic mutations in ~10,000 tumor samples from TCGA. We observed that around 40% of the 19,439 mutant genes in TCGA are also altered in NDD samples. 1504 of the 14,133 distinct NDD mutations are present in TCGA. On the other hand, TCGA contains 1060 distinct driver mutations, whereas known cancer driver mutations and NDD only share 23 mutations. This result suggests that common mutations across the two pathologies do exist, although they are typically less potent than cancer drivers. Especially, PTEN and PI3Kα possess a range of mutations scattered through their protein sequences that are either common or disease specific. This work argues for searches for such mutations even in undiagnosed family members and follow their combination in the offspring. It further supports consideration of cancer pharmacologies in NDD-patients.

CHAPTER 6

DISCUSSION

This dissertation brings together research on how mutations contribute to the emergence and development of two different phenotypes: Cancers and NDDs. In order to understand the relationships between mutations occurring in *cis* or *trans* across pan-cancer genomes, our initial goal was to detect double mutations in the same or different gene. Then we broadened our scope to determine the similarities and differences between cancer and NDDs by employing mutation and transcriptomic data. We designed Chapter 3 to discover latent driver mutations based on the premise that *in cis*, latent, and weak mutations can cooperate to enhance the oncogenic signal. In Chapter 4, we identified co-occurring double mutations in different genes that additively can promote tumorigenesis through single or multiple pathways. They are mostly in primary tumors. Rare occurrences can be a signature of metastatic tumors. In Chapter 5, we aim to answer the following questions: (i) How can disorders with radically diverse clinical manifestations be caused by the same biological pathways, proteins, and mutations? (ii) Why do those who have conditions like autism and schizophrenia have a higher risk of developing cancer? Utilizing mutations, transcriptomic information, and protein-protein interaction (PPI) networks all at once, we take on this challenging task.

In the first part of this dissertation, we scan the cancer genome landscapes aiming to identify infrequent driver mutations, i.e., latent drivers. We identified 155 significant, same gene double mutations which are composed of mostly one rare and one frequent. Frequent mutations have been cataloged as strong drivers [5,25,224]. Rare drivers can also be strong drivers. We newly cataloged 140 latent drivers. Even though they may be cancer-wide, coupling with another mutation increases their cancer-type specificity. The load of double mutations in tumor suppressors is significantly higher than in oncogenes, indicating their relative robustness to functional loss.

In our definition, mutations which are statistically frequent and thus labeled as oncogenic hotspots in the literature are strong drivers. Oncogenic mutations in the long tails of the distributions are statistically rare. They can be strong or weak drivers. Mutations that are rare [25] and not yet labeled as oncogenic can be latent drivers. They may or may not be allosteric [225]. Rare drivers can be as potent as frequent drivers. Their low statistical frequencies may simply be an outcome of the computational strategy that has been employed in the calculation [226,227]. They may be tissue, or cell specific, harbored in specific cancers. Apart from repressors, under physiological conditions, the wild-type inactive state is more highly populated than the active state. Driver mutations, whether frequent or rare, destabilize the inactive state and/or stabilize the active state making the active state more populated than the inactive state. Two or multiple driver mutations can destabilize the inactive state to a greater extent than single driver mutation as compared to the active state, shifting the population toward the active state. Especially, the conformational change that they promote may also involve steric hindrance at the drug binding site. However, an allosteric mutation away from the binding site may restore drug efficacy against highly resistant mutants, as observed in *BCR-ABL1* [228]. A latent driver also either destabilizes the inactive state and/or stabilizes the active state, but the relative difference between the states can be smaller. Consequently, on their own their contribution to protein activation is relatively small, hindering their identification. However, the additive contributions of strong drivers or of latent drivers, strong or weak, can increase the population of the active conformations leading to the ensemble being fully activated. Given that the mechanism described here depends on the positions of the constituents in the 3D protein structure and their distance from one another in addition to other factors, it is entirely plausible that it cannot apply to all doublets.

With the sparsity of patient treatment datasets, cell lines or patient-derived tumor xenografts are a useful clinical interpretation resource. We found significant differences in the response to PI3K inhibitors in tumor models that differ in the presence or absence of double mutations in *PIK3CA*, which is in line with recent experimental work [29]. Tumor growth is extremely fast in double mutant *PIK3CA* compared to the single mutant. Recent mechanistic studies suggest that the increased protein activity or acquired drug resistance is due to the mutation combinations. Zhang et al. [40] suggested that combinations of strong and weak drivers can enhance PI3K activity and explain the phenotypic differences in *PIK3CA* double mutant tumors [36] that we observed prominently in breast and uterus tumors. Here we further extended the analysis to combinations of less frequent mutations not catalogued as drivers, which we view as potential latent drivers. Among them, doublets with mutation at position R88 are depleted in breast but not in uterus cancers, suggesting that potential latent driver mutations pairing with R88 are important signatures of uterus tumors.

Not limited to *PIK3CA*, numerous other significant double mutations with possible prognostic or therapeutic impact have also been identified (i.e. *EGFR* in the lung in line with previous studies [30]). To fully understand mutational frequencies requires detailed functional data related to specific mutations, their combinations, and the proteins that harbor them. We, and others, have been aiming to reveal the mechanisms of oncogenic mutations in key protein nodes in the network. The paramount principle that guides us is that the mechanisms of the mutations mimic the physiological activation [229]. However, whereas physiological activation is regulated, taking place following some signaling event, e.g., hormone binding to the extracellular domain of a receptor tyrosine kinase in the case of PI3K, with the signaling propagating downstream through a series of cascading events, oncogenic activation is dysregulated. We thus suggest that the single mutations which are components of doublets can act in one of two ways: their effects can be complementary in relieving the autoinhibition [224,230,231], or can enhance the same effect, for example involving not one positive charge but two for membrane binding. Consider for example PI3K, whose physiological activation involves binding of the phosphorylated C-terminal motif of insulin receptor to the nSH2, resulting in breaking of the interaction of the nSH2 with the helical domain and relieving the autoinhibition, and binding of active Ras, which assists in binding and properly positioning the PI3K on the membrane. E542K and E545K hotspots mimic the action of the first, and H1047R the second. With all being strong hotspots, their co-occurrence can trigger oncogene induced senescence (OIS). However, a combination with more moderate mutations can powerfully activate this lipid kinase. Relieving the autoinhibition is a common physiological activation mechanism that oncogenic mutations adopt [230]. Not all mutations form pairs. One example is *BRAF* V600E. This has been attributed to its being a strong hotspot. Mutant *BRAF* V600E has been postulated to be activated as a monomer independent of Ras activation [232] and shown to be able to phosphorylate MEK [233–235]. However, as we noted above, recent data suggest that even though the mutant is activated as a monomer, a dimeric BRAF is still required to phosphorylate MEK in cells [234,236–238]. Mechanistic arguments clarify that the despite the activating mutation, for cell growth BRAF V600E still requires a collaboration with a Raf partner to have MEK appropriately positioned and retained in the assembly, just as in the case of physiological *BRAF* [236], an observation which is of vital importance in drug discovery aiming at targeting dimerization. This example serves to illustrate the importance of knowledge of the functional activation mechanism which statistics alone is unable to provide [239]. Combined, they may better forecast treatment outcomes. The sensitivity or responsivity of drug action to a targeted cancer depends on how much the tumor relies on the particular oncogene and the cellular pathway with which it is associated. In *PIK3CA*, a combination of a driver mutation with a weak driver, or strong latent driver, particularly under different mechanisms of actions, have a good, albeit temporary, therapeutic response.

In the second part of this dissertation, we screen mutation profiles of pan-cancer tumors to identify double mutations in different genes. We found 4352 statistically significant different gene double mutations that alter non-redundant pathways and interactions and promote cancer-specific tumorigenesis. The genes harboring these double mutations are observed to be involved in exclusive pathways implicating a compensatory relation between the double mutation constituents. Several studies looking for gene-level epistatic interactions have provided fundamental insights into the genetic architecture of cells as well as identified potential new therapeutic interventions [240–242]. However, either synthetic-lethal or cooperating interactions are shown to be context dependent in most of the cases, i.e., the epistatic relations are not preserved across different cancer-types. For example, patients with advanced pancreatic cancer and pathogenic mutations in BRCA1, BRCA2, and PALB2 have shown therapeutic activity with PARP inhibitors [243,244].

The term "epistasis" describes how a mutation depends on one or more additional mutations as well as the overall genetic background. Epistasis has been postulated as a significant contribution to variances in disease outcome in the context of human disorders, where it complicates the range of disease symptoms [245]. Our competence to forecast phenotypic outcomes is constrained by the nonadditive interaction between mutations and the inadequate understanding of the underlying physiological effects.

There are various classes of epistatic interactions in cells depending on the context. If two mutations, when present together, provide a higher phenotypic outcome than would be expected from each of their individual effects, this is referred to as positive epistasis. The outcome synergistic and arises between the mutations in the genes belonging to non-redundant complementary pathways. On the contrary, two genes are synthetic lethal if mutation in both genes impairs cell viability and result in cell death. Synthetic lethality usually arises due to the co-occurring mutations in the genes that are involved in the upstream or downstream of the same pathway or in the protein-protein interfaces. Therefore, if a mutation occurs in one of the interacting gene pairs, the other one needs to be in wild-type form to be able to sustain cell viability [246]. These interactions are attractive for the development of cancer treatments because they allow the targeting of a gene whose synthetic lethal partner is irreversibly inactivated in cancer cells but expresses wild-type in healthy cells.

Here, we carry out analysis of co-occurring mutations in trans. We adopted the fp-growth tree method, which has been developed for database systems, to find the association rules between single point mutations, deletions, and amplifications events. Our analysis revealed association with metastatic tumors. We also discover that PIK3CA and ESR1 mutations co-occur in metastatic tumors and double mutations in EGFR are more frequent in metastatic tumors compared to primary tumors. Mutations affecting the same pathway are usually mutually exclusive in a tumor to prevent functional redundancy, synthetic lethality [37,178,196] and OIS [106], considering the additivity of the mutational impact. We also observe this pattern in protein interfaces that double hotspot mutations in the same interface are exclusive i.e., CDKN2A R58, R80, H83 and R84. Driver mutations co-occurrence (DCO) networks can be helpful in identifying synergy and drug response [247].

Some of the in trans double mutations that are prominent in specific tissues or cancer subtypes are associated with poor overall survival. For example, co-occurring mutations in IDH2/NPM1, IDH1/NPM1, DNMT3A/NPM1 pairs are significantly prevalent in AML tumors which have been recently associated with poorer overall survival [197]. Another combination is mutated RNF43 together with BRAFV600E which significantly co-occur in our dataset. This cooperativity was defined as a marker of aggressive right-sided colorectal cancer (RCRC) subtype [198]. Co-occurring mutations in PIK3CA and ARID1A in uterine endometrioid carcinoma (UEC) subtype raise in sensitivity to PI3K inhibitors [199].

In the third part of this dissertation, we analyzed genomes of NDD and cancer patients asking how and what makes these pathologies similar or different. We comprehensively analyzed mutations, transcriptomic data, and PPI networks. We dissected commonly mutated proteins in NDD and cancer to comprehend why some mutations can promote cancer while others can abet NDDs, and why the same mutations can support both phenotypes. The analysis was carried out under our premise that cancer mutations are connected to elevated signaling levels, while NDD mutations encode sustained but low levels. We further surmised that signaling levels are largely determined by the total number of molecules that the mutations activate, either alone or in combination, along with the cell type-specific expression levels of the mutant protein and other proteins in the relevant pathways, the timing of the emergence of the mutation (inherited or during embryonic development, or sporadic), as well as additional factors [62]. Ample data indicate that even high expression levels of an unmutated protein can provoke cancer. Dysfunction of proteins that are involved in the MAPK and PI3K/AKT/mTOR pathways which harbor germline or somatic mutations, is particularly crucial to determine since these pathways control proliferation and differentiation, thus clinical outcome.

Cancer involves uncontrolled cell proliferation whereas NDDs are connected to anomalies in the development of the nervous system. Cell proliferation and differentiation take place in both cancer and NDDs. Since NDDs are mostly related to dysregulated differentiation, mutations in genes regulating chromatin organization rank high. Risk genes for NDDs include more than a third of the cancer driver genes, and NDDs and cancer share the same cellular pathways, especially MAPK and PI3K/AKT/mTOR, the hallmark of cell division and growth [151,248], thus proliferation and differentiation [62,206]. In brain cells, embryonic mutations in both pathways give rise to NDDs[152]. Hundreds of genes are implicated in NDDs; however, they are involved in few conserved pathways regulating transcription, including chromatin accessibility, and synaptic signaling. PI3K/mTOR and Ras/MAPK are frequently linked with synaptic dysregulation [41,43,62,249,250]. Proteins in the Wnt, BMP/TGF-β (bone morphogenetic protein/transforming growth factor β), SHH (sonic hedgehog), FGF (fibroblast growth factor), and RA (retinoic acid) pathways, are also involved in autistic brain development [251]. Gene expression profiles of 22 cancer types and frontal cortical tissues from ASD patients identified similarities in genes and pathways [154].

NDDs share phenotypic and clinical commonalities. Tumor suppressor phosphatase and tensin homolog (PTEN), which carries inherited (germline) and *de novo* mutations in NDD patients, is related to cancers and several NDDs, collectively named PTEN hamartoma tumor syndrome (PHTS). The NDDs include phenotypes, such as Cowden syndrome (CS), Bannayan-Riley-Ruvalcaba syndrome (BRRS), Proteus syndrome (PS), Proteus-like syndrome (PSL), macrocephaly, and ASD. NDDs often overlap mutation-wise and genome-wise [134–136]. Among these deletions, and duplications of the 16p11.2 region are common. About 48% of deletion carriers and 63% of duplication carriers have at least one psychiatric diagnosis [138,139]. RASopathies, which include Noonan syndrome (NS), cardiofaciocutaneous (CFC) syndrome, neurofibromatosis type 1 (NF1), and Legius syndrome (LS), are NDDs that result from overactivation of the MAPK pathway due to the germline mutations and (or) overexpression in embryogenesis [140,142]. Their phenotypic overlaps may emerge due to shared proteins/pathways as in the case of *PIK3CA*-related overgrowth spectrum (PROS), PS, and CS which share phenotypic characteristics with RASopathies [141]. The commonality of cancer and RASopathies prompted MEK inhibitors and Ras-targeted therapies for some RASopathies like selumetinib for NF1patients [144–146].

Although there is a strong association between PTEN germline mutations and cancer–PHTS–they have also been described in patients with ASDs [133]. PTEN mutations linked to ASD can lead to an unstable but still catalytically active gene product [252]. C124S, G129R, H118P, H123Q, E157G, F241S, D252G, N276S, and D326N are autism-related; A39P, N48K, L108P, L112P, and R130L are PHTS-related mutations [220]. AKT signaling was suppressed in all seven ASD-related PTEN mutations. However, AKT signaling was impaired by all five PTEN mutations in severe PHTS cases, suggesting that variants with partial loss of PTEN function are predominantly in ASD patients [220]. Thus, catalytically inactive PTEN mutant is connected to tumor phenotypes, partially active PTEN to ASD [253,254] .

Dysregulation of the PI3K/AKT/mTOR pathway is a primary factor in NDDs, including megalencephaly (also known as “large brain”), microcephaly (sometimes known as “small brain”), ASD, intellectual disability, schizophrenia, and epilepsy [255]. Mosaic gain-of-function mutations in the *PIK3CA* gene lead to PROS, with clinical outcomes such as excessive tissue growth, blood vessel abnormalities, and scoliosis [256,257]. Among ~200 individuals with *PIK3CA* mosaic mutations, highly activating hotspot mutations were associated with severe brain and/or body overgrowth, whilst fewer activating mutations were linked to more mild somatic overgrowth and mostly brain overgrowth [258,259]. R88Q, V344M, and G914R mutations were identified in PI3Kα patients with macrocephaly and developmental delay or ASD [260].

We further pursued the complex relationship between genotype and phenotype by constructing disease-specific networks of ASD and breast cancer. We observed distinct protein-protein interactions (PPIs) in shared pathways controlling the cell cycle. These rewired interactions could be a reason why shared pathways have different signal strengths in ASD and cancers. Under physiological conditions, MAPK and PI3K/AKT/mTOR pathways coregulate the cell cycle through feedback loops to control cell functions, including growth, and division. In cancers, they are frequently hyperactivated [261–263] The PI3K/AKT pathway is also critical in early embryonic development and maintenance of stem cell pluripotency through inhibition of the MAPK proliferation pathway [264–267]. The strength of the signaling perturbations induced by the mutations is manifested in weak/moderate and strong signaling changes, epitomized by ASD and breast cancer, respectively. Strong signals enhance proliferation, and weak/moderate signals may drive cell cycle exit in differentiation [268].

TF complexes are primarily involved in cell cycle regulation through their targets, such as E2F mediating CDK that accelerates proliferation [269,270]. In breast cancer-specific networks, CDK4 interacts with MAPK1, JAK3, and p53, promoting proliferation [271]. In the ASD-specific network, TF complexes such as forkhead box protein G1 (FOXG1) and sex determining region Y-box 2 (SOX2), also implicated in microcephaly, play critical roles in lineage determination, neural stem/progenitor cell proliferation, and maintenance of pluripotency [272,273]. In NDDs, these TFs can promote premature senescence and dysregulated differentiation via distinct pathways such as Wnt and Hippo [274]. In a study of the English population half of decedents with intellectual disabilities and cancer were at stage IV when diagnosed [275], which suggested involvement of the canonical Wnt pathway during brain morphogenesis, and non-canonical in cancer cell migration and metastasis [276]. Cancer onset in NDD can be undetected until stage IV since the slow cell division in the NDDs retards mutational accumulation [275]. Where statistics are available, the mortality of cancer patients with intellectual disabilities was reported to be approximately 1.5 times higher than the general population [277]. These results suggest that cancer initiation and progression differ in individuals with NDD than in the broad apparent NDD-free population with different outcomes via common pathways.

The expression scores of TFs were grouped based on proliferation and differentiation. TFs enhancing proliferation were mainly overexpressed in cancers while relatively low-expressed in ASD. Proliferating cells are more vulnerable to mutations than differentiating ones, both since dividing cells have less time to repair DNA damage than quiescent cells, and with more replication cycles there is a higher chance for mutations [278] . As to TFs in the differentiation state, ASD has relatively higher expression profiles, while there are significantly low-expression profiles in cancers. In cancers, high expression couples with accumulation of mutations, cell growth, and metastasis [279].

Finally, immunity could be viewed as a common factor in NDDs and cancer [41,62]. Multiple pathways related to immunity can be dysregulated in NDD due to the coevolution of the immune and nervous systems [41,280]. Signaling pathways related to immunity, such as Wnt, Notch, JAK/STAT, and Hippo, also play roles in cancer metastasis and drug resistance [274,281,282].

We believe the findings of this dissertation offer a new perspective to identify double mutations in the same or different genes. Extending the gene-level interactions with the mutation-level interactions would help to discern epistatic interactions and provide a higher resolution for the genetic architecture of the cells. These types of interactions play a vital role in identifying disease subtype, predicting therapeutic response or resistance.
Recently, the mutations in cancer related genes and pathways have implicated to be involved in neurodevelopmental disorders [62,153]. Therefore, investigating the role of mutations, either single or double, interaction of mutations and enriched pathways would give valuable information about the etiology of neurodevelopmental disorders and its similarity/difference to cancer. As noted earlier, and reemphasized by our work, advanced computational

approaches are crucial to reveal and forecast such important patterns in cancer progression and the manifestation of NDDs. Vast amounts of data are being generated. Interpreting it and extracting patterns is vital for precision oncology. Co-occurring mutational signatures for metastatic tumors, coupled with identification of the respective proteins and pathways, can create interactive maps. Linking them with drugs can create immensely useful tool for the attending physicians [120,200].

REFERENCES

1. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer Genome Landscapes. Science (80-). 2013;339:1546–58.

2. McFarland CD, Korolev KS, Kryukov G V., Sunyaev SR, Mirny LA. Impact of deleterious passenger mutations on cancer progression. Proc Natl Acad Sci. 2013;110:2910–5.

3. McFarland CD, Yaglom JA, Wojtkowiak JW, Scott JG, Morse DL, Sherman MY, et al. The Damaging Effect of Passenger Mutations on Cancer Progression. Cancer Res. 2017;77:4763–72.

4. Nussinov R, Jang H, Tsai C-J, Cheng F. Review: Precision medicine and driver mutations: Computational methods, functional assays and conformational principles for interpreting cancer drivers. Panchenko ARR, editor. PLOS Comput. Biol. 2019. p. e1006658.

5. Nussinov R, Tsai C-J. 'Latent drivers' expand the cancer mutational landscape. Curr Opin Struct Biol. 2015;32:25–32.

6. Castro-Giner F, Ratcliffe P, Tomlinson I. The mini-driver model of polygenic cancer evolution. Nat Rev Cancer. 2015;15:680–5.

7. Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. Nat Rev Cancer. 2018;18:696–705.

8. An O, Dall'Olio GM, Mourikis TP, Ciccarelli FD. NCG 5.0: updates of a manually curated repository of cancer genes and associated properties from cancer mutational screenings. Nucleic Acids Res. 2016;44:D992–9.

9. Tamborero D, Rubio-Perez C, Deu-Pons J, Schroeder MP, Vivancos A, Rovira A, et al. Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. Genome Med. 2018;10:25.

10. Mularoni L, Sabarinathan R, Deu-Pons J, Gonzalez-Perez A, López-Bigas N. OncodriveFML: a general framework to identify coding and non-coding regions with cancer driver mutations. Genome Biol. 2016;17:128.

11. Arnedo-Pac C, Mularoni L, Muiños F, Gonzalez-Perez A, Lopez-Bigas N. OncodriveCLUSTL: a sequence-based clustering method to identify cancer drivers. Schwartz R, editor. Bioinformatics. 2019;35:4788–90.

12. Gundem G, Perez-Llamas C, Jene-Sanz A, Kedzierska A, Islam A, Deu-Pons J, et al. IntOGen: integration and data mining of multidimensional oncogenomic data. Nat Methods. 2010;7:92–3.

13. Martínez-Jiménez F, Muiños F, Sentís I, Deu-Pons J, Reyes-Salazar I, Arnedo-Pac C, et al. A compendium of mutational cancer driver genes. Nat Rev Cancer. 2020;20:555–72.

14. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell. 2018;173:371-385.e18.

15. Ramos AH, Lichtenstein L, Gupta M, Lawrence MS, Pugh TJ, Saksena G, et al. Oncotator: Cancer Variant Annotation Tool. Hum Mutat. 2015;36:E2423–9.

16. Chen S, He X, Li R, Duan X, Niu B. HotSpot3D web server: an integrated resource for mutation analysis in protein 3D structures. Elofsson A, editor. Bioinformatics. 2020;36:3944–6.

17. Niu B, Scott AD, Sengupta S, Bailey MH, Batra P, Ning J, et al. Protein-structure-guided discovery of functional mutations across 19 cancer types. Nat Genet. 2016;48:827–37.

18. Dincer C, Kaya T, Keskin O, Gursoy A, Tuncbag N. 3D spatial organization and network-guided comparison of mutation profiles in Glioblastoma reveals similarities across patients. PLoS Comput Biol. 2019;15:e1006789.

19. Porta-Pardo E, Valencia A, Godzik A. Understanding oncogenicity of cancer driver genes and mutations in the cancer genomics era. FEBS Lett. 2020. p. 4233–46.

20. Porta-Pardo E, Hrabe T, Godzik A. Cancer3D: understanding cancer mutations through protein structures. Nucleic Acids Res. 2015;43:D968–73.

21. Evans P, Avey S, Yong Kong, Krauthammer M. Adjusting for Background Mutation Frequency Biases Improves the Identification of Cancer Driver Genes. IEEE Trans Nanobioscience. 2013;12:150–7.

22. Brown A-L, Li M, Goncearenco A, Panchenko AR. Finding driver mutations in cancer: Elucidating the role of background mutational processes. Nussinov R, editor. PLOS Comput Biol. 2019;15:e1006981.

23. Zhang Z, Miteva MA, Wang L, Alexov E. Analyzing Effects of Naturally Occurring Missense Mutations. Comput Math Methods Med. 2012;2012:1–15.

24. Alexov E. Advances in Human Biology: Combining Genetics and Molecular Biophysics to Pave the Way for Personalized Diagnostics and Medicine. Adv Biol. 2014.

25. Nussinov R, Tsai C-J, Jang H. Why Are Some Driver Mutations Rare? Trends Pharmacol Sci. 2019;40:919–29.

26. Donehower LA, Soussi T, Korkut A, Liu Y, Schultz A, Cardenas M, et al. Integrated Analysis of TP53 Gene and Pathway Alterations in The Cancer Genome Atlas. Cell Rep. 2019;28:1370-1384.e5.

27. Bozic I, Antal T, Ohtsuki H, Carter H, Kim D, Chen S, et al. Accumulation of driver and passenger mutations during tumor progression. Proc Natl Acad Sci. 2010;107:18545–50.

28. Risques RA, Kennedy SR. Aging and the rise of somatic cancer-associated mutations in normal tissues. Williams SM, editor. PLOS Genet. 2018;14:e1007108.

29. Vasan N, Razavi P, Johnson JL, Shao H, Shah H, Antoine A, et al. Double PIK3CA mutations in cis increase oncogenicity and sensitivity to PI3Kα inhibitors. Science (80-). 2019;366:714–23.

30. Gorelick AN, Sánchez-Rivera FJ, Cai Y, Bielski CM, Biederstedt E, Jonsson P, et al. Phase and context shape the function of composite oncogenic mutations. Nature. 2020;582:100–3.

31. Saito Y, Koya J, Araki M, Kogure Y, Shingaki S, Tabata M, et al. Landscape and function of multiple mutations within individual oncogenes. Nature. 2020;582:95–9.

32. Saito Y, Koya J, Kataoka K. Multiple mutations within individual oncogenes. Cancer Sci. 2021;112:483–9.

33. Boca SM, Kinzler KW, Velculescu VE, Vogelstein B, Parmigiani G. Patient-oriented gene set analysis for cancer mutation data. Genome Biol. 2010;11:R112.

34. Campbell PJ. Cliques and Schisms of Cancer Genes. Cancer Cell. 2017.

35. Nussinov R, Tsai C-J, Jang H. Are Parallel Proliferation Pathways Redundant? Trends Biochem Sci. 2020;45:554–63.

36. Nussinov R, Zhang M, Maloney R, Jang H. Drugging multiple same-allele driver mutations in cancer. Expert Opin Drug Discov. 2021;16:823–8.

37. van de Haar J, Canisius S, Yu MK, Voest EE, Wessels LFA, Ideker T. Identifying Epistasis in Cancer Genomes: A Delicate Affair. Cell. 2019;177:1375–83.

38. Cisowski J, Sayin VI, Liu M, Karlsson C, Bergo MO. Oncogene-induced senescence underlies the mutual exclusive nature of oncogenic KRAS and BRAF. Oncogene.

2016;35:1328–33.

39. Cavalli G, Biavasco R, Borgiani B, Dagna L. Oncogene-Induced Senescence as a New Mechanism of Disease: The Paradigm of Erdheimâ€“Chester Disease. Front Immunol. 2014;5.

40. Zhang M, Jang H, Nussinov R. PI3K Driver Mutations: A Biophysical Membrane-Centric Perspective. Cancer Res. 2021;81:237–47.

41. Nussinov R, Tsai C-J, Jang H. Neurodevelopmental disorders, immunity, and cancer are connected. iScience. 2022;25:104492.

42. Yavuz BR, Arici MK, Demirel HC, et al. Neurodevelopmental disorders and cancer networks share pathways; but differ in mechanisms, signaling strength, and outcome (preprint), bioRxiv (2023).

43. Nussinov, R, Yavuz, BR, Arici, MK, et al. Neurodevelopmental disorders, like cancer, are connected to impaired chromatin remodelers, PI3K/mTOR, and PAK1-regulated MAPK. Biophys Rev 15, 163–181 (2023). https://doi.org/10.1007/s12551-023-01054-9.

44. McLendon R, Friedman A, Bigner D, Van Meir EG, Brat DJ, Mastrogianakis GM, et al. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature. 2008;455:1061–8.

45. Sweeney SM, Cerami E, Baras A, Pugh TJ, Schultz N, Stricker T, et al. AACR project genie: Powering precision medicine through an international consortium. Cancer Discov. 2017.

46. Auvil JG. Therapeutically Applicable Research to Generate Effective Treatments (TARGET). AACR Annu Meet. 2018.

47. Zhang J, Bajari R, Andric D, Gerthoffert F, Lepsa A, Nahal-Bose H, et al. The International Cancer Genome Consortium Data Portal. Nat. Biotechnol. 2019.

48. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature. 2020;578:94–101.

49. Campbell PJ, Getz G, Korbel JO, Stuart JM, Jennings JL, Stein LD, et al. Pan-cancer analysis of whole genomes. Nature. 2020;578:82–93.

50. Diehl AG, Boyle AP. Deciphering ENCODE. Trends Genet. 2016;32:238–49.

51. Aganezov S, Yan SM, Soto DC, Kirsche M, Zarate S, Avdeyev P, et al. A complete reference genome improves analysis of human genetic variation. Science (80-). American Association for the Advancement of Science; 2022;376.

52. Wang S, Cang S, Liu D. Third-generation inhibitors targeting EGFR T790M mutation in advanced non-small cell lung cancer. J Hematol Oncol. 2016;9:34.

53. Engelbrecht C, Urban M, Schoeman M, Paarwater B, van Coller A, Abraham DR, et al. Clinical Utility of Whole Exome Sequencing and Targeted Panels for the Identification of Inborn Errors of Immunity in a Resource-Constrained Setting. Front Immunol. 2021;12.

54. Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, Ellrott K, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet. 2013;45:1113–20.

55. Gonzalez-Perez A, Tamborero D, Lopez-Bigas N, Margolin A, Ding L, Gordenin D, et al. Thread 1: Mutational drivers. Nat Genet. 2013.

56. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. Cancer Discov. 2012;2:401–4.

57. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. Sci Signal. 2013;6.

58. National Cancer Institute. Genomic Data Commons Data Portal. Harmon. Cancer

Datasets. 2019.

59. Gerasimavicius L, Livesey BJ, Marsh JA. Loss-of-function, gain-of-function and dominant-negative mutations have profoundly different effects on protein structure. Nat Commun. Nature Publishing Group; 2022;13:3895.

60. Stenson PD, Mort M, Ball E V., Evans K, Hayden M, Heywood S, et al. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. Hum Genet. Springer Verlag; 2017;136:665–77.

61. Cancer.gov. NCI Dictionary of Cancer Terms - National Cancer Institute. Https://Www.Cancer.Gov/Publications/Dictionaries/Cancer-Terms. 2016.

62. Nussinov R, Tsai CJ, Jang H. How can same-gene mutations promote both cancer and developmental disorders? Sci Adv. American Association for the Advancement of Science; 2022;8:2059.

63. Pon JR, Marra MA. Driver and Passenger Mutations in Cancer. Annu Rev Pathol Mech Dis. 2015;10:25–50.

64. Reva B. Revealing selection in cancer using the predicted functional impact of cancer mutations. Application to nomination of cancer drivers. BMC Genomics. 2013;14:S8.

65. Sakoparnig T, Fried P, Beerenwinkel N. Identification of Constrained Cancer Driver Genes Based on Mutation Timing. Wang E, editor. PLoS Comput Biol. 2015;11:e1004027.

66. Mao Y, Chen H, Liang H, Meric-Bernstam F, Mills GB, Chen K. CanDrA: Cancer-Specific Driver Missense Mutation Annotation with Optimized Features. Adamovic T, editor. PLoS One. 2013;8:e77945.

67. Poulos RC, Wong JWH. Finding cancer driver mutations in the era of big data research. Biophys Rev. 2019;11:21–9.

68. Cheng F, Liang H, Butte AJ, Eng C, Nussinov R. Personal mutanomes meet modern oncology drug discovery and precision health. Pharmacol. Rev. 2019.

69. Mohsen H, Gunasekharan V, Qing T, Seay M, Surovtseva Y, Negahban S, et al. Network propagation-based prioritization of long tail genes in 17 cancer types. Genome Biol. 2021;22:287.

70. Nussinov R, Jang H, Tsai C-J, Cheng F. Precision medicine review: rare driver mutations and their biophysical classification. Biophys Rev. 2019;11:5–19.
71. Pierotti MA, Frattini M, Molinari F, Sozzi G, Croce CM. Oncogenes. John Wiley & Sons, Ltd; :1–22.

72. Osborne C, Wilson P, Tripathy D. Oncogenes and Tumor Suppressor Genes in Breast Cancer: Potential Diagnostic and Therapeutic Applications. Oncologist. 2004;9:361–77.

73. Zhu K, Liu Q, Zhou Y, Tao C, Zhao Z, Sun J, et al. Oncogenes and tumor suppressor genes: comparative genomics and network perspectives. BMC Genomics. 2015;16:S8.

74. Bowden GT, Schneider B, Domann R, Kulesz-Martin M. Oncogene activation and tumor suppressor gene inactivation during multistage mouse skin carcinogenesis. Cancer Res. 1994.

75. Jia P, Zhao Z. Characterization of Tumor-Suppressor Gene Inactivation Events in 33 Cancer Types. Cell Rep. 2019.

76. Carter H, Chen S, Isik L, Tyekucheva S, Velculescu VE, Kinzler KW, et al. Cancer-Specific High-Throughput Annotation of Somatic Mutations: Computational Prediction of Driver Missense Mutations. Cancer Res. 2009;69:6660–7.

77. Davoli T, Xu AW, Mengwasser KE, Sack LM, Yoon JC, Park PJ, et al. Cumulative Haploinsufficiency and Triplosensitivity Drive Aneuploidy Patterns and Shape the Cancer Genome. Cell. 2013;155:948–62.

78. Holland-Frei Cancer Medicine. Holland-Frei Cancer Med. 2017.

79. Botezatu A, Iancu I V., Popa O, Plesa A, Manda D, Huica I, et al. Mechanisms of Oncogene Activation. New Asp Mol Cell Mech Hum Carcinog. InTech; 2016.

80. Kumar RD, Searleman AC, Swamidass SJ, Griffith OL, Bose R. Statistically identifying tumor suppressors and oncogenes from pan-cancer genome-sequencing data. Bioinformatics. 2015;btv430.

81. Zhao M, Kim P, Mitra R, Zhao J, Zhao Z. TSGene 2.0: an updated literature-based knowledgebase for tumor suppressor genes. Nucleic Acids Res. 2016;44:D1023–31.

82. Sun W, Yang J. Functional Mechanisms for Human Tumor Suppressors. J Cancer. 2010;136–40.

83. Chernoff J. The two-hit theory hits 50. Welch M, editor. Mol Biol Cell. 2021;32.

84. Knudson AG. Mutation and Cancer: Statistical Study of Retinoblastoma. Proc Natl Acad Sci. 1971;68:820–3.

85. Morrill SA, Amon A. Why haploinsufficiency persists. Proc Natl Acad Sci. 2019;116:11866–71.

86. El Tekle G, Bernasocchi T, Unni AM, Bertoni F, Rossi D, Rubin MA, et al. Co-occurrence and mutual exclusivity: what cross-cancer mutation patterns can tell us. Trends in Cancer. 2021;7:823–36.

87. Chakravarty D, Gao J, Phillips S, Kundra R, Zhang H, Wang J, et al. OncoKB: A Precision Oncology Knowledge Base. JCO Precis Oncol. 2017;1–16.

88. Olafsson S, Anderson CA. Somatic mutations provide important and unique insights into the biology of complex diseases. Trends Genet. 2021;37:872–81.

89. Cairns J. Mutation selection and the natural history of cancer. Nature. 1975;255:197–200.

90. Nowell PC. The Clonal Evolution of Tumor Cell Populations. Science (80-). 1976;194:23–8.

91. Roizen MF. Hallmarks of Cancer: The Next Generation. Yearb Anesthesiol Pain Manag. 2012;2012:13.

92. Pérez-Pérez JM, Candela H, Micol JL. Understanding synergy in genetic interactions. Trends Genet. 2009;25:368–76.

93. Lehner B. Molecular mechanisms of epistasis within and between genes. Trends Genet. 2011;27:323–31.

94. Domingo J, Baeza-Centurion P, Lehner B. The Causes and Consequences of Genetic Interactions (Epistasis). Annu Rev Genomics Hum Genet. 2019;20:433–60.

95. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, et al. Universal Patterns of Selection in Cancer and Somatic Tissues. Cell. 2017;171:1029-1041.e21.

96. Network CGA. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012;487:330–7.

97. Ahronian LG, Sennott EM, Van Allen EM, Wagle N, Kwak EL, Faris JE, et al. Clinical Acquired Resistance to RAF Inhibitor Combinations in BRAF-Mutant Colorectal Cancer through MAPK Pathway Alterations. Cancer Discov. 2015;5:358–67.

98. Rajagopalan H, Bardelli A, Lengauer C, Kinzler KW, Vogelstein B, Velculescu VE. RAF/RAS oncogenes and mismatch-repair status. Nature. 2002;418:934–934.
99. Zeng S, Shen W, Liu L. Senescence and Cancer. Cancer Transl Med. 2018;4:70.

100. Wang B, Kohli J, Demaria M. Senescent Cells in Cancer Therapy: Friends or Foes? Trends in Cancer. 2020.

101. Liu X, Ding J, Meng L. Oncogene-induced senescence: a double edged sword in

cancer. Acta Pharmacol Sin. 2018;39:1553–8.

102. Zhu H, Blake S, Kusuma FK, Pearson RB, Kang J, Chan KT. Oncogene-induced senescence: From biology to therapy. Mech Ageing Dev. 2020;187:111229.

103. Jung SH, Hwang HJ, Kang D, Park HA, Lee HC, Jeong D, et al. mTOR kinase leads to PTEN-loss-induced cellular senescence by phosphorylating p53. Oncogene. 2019.

104. Chan KT, Blake S, Zhu H, Kang J, Trigos AS, Madhamshettiwar PB, et al. A functional genetic screen defines the AKT-induced senescence signaling network. Cell Death Differ. 2020;27:725–41.

105. Tu Z, Aird KM, Zhang R. RAS, cellular senescence and transformation. Small GTPases. 2012;3:163–7.

106. Cisowski J, Bergo MO. What makes oncogenes mutually exclusive? Small GTPases. 2017.

107. Deng Y, Luo S, Deng C, Luo T, Yin W, Zhang H, et al. Identifying mutual exclusivity across cancer genomes: computational approaches to discover genetic interaction and reveal tumor vulnerability. Brief Bioinform. 2019;20:254–66.

108. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. Cell. 2018;173:400-416.e11.

109. Hua X, Hyland PL, Huang J, Song L, Zhu B, Caporaso NE, et al. MEGSA: A Powerful and Flexible Framework for Analyzing Mutual Exclusivity of Tumor Mutations. Am J Hum Genet. 2016;98:442–55.

110. Ciriello G, Cerami E, Aksoy BA, Sander C, Schultz N. Using MEMo to discover mutual exclusivity modules in cancer. Curr Protoc Bioinforma. 2013.

111. Zhang J, Wu L-Y, Zhang X-S, Zhang S. Discovery of co-occurring driver pathways

in cancer. BMC Bioinformatics. 2014;15:271.

112. Mina M, Raynaud F, Tavernari D, Battistello E, Sungalee S, Saghafinia S, et al. Conditional Selection of Genomic Alterations Dictates Cancer Evolution and Oncogenic Dependencies. Cancer Cell. 2017;32:155-168.e6.

113. Kim. MEMCover: integrated analysis of mutual exclusivity and functional network reveals dysregulated pathways across multiple cancer types. Bioinformatics. 2015.

114. Jubb HC, Pandurangan AP, Turner MA, Ochoa-Montaño B, Blundell TL, Ascher DB. Mutations at protein-protein interfaces: Small changes over big surfaces have large impacts on human health. Prog. Biophys. Mol. Biol. 2017.

115. Cheng F, Liang H, Butte AJ, Eng C, Nussinov R. Personal Mutanomes Meet Modern Oncology Drug Discovery and Precision Health. Gottesman MM, editor. Pharmacol Rev. 2019;71:1–19.

116. Ivanov AA, Revennaugh B, Rusnak L, Gonzalez-Pecchi V, Mo X, Johns MA, et al. The OncoPPi Portal: an integrative resource to explore and prioritize protein–protein interactions for cancer target discovery. Berger B, editor. Bioinformatics. 2018;34:1183–91.

117. Ivanov AA. Explore protein–protein interactions for cancer target discovery using the oncoPPi portal. Methods Mol Biol. 2020.

118. Chen C, Shi C, Huang X, Zheng J, Zhu Z, Li Q, et al. Molecular Profiles and Metastasis Markers in Chinese Patients with Gastric Carcinoma. Sci Rep. 2019;9:13995.

119. Zhang Y, Chen Y, Yang C, Seger N, Hesla AC, Tsagkozis P, et al. TERT promoter mutation is an objective clinical marker for disease progression in chondrosarcoma. Mod Pathol. 2021;34:2020–7.

120. Zhou Y, Zhao J, Fang J, Martin W, Li L, Nussinov R, et al. My personal mutanome: a computational genomic medicine platform for searching network perturbing alleles linking genotype to phenotype. Genome Biol. 2021;22:53.

121. Chiang H-L, Liu C-J, Hu Y-W, Chen S-C, Hu L-Y, Shen C-C, et al. Risk of cancer in children, adolescents, and young adults with autistic disorder. J Pediatr. 2015;166:418-23.e1.

122. Liu Q, Yin W, Meijsen JJ, Reichenberg A, Gådin JR, Schork AJ, et al. Cancer risk in individuals with autism spectrum disorder. Ann Oncol. 2022;33:713–9.

123. Kao H-T, Buka SL, Kelsey KT, Gruber DF, Porton B. The correlation between rates of cancer and autism: an exploratory ecological investigation. PLoS One. 2010;5:e9372.

124. Chen M-H, Tsai S-J, Su T-P, Li C-T, Lin W-C, Cheng C-M, et al. Cancer risk in patients with bipolar disorder and unaffected siblings of such patients: A nationwide population-based study. Int J Cancer. 2022;150:1579–86.

125. Singh G, Driever PH, Sander JW. Cancer risk in people with epilepsy: the role of antiepileptic drugs. Brain. 2005;128:7–17.

126. Singh G, Fletcher O, Bell GS, McLean AE, Sander JW. Cancer mortality amongst people with epilepsy: a study of two cohorts with severe and presumed milder epilepsy. Epilepsy Res. 2009;83:190–7.

127. Pareja F, Ptashkin RN, Brown DN, Derakhshan F, Selenica P, da Silva EM, et al. Cancer-Causative Mutations Occurring in Early Embryogenesis. Cancer Discov. 2022;12:949–57.

128. Brunet T, Jech R, Brugger M, Kovacs R, Alhaddad B, Leszinski G, et al. De novo variants in neurodevelopmental disorders-experiences from a tertiary care center. Clin Genet. 2021;100:14–28.

129. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature. 2014;515:216–21.

130. Study DDD. Prevalence and architecture of de novo mutations in developmental disorders. Nature. 2017;542:433–8.

131. Chau KK, Zhang P, Urresti J, Amar M, Pramod AB, Chen J, et al. Full-length isoform transcriptome of the developing human brain provides further insights into autism. Cell Rep. 2021;36:109631.

132. Nussinov R, Tsai C-J, Jang H. Anticancer drug resistance: An update and perspective. Drug Resist Updat Rev Comment Antimicrob Anticancer Chemother. 2021;59:100796.

133. Cummings K, Watkins A, Jones C, Dias R, Welham A. Behavioural and psychological features of PTEN mutations: a systematic review of the literature and meta-analysis of the prevalence of autism spectrum disorder characteristics. J Neurodev Disord. 2022;14:1.

134. Frazier TW, Jaini R, Busch RM, Wolf M, Sadler T, Klaas P, et al. Cross-level analysis of molecular and neurobehavioral function in a prospective series of patients with germline heterozygous PTEN mutations with and without autism. Mol. Autism. 2021.

135. Orrico A, Galli L, Buoni S, Orsi A, Vonella G, Sorrentino V. Novel PTEN mutations in neurodevelopmental disorders and macrocephaly. Clin Genet. 2009;75:195–8.

136. Skelton PD, Stan R V, Luikart BW. The Role of PTEN in Neurodevelopment. Mol Neuropsychiatry. 2020;5:60–71.

137. Frazier TW. Autism Spectrum Disorder Associated with Germline Heterozygous Mutations. Cold Spring Harb Perspect Med. 2019;9.

138. Niarchou M, Chawner SJRA, Doherty JL, Maillard AM, Jacquemont S, Chung WK, et al. Psychiatric disorders in children with 16p11.2 deletion and duplication. Transl Psychiatry. 2019;9:8.

139. Walsh KM, Bracken MB. Copy number variation in the dosage-sensitive 16p11.2 interval accounts for only a small proportion of autism incidence: a systematic review and meta-analysis. Genet Med. 2011;13:377–84.

140. Gross AM, Frone M, Gripp KW, Gelb BD, Schoyer L, Schill L, et al. Advancing RAS/RASopathy therapies: An NCI-sponsored intramural and extramural collaboration for the study of RASopathies. Am. J. Med. Genet. Part A. 2020. p. 866–76.

141. Simanshu DK, Nissley D V, McCormick F. RAS Proteins and Their Regulators in Human Disease. Cell. 2017. p. 17–33.

142. Hebron KE, Hernandez ER, Yohe ME. The RASopathies: from pathogenetics to therapeutics. Dis Model Mech. 2022;15.

143. Dunnett-Kane V, Burkitt-Wright E, Blackhall FH, Malliri A, Evans DG, Lindsay CR. Germline and sporadic cancers driven by the RAS pathway: parallels and contrasts. Ann Oncol. 2020;31:873–83.

144. Andelfinger G, Marquis C, Raboisson M-J, Théoret Y, Waldmüller S, Wiegand G, et al. Hypertrophic Cardiomyopathy in Noonan Syndrome Treated by MEK-Inhibition. J Am Coll Cardiol. 2019;73:2237–9.

145. Cox AD, Der CJ, Philips MR. Targeting RAS Membrane Association: Back to the Future for Anti-RAS Drug Discovery? Clin. Cancer Res. 2015. p. 1819–27.

146. Dombi E, Baldwin A, Marcus LJ, Fisher MJ, Weiss B, Kim A, et al. Activity of Selumetinib in Neurofibromatosis Type 1-Related Plexiform Neurofibromas. N Engl J Med. 2016;375:2550–60.

147. Muyas F, Zapata L, Guigó R, Ossowski S. The rate and spectrum of mosaic mutations during embryogenesis revealed by RNA sequencing of 49 tissues. Genome Med. 2020.

148. Jansen AML, Goel A. Mosaicism in Patients With Colorectal Cancer or Polyposis Syndromes: A Systematic Review. Clin. Gastroenterol. Hepatol. 2020.

149. Ye Y, Cho MT, Retterer K, Alexander N, Ben-Omran T, Al-Mureikhi M, et al. De novo POGZ mutations are associated with neurodevelopmental disorders and

microcephaly . Mol Case Stud. 2015.

150. Matsumura K, Seiriki K, Okada S, Nagase M, Ayabe S, Yamada I, et al. Pathogenic POGZ mutation causes impaired cortical development and reversible autism-like phenotypes. Nat Commun. 2020.

151. Yaeger R, Corcoran RB. Targeting Alterations in the RAF-MEK Pathway. Cancer Discov. 2019;9:329–41.

152. Borrie SC, Brems H, Legius E, Bagni C. Cognitive Dysfunctions in Intellectual Disabilities: The Contributions of the Ras-MAPK and PI3K-AKT-mTOR Pathways. Annu Rev Genomics Hum Genet. 2017;18:115–42.

153. Parenti I, Rabaneda LG, Schoen H, Novarino G. Neurodevelopmental Disorders: From Genetics to Functional Pathways. Trends Neurosci. 2020;43:608–21.

154. Forés-Martos J, Catalá-López F, Sánchez-Valle J, Ibáñez K, Tejero H, Palma-Gudiel H, et al. Transcriptomic metaanalyses of autistic brains reveals shared gene expression and biological pathway abnormalities with cancer. Mol Autism. 2019;10:17.

155. Turner TN, Yi Q, Krumm N, Huddleston J, Hoekzema K, F Stessman HA, et al. denovo-db: a compendium of human de novo variants. Nucleic Acids Res. 2017;45:D804–11.

156. Bragin E, Chatzimichali EA, Wright CF, Hurles ME, Firth H V, Bevan AP, et al. DECIPHER: database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. Nucleic Acids Res. 2014;42:D993–1000.

157. Abrahams BS, Arking DE, Campbell DB, Mefford HC, Morrow EM, Weiss LA, et al. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). Mol Autism. BioMed Central; 2013;4:1–3.

158. Piñero J, Queralt-Rosinach N, Bravo À, Deu-Pons J, Bauer-Mehren A, Baron M, et al. DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. Database . 2015;2015:bav028.

159. Yavuz BR, Tsai CJ, Nussinov R, Tuncbag N. Pan-cancer clinical impact of latent drivers from double mutations. Commun Biol. 2023.

160. Islam SMA, Wu Y, Díaz-Gay M, Bergstrom EN, He Y, Barnes M, et al. Uncovering novel mutational signatures by de novo extraction with SigProfilerExtractor. bioRxiv. 2020.

161. van der Meer D, Barthorpe S, Yang W, Lightfoot H, Hall C, Gilbert J, et al. Cell Model Passports—a hub for clinical, genetic and functional datasets of preclinical cancer models. Nucleic Acids Res. 2019;47:D923–9.

162. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. 2012;41:D955–61.

163. Gao H, Korn JM, Ferretti S, Monahan JE, Wang Y, Singh M, et al. High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. Nat Med. 2015;21:1318–25.

164. Chen Z, Feng J, Saldivar J-S, Gu D, Bockholt A, Sommer SS. EGFR somatic doublets in lung cancer are frequent and generally arise from a pair of driver mutations uncommonly seen as singlet mutations: one-third of doublets occur at five pairs of amino acids. Oncogene. 2008;27:4336–43.

165. Gao J, Chang MT, Johnsen HC, Gao SP, Sylvester BE, Sumer SO, et al. 3D clusters of somatic mutations in cancer reveal numerous rare mutations as functional targets. Genome Med. 2017;9:4.

166. Yao Z, Yaeger R, Rodrik-Outmezguine VS, Tao A, Torres NM, Chang MT, et al. Tumours with class 3 BRAF mutants are sensitive to the inhibition of activated RAS. Nature. 2017;548:234–8.

167. Zhang M, Jang H, Nussinov R. The mechanism of PI3Kα activation at the atomic level. Chem Sci. 2019;10:3671–80.

168. Huang CH, Mandelker D, Schmidt-Kittler O, Samuels Y, Velculescu VE, Kinzler KW, et al. The structure of a human p110alpha/p85alpha complex elucidates the effects of oncogenic PI3Kalpha mutations. Science (80-). 2007.

169. Burke JE, Perisic O, Masson GR, Vadas O, Williams RL. Oncogenic mutations mimic and enhance dynamic events in the natural activation of phosphoinositide 3-kinase p110α (PIK3CA). Proc Natl Acad Sci. 2012;109:15259–64.

170. Zhang M, Jang H, Nussinov R. Structural Features that Distinguish Inactive and Active PI3K Lipid Kinases. J Mol Biol. 2020;432:5849–59.

171. Rodrigues CHM, Pires DEV, Ascher DB. DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. Nucleic Acids Res. 2018;46:W350–5.

172. Mackay TFC. Epistasis and quantitative traits: using model organisms to study gene–gene interactions. Nat Rev Genet. 2014;15:22–33.

173. Liu X, Lu S, Song K, Shen Q, Ni D, Li Q, et al. Unraveling allosteric landscapes of allosterome with ASD. Nucleic Acids Res. 2020;48:D394–401.

174. Nussinov R, Zhang M, Tsai C-J, Jang H. Phosphorylation and Driver Mutations in PI3Kα and PTEN Autoinhibition. Mol Cancer Res. 2021;19:543–8.

175. Suda K, Onozato R, Yatabe Y, Mitsudomi T. EGFR T790M Mutation: A Double Role in Lung Cancer Cell Survival? J Thorac Oncol. 2009;4:1–4.

176. Wang S, Cang S, Liu D. Third-generation inhibitors targeting EGFR T790M mutation in advanced non-small cell lung cancer. J Hematol Oncol. 2016;9:34.

177. Yan F, Liu X, Zhang S, Su J, Zhang Q, Chen J. Effect of double mutations T790M/L858R on conformation and drug-resistant mechanism of epidermal growth factor receptor explored by molecular dynamics simulations. RSC Adv. Royal Society of Chemistry; 2018;8:39797–810.

178. Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, et al. Oncogenic Signaling Pathways in The Cancer Genome Atlas. Cell. 2018.

179. Kishore J, Goel M, Khanna P. Understanding survival analysis: Kaplan-Meier estimate. Int J Ayurveda Res. 2010;1:274.

180. Han J, Pei J, Yin Y. Mining frequent patterns without candidate generation. ACM SIGMOD Rec. 2000;29:1–12.

181. Langville AN, Meyer CD. A Survey of Eigenvector Methods for Web Information Retrieval. SIAM Rev. 2005;47:135–61.

182. Hagberg AA, Schult DA, Swart PJ. Exploring network structure, dynamics, and function using NetworkX. 7th Python Sci Conf (SciPy 2008). 2008.

183. Su M-G, Weng JT-Y, Hsu JB-K, Huang K-Y, Chi Y-H, Lee T-Y. Investigation and identification of functional post-translational modification sites associated with drug binding and protein-protein interactions. BMC Syst Biol. 2017;11:132.

184. Duan G, Walther D. The Roles of Post-translational Modifications in the Context of Protein Interaction Networks. Radivojac P, editor. PLOS Comput Biol. 2015;11:e1004049.

185. Meyer MJ, Beltrán JF, Liang S, Fragoza R, Rumack A, Liang J, et al. Interactome INSIDER: a structural interactome browser for genomic studies. Nat Methods. 2018;15:107–14.

186. Kuleshov M V., Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016;44:W90–7.

187. Alon U. Network motifs: Theory and experimental approaches. Nat. Rev. Genet. 2007.

188. Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, et al. Oncogenic Signaling Pathways in The Cancer Genome Atlas. Cell. 2018;173:321-337.e10.

189. Yavuz BR, Tsai C-J, Nussinov R, Tuncbag N. Discovery of Latent Drivers from Double Mutations in Pan-Cancer Data Reveal their Clinical Impact. bioRxiv. 2021;2021.04.02.438239.

190. Theodorou V, Stark R, Menon S, Carroll JS. GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility. Genome Res. 2013;23:12–22.

191. Priestley P, Baber J, Lolkema MP, Steeghs N, de Bruijn E, Shale C, et al. Pan-cancer whole-genome analyses of metastatic solid tumours. Nature. 2019;575:210–6.

192. Shi Y, Hata A, Lo RS, Massagué J, Pavletich NP. A structural basis for mutational inactivation of the tumour suppressor Smad4. Nature. 1997;388:87–93.

193. Sporn MB, Liby KT. NRF2 The good the bad and the importance of the context. Nat Rev Cancer. 2013.

194. Jaramillo MC, Zhang DD. The emerging role of the Nrf2–Keap1 signaling pathway in cancer. Genes Dev. 2013;27:2179–91.

195. Tao S, Wang S, Moghaddam SJ, Ooi A, Chapman E, Wong PK, et al. Oncogenic KRAS Confers Chemoresistance by Upregulating NRF2. Cancer Res. 2014;74:7430–41.

196. Ciriello G, Cerami E, Sander C, Schultz N. Mutual exclusivity analysis identifies oncogenic network modules. Genome Res. 2012;22:398–406.

197. Dunlap JB, Leonard J, Rosenberg M, Cook R, Press R, Fan G, et al. The combination of NPM1, DNMT3A, and IDH1/2 mutations leads to inferior overall survival in AML. Am J Hematol. 2019;94:913–20.

198. Matsumoto A, Shimada Y, Nakano M, Oyanagi H, Tajima Y, Nakano M, et al. RNF43 mutation is associated with aggressive tumor biology along with BRAF V600E mutation in right-sided colorectal cancer. Oncol Rep. 2020.

199. De P, Dey N. Mutation-driven signals of ARID1A and PI3K pathways in ovarian carcinomas: Alteration is an opportunity. Int. J. Mol. Sci. 2019.

200. Liu C, Han Z, Zhang Z-K, Nussinov R, Cheng F. A network-based deep learning methodology for stratification of tumor mutations. Bioinformatics. 2021.

201. Stephenson JD, Laskowski RA, Nightingale A, Hurles ME, Thornton JM. VarMap: a web tool for mapping genomic coordinates to protein sequence and structure and retrieving protein structural annotations. Bioinformatics. 2019;35:4854–6.

202. Zhou X, Edmonson MN, Wilkinson MR, Patel A, Wu G, Liu Y, et al. Exploring genomic alteration in pediatric cancer using ProteinPaint. Nat Genet. 2016;48:4–6.

203. Lee JO, Yang H, Georgescu MM, Di Cristofano A, Maehama T, Shi Y, et al. Crystal structure of the PTEN tumor suppressor: implications for its phosphoinositide phosphatase activity and membrane association. Cell. 1999;99:323–34.

204. Miller MS, Schmidt-Kittler O, Bolduc DM, Brower ET, Chaves-Moreira D, Allaire M, et al. Structural basis of nSH2 regulation and lipid binding in PI3Kα. Oncotarget. 2014. p. 5198–208.

205. Delano WL. The PyMOL Molecular Graphics System. CCP4 Newsl protein Crystallogr. 2002.

206. Qi H, Dong C, Chung WK, Wang K, Shen Y. Deep Genetic Connection Between Cancer and Developmental Disorders. Hum Mutat. 2016;37:1042–50.

207. Jiang C-C, Lin L-S, Long S, Ke X-Y, Fukunaga K, Lu Y-M, et al. Signalling pathways in autism spectrum disorder: mechanisms and therapeutic implications. Signal Transduct Target Ther. 2022;7:229.

208. Stratton M. Patterns of somatic mutation in human cancer genomes. Eur. J. Cancer Suppl. 2008. p. 6.

209. Huang K-L, Mashl RJ, Wu Y, Ritter DI, Wang J, Oh C, et al. Pathogenic Germline Variants in 10,389 Adult Cancers. Cell. 2018;173:355-370.e14.

210. Xu X, Zhou Y, Feng X, Li X, Asad M, Li D, et al. Germline genomic patterns are associated with cancer risk, oncogenic pathways, and clinical outcomes. Sci Adv. 2020;6.

211. Qing T, Mohsen H, Marczyk M, Ye Y, O'Meara T, Zhao H, et al. Germline variant burden in cancer genes correlates with age at diagnosis and somatic mutation burden. Nat Commun. 2020;11:2438.

212. Rashed WM, Marcotte EL, Spector LG. Germline Mutations as a Cause of Childhood Cancer. JCO Precis Oncol. 2022;6:e2100505.

213. Liu M, Liu X, Suo P, Gong Y, Qu B, Peng X, et al. The contribution of hereditary cancer-related germline mutations to lung cancer susceptibility. Transl Lung Cancer Res. 2020;9:646–58.

214. Li B, Li K, Tian D, Zhou Q, Xie Y, Fang Z, et al. De novo mutation of cancer-related genes associates with particular neurodevelopmental disorders. J Mol Med . 2020;98:1701–12.

215. Koire A, Katsonis P, Kim YW, Buchovecky C, Wilson SJ, Lichtarge O. A method to delineate de novo missense variants across pathways prioritizes genes linked to autism. Sci Transl Med. 2021;13.

216. Nussinov R, Tsai C-J, Jang H. Allostery, and how to define and measure signal transduction. Biophys Chem. 2022;283:106766.

217. Nussinov R, Tsai C-J, Jang H. A New View of Activating Mutations in Cancer. Cancer Res. 2022;82:4114–23.

218. Pejaver V, Urresti J, Lugo-Martinez J, Pagel KA, Lin GN, Nam H-J, et al. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. Nat Commun. 2020;11:5918.

219. Jang H, Smith IN, Eng C, Nussinov R. The mechanism of full activation of tumor suppressor PTEN at the phosphoinositide-enriched membrane. iScience. 2021;24:102438.

220. Spinelli L, Black FM, Berg JN, Eickholt BJ, Leslie NR. Functionally distinct groups of inherited PTEN mutations in autism and tumour syndromes. J Med Genet. 2015;52:128–34.

221. Portelli S, Barr L, de Sá AGC, Pires DE V, Ascher DB. Distinguishing between PTEN clinical phenotypes through mutation analysis. Comput Struct Biotechnol J. 2021;19:3097–109.

222. Mighell TL, Evans-Dutson S, O'Roak BJ. A Saturation Mutagenesis Approach to Understanding PTEN Lipid Phosphatase Activity and Genotype-Phenotype Relationships. Am J Hum Genet. 2018;102:943–55.

223. Wong CW, Or PMY, Wang Y, Li L, Li J, Yan M, et al. Identification of a PTEN mutation with reduced protein stability, phosphatase activity, and nuclear localization in Hong Kong patients with autistic features, neurodevelopmental delays, and macrocephaly. Autism Res. 2018. p. 1098–109.

224. Nussinov R, Jang H, Tsai C-J, Cheng F. Review: Precision medicine and driver mutations: Computational methods, functional assays and conformational principles for interpreting cancer drivers. Panchenko ARR, editor. PLOS Comput. Biol. 2019. p. e1006658.

225. Guarnera E, Berezovsky IN. Allosteric drugs and mutations: chances, challenges, and necessity. Curr Opin Struct Biol. 2020;62:149–57.

226. Tan ZW, Guarnera E, Tee W-V, Berezovsky IN. AlloSigMA 2: paving the way to designing allosteric effectors and to exploring allosteric effects of mutations. Nucleic Acids Res. 2020;48:W116–24.

227. Tan ZW, Tee W-V, Guarnera E, Booth L, Berezovsky IN. AlloMAPS: allosteric mutation analysis and polymorphism of signaling database. Nucleic Acids Res. 2019;47:D265–70.

228. Teng M, Luskin MR, Cowan-Jacob SW, Ding Q, Fabbro D, Gray NS. The Dawn of Allosteric BCR-ABL1 Drugs: From a Phenotypic Screening Hit to an Approved Drug. J Med Chem. American Chemical Society; 2022;65:7581–94.

229. Nussinov R, Zhang M, Maloney R, Tsai C-J, Yavuz BR, Tuncbag N, et al. Mechanism of activation and the rewired network: New drug design concepts. Med Res Rev. United States; 2022;42:770–99.

230. Nussinov R, Tsai C-J, Jang H. Autoinhibition can identify rare driver mutations and advise pharmacology. FASEB J. United States; 2020;34:16–29.

231. Nussinov R, Zhang M, Tsai C-J, Liao T-J, Fushman D, Jang H. Autoinhibition in Ras effectors Raf, PI3Kα, and RASSF5: a comprehensive review underscoring the challenges in pharmacological intervention. Biophys Rev. 2018;10:1263–82.

232. Poulikakos PI, Persaud Y, Janakiraman M, Kong X, Ng C, Moriceau G, et al. RAF inhibitor resistance is mediated by dimerization of aberrantly spliced BRAF(V600E). Nature. 2011;480:387–90.

233. Yao Z, Torres NM, Tao A, Gao Y, Luo L, Li Q, et al. BRAF Mutants Evade ERK-Dependent Feedback by Different Mechanisms that Determine Their Sensitivity to Pharmacologic Inhibition. Cancer Cell. 2015;28:370–83.

234. Röring M, Herr R, Fiala GJ, Heilmann K, Braun S, Eisenhardt AE, et al. Distinct requirement for an intact dimer interface in wild-type, V600E and kinase-dead B-Raf signalling. EMBO J. John Wiley & Sons, Ltd; 2012;31:2629–47.

235. Freeman AK, Ritt DA, Morrison DK. Effects of Raf Dimerization and Its Inhibition on Normal and Disease-Associated Raf Signaling. Mol Cell. 2013;49:751–8.

236. Yuan TL, Cantley LC. PI3K pathway alterations in cancer: variations on a theme.

Oncogene. 2008;27:5497–510.

237. Thevakumaran N, Lavoie H, Critton DA, Tebben A, Marinier A, Sicheri F, et al. Crystal structure of a BRAF kinase domain monomer explains basis for allosteric regulation. Nat Struct Mol Biol. 2015;22:37–43.

238. Diedrich B, Rigbolt KT, Röring M, Herr R, Kaeser-Pebernard S, Gretzmeier C, et al. Discrete cytosolic macromolecular BRAF complexes exhibit distinct activities and composition. EMBO J. John Wiley & Sons, Ltd; 2017;36:646–63.

239. Nussinov R, Jang H, Tsai C-J. The structural basis for cancer treatment decisions. Oncotarget. 2014;5:7285–302.

240. Henkel L, Rauscher B, Boutros M. Context-dependent genetic interactions in cancer. Curr Opin Genet Dev. 2019.

241. Park S, Lehner B. Cancer type-dependent genetic interactions between cancer driver alterations indicate plasticity of epistasis across cell types. Mol Syst Biol. 2015;11:824.

242. Mina M, Iyer A, Ciriello G. Epistasis and evolutionary dependencies in human cancers. Curr Opin Genet Dev. Curr Opin Genet Dev; 2022;77:101989.

243. Zhu H, Wei M, Xu J, Hua J, Liang C, Meng Q, et al. PARP inhibitors in pancreatic cancer: molecular mechanisms and clinical applications. Mol Cancer. 2020;19:49.

244. Brown TJ, Reiss KA. PARP Inhibitors in Pancreatic Cancer. Cancer J. (United States). 2021.

245. Rauscher R, Bampi GB, Guevara-Ferrer M, Santos LA, Joshi D, Mark D, et al. Positive epistasis between disease-causing missense mutations and silent polymorphism with effect on mRNA translation velocity. Proc Natl Acad Sci. 2021;118.

246. O'Neil NJ, Bailey ML, Hieter P. Synthetic lethality and cancer. Nat Rev Genet. 2017;18:613–23.

247. Mateo L, Duran-Frigola M, Gris-Oliver A, Palafox M, Scaltriti M, Razavi P, et al. Personalized cancer therapy prioritization based on driver alteration co-occurrence patterns. Genome Med. 2020;12:78.

248. Zhao J, Luo Z. Discovery of Raf Family Is a Milestone in Deciphering the Ras-Mediated Intracellular Signaling Pathway. Int J Mol Sci. 2022;23.

249. Sahin M, Sur M. Genes, circuits, and precision therapies for autism and related neurodevelopmental disorders. Science (80-). 2015;350.

250. Longo F, Klann E. Reciprocal control of translation and transcription in autism spectrum disorder. EMBO Rep. 2021;22:e52110.

251. Kumar S, Reynolds K, Ji Y, Gu R, Rai S, Zhou CJ. Impaired neurodevelopmental pathways in autism spectrum disorder: a review of signaling mechanisms and crosstalk. J Neurodev Disord. 2019;11:10.

252. Chang C-J, Mulholland DJ, Valamehr B, Mosessian S, Sellers WR, Wu H. PTEN nuclear localization is regulated by oxidative stress and mediates p53-dependent tumor suppression. Mol Cell Biol. Molecular and Medical Pharmacology, UCLA School of Medicine, Los Angeles, CA 90095, USA.: Informa UK Limited; 2008;28:3281–9.

253. Papa A, Wan L, Bonora M, Salmena L, Song MS, Hobbs RM, et al. Cancer-Associated PTEN Mutants Act in a Dominant-Negative Manner to Suppress PTEN Protein Function. Cell. 2014. p. 595–610.

254. Rodríguez-Escudero I, Oliver MD, Andrés-Pons A, Molina M, Cid VJ, Pulido R. A comprehensive functional analysis of PTEN mutations: implications in tumor- and autism-related syndromes. Hum Mol Genet. 2011;20:4132–42.

255. Wang L, Zhou K, Fu Z, Yu D, Huang H, Zang X, et al. Brain Development and Akt Signaling: the Crossroads of Signaling Pathway and Neurodevelopmental Diseases. J Mol Neurosci. 2017;61:379–84.

256. Crunkhorn S. Genetic disorders: PI3K inhibitor reverses overgrowth syndrome. Nat

Rev Drug Discov. 2018;17:545.

257. Venot Q, Blanc T, Rabia SH, Berteloot L, Ladraa S, Duong J-P, et al. Targeted therapy in patients with PIK3CA-related overgrowth syndrome. Nature. 2018;558:540–6.

258. Mirzaa G, Timms AE, Conti V, Boyle EA, Girisha KM, Martin B, et al. PIK3CA-associated developmental disorders exhibit distinct classes of mutations with variable expression and tissue distribution. JCI Insight. 2016.

259. Dobyns WB, Mirzaa GM. Megalencephaly syndromes associated with mutations of core components of the PI3K-AKT-MTOR pathway: PIK3CA, PIK3R2, AKT3, and MTOR. Am J Med Genet C Semin Med Genet. 2019;181:582–90.

260. Yeung KS, Tso WWY, Ip JJK, Mak CCY, Leung GKC, Tsang MHY, et al. Identification of mutations in the PI3K-AKT-mTOR signalling pathway in patients with macrocephaly and developmental delay and/or autism. Mol Autism. BioMed Central; 2017;8:1–11.

261. Ersahin T, Tuncbag N, Cetin-Atalay R. The PI3K/AKT/mTOR interactive pathway. Mol Biosyst. 2015;11:1946–54.

262. Vanhaesebroeck B, Guillermet-Guibert J, Graupera M, Bilanges B. The emerging mechanisms of isoform-specific PI3K signalling. Nat Rev Mol Cell Biol. Nature Publishing Group; 2010;11:329–41.

263. Thorpe LM, Yuzugullu H, Zhao JJ. PI3K in cancer: divergent roles of isoforms, modes of activation and therapeutic targeting. Nat Rev Cancer. Nature Publishing Group; 2014;15:7–24.

264. Bi L, Okabe I, Bernard DJ, Wynshaw-Boris A, Nussbaum RL. Proliferative Defect and Embryonic Lethality in Mice Homozygous for a Deletion in the p110α Subunit of Phosphoinositide 3-Kinase *. J Biol Chem. Elsevier; 1999;274:10963–8.

265. Peng X-D, Xu P-Z, Chen M-L, Hahn-Windgassen A, Skeen J, Jacobs J, et al. Dwarfism, impaired skin development, skeletal muscle atrophy, delayed bone

development, and impeded adipogenesis in mice lacking Akt1 and Akt2. Genes Dev. Cold Spring Harbor Laboratory Press; 2003;17:1352–65.

266. Hall M. Faculty Opinions recommendation of mTOR is essential for growth and proliferation in early mouse embryos and embryonic stem cells. Fac. Opin. – Post-Publication Peer Rev. Biomed. Lit. 2004.

267. Yu JSL, Cui W. Proliferation, survival and metabolism: the role of PI3K/AKT/mTOR signalling in pluripotency and cell fate determination. Development. The Company of Biologists; 2016;143:3050–60.

268. Eastman AE, Chen X, Hu X, Hartman AA, Pearlman Morales AM, Yang C, et al. Resolving Cell Cycle Speed in One Snapshot with a Live-Cell Fluorescent Reporter. Cell Rep. 2020;31:107804.

269. DeGregori J, Leone G, Miron A, Jakoi L, Nevins JR. Distinct roles for E2F proteins in cell growth control and apoptosis. Proc Natl Acad Sci U S A. 1997;94:7245–50.

270. Tadesse S, Caldon EC, Tilley W, Wang S. Cyclin-Dependent Kinase 2 Inhibitors in Cancer Therapy: An Update. J Med Chem. 2019;62:4233–51.

271. Scheiblecker L, Kollmann K, Sexl V. CDK4/6 and MAPK-Crosstalk as Opportunity for Cancer Treatment. Pharm. 2020;13.

272. Hou P-S, hAilín DÓ, Vogel T, Hanashima C. Transcription and Beyond: Delineating FOXG1 Function in Cortical Development and Disorders. Front Cell Neurosci. 2020;14:35.

273. Li C, Ito H, Fujita K, Shiwaku H, Qi Y, Tagawa K, et al. Sox2 transcriptionally regulates PQBP1, an intellectual disability-microcephaly causative gene, in neural stem progenitor cells. PLoS One. 2013;8:e68627.

274. Nussinov R, Tsai C-J, Jang H, Korcsmáros T, Csermely P. Oncogenic KRAS signaling and YAP1/β-catenin: Similar cell cycle control in tumor initiation. Semin Cell Dev Biol. 2016;58:79–85.

275. Heslop P, Cook A, Sullivan B, Calkin R, Pollard J, Byrne V. Cancer in deceased adults with intellectual disabilities: English population-based study using linked data from three sources. BMJ Open. 2022;12:e056974.

276. Corda G, Sala A. Non-canonical WNT/PCP signalling in cancer: Fzd6 takes centre stage. Oncogenesis. 2017;6:e364.

277. Cuypers M, Schalk BWM, Boonman AJN, Naaldenberg J, Leusink GL. Cancer-related mortality among people with intellectual disabilities: A nationwide population-based cohort study. Cancer. 2022;128:1267–74.

278. Bielas JH, Heddle JA. Proliferation is necessary for both repair and mutation in transgenic mouse cells. Proc Natl Acad Sci U S A. 2000;97:11391–6.

279. Demeter M, Derényi I, Szöllősi GJ. Trade-off between reducing mutational accumulation and increasing commitment to differentiation determines tissue organization. Nat Commun. Nature Publishing Group; 2022;13:1–10.

280. Zengeler KE, Lukens JR. Innate immunity at the crossroads of healthy brain maturation and neurodevelopmental disorders. Nat Rev Immunol. 2021;21:454–68.

281. Clara JA, Monge C, Yang Y, Takebe N. Targeting signalling pathways and the immune microenvironment of cancer stem cells - a clinical update. Nat Rev Clin Oncol. 2020;17:204–32.

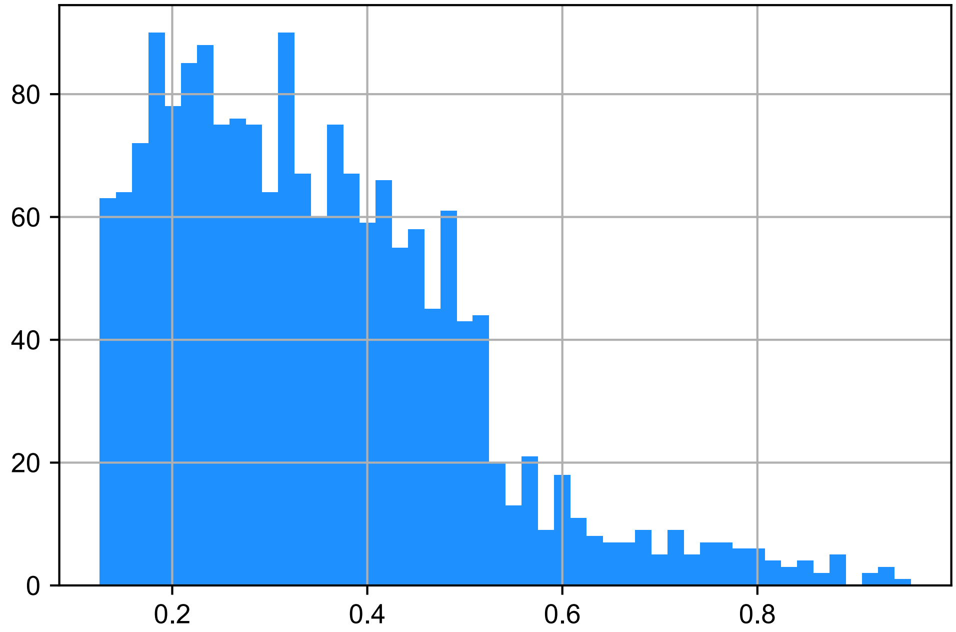
282. Pisibon C, Ouertani A, Bertolotto C, Ballotti R, Cheli Y. Immune Checkpoints in Cancers: From Signaling to the Clinic. Cancers . Multidisciplinary Digital Publishing Institute; 2021;13:4573.

283. Yavuz BR, Tsai C-J, Nussinov R, Tuncbağ N. Latent driver discovery-figure source data [Internet]. 2022. Available from: https://doi.org/10.6084/m9.figshare.21788192.v3

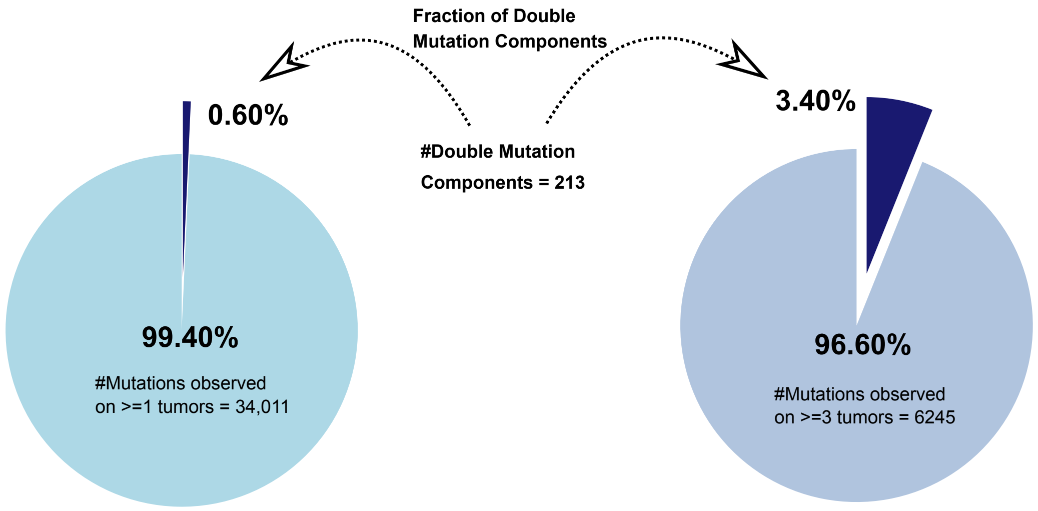
APPENDICES

APPENDIX A

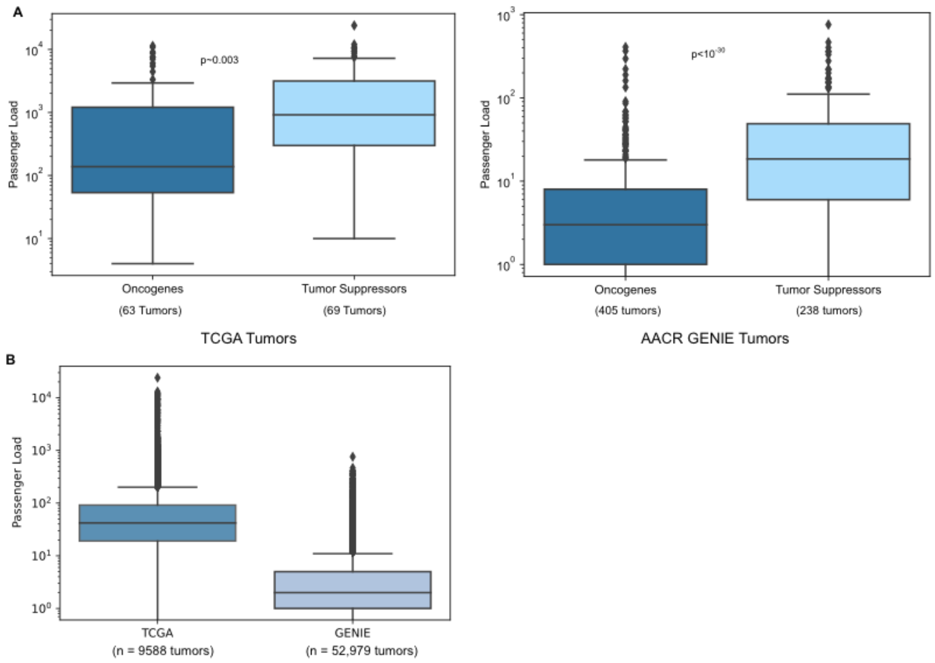
ADDITIONAL CHAPTERS OF FIGURE 3



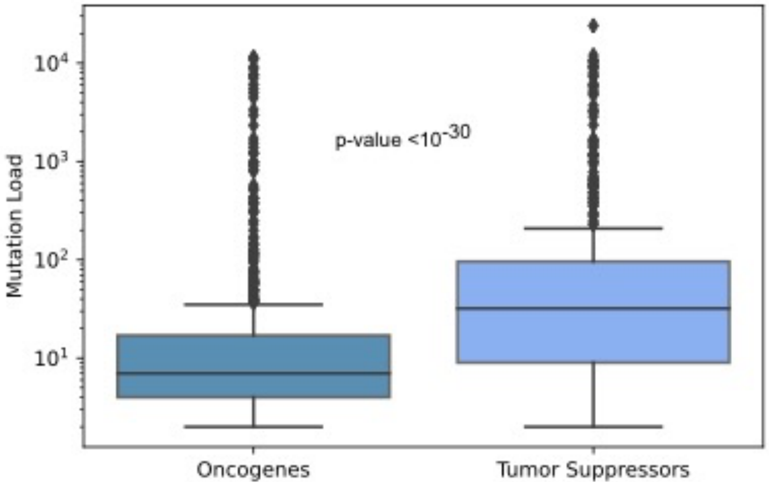
Appendix A - Figure 1: Histogram depicting VAF values of 213 double mutation constituents on 821 double mutant tumors. Double mutation constituents have VAF values accumulated between 0.2 and 0.4 on the double mutant tumors.



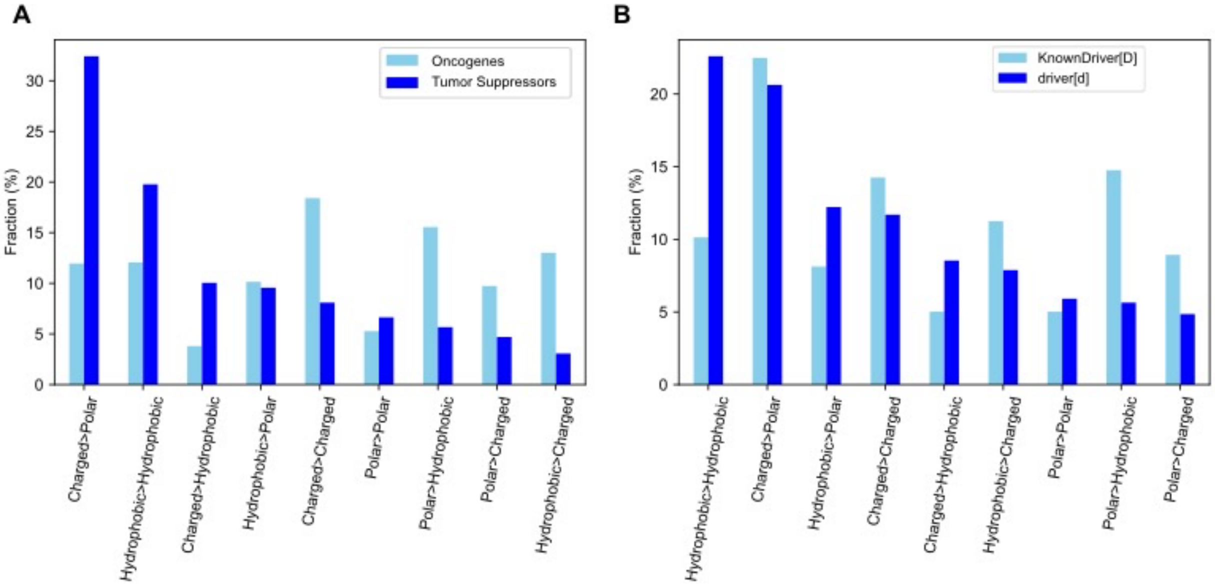
Appendix A - Figure 2: Fraction of 213 double mutation components among all mutations that are observed on at least one tumor (n=34,011)(on the left), and three tumors (n=6245)(on the right) on the 53 genes harboring at least one double mutation.



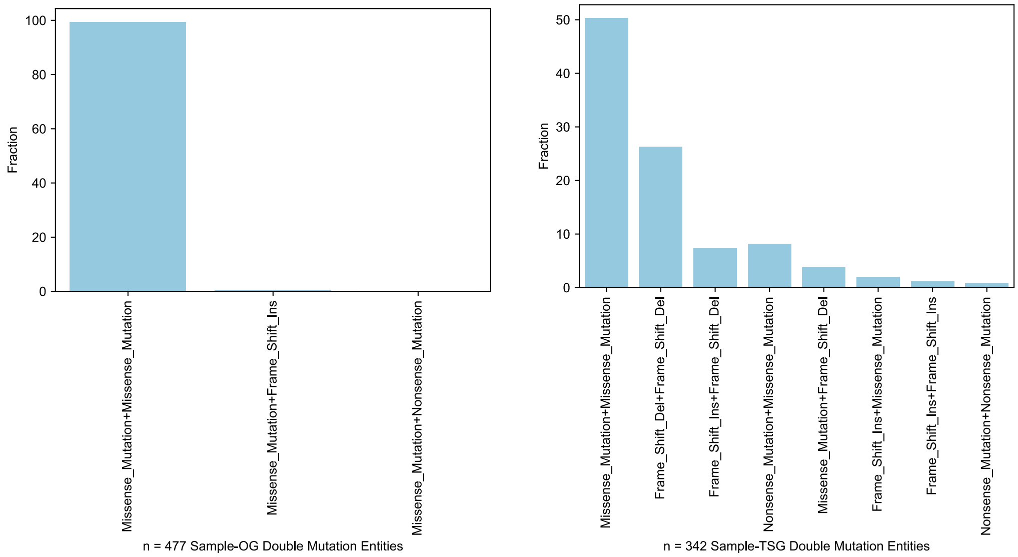
Appendix A - Figure 3: (A) Comparison of the passenger mutation loads of tumors from TCGA and GENIE data sets separately shows that the passenger load of the tumors carrying at least one tumor suppressor doublet is significantly higher than its counterpart in oncogenes. (B) Passenger mutation loads of the tumors among TCGA and GENIE data.



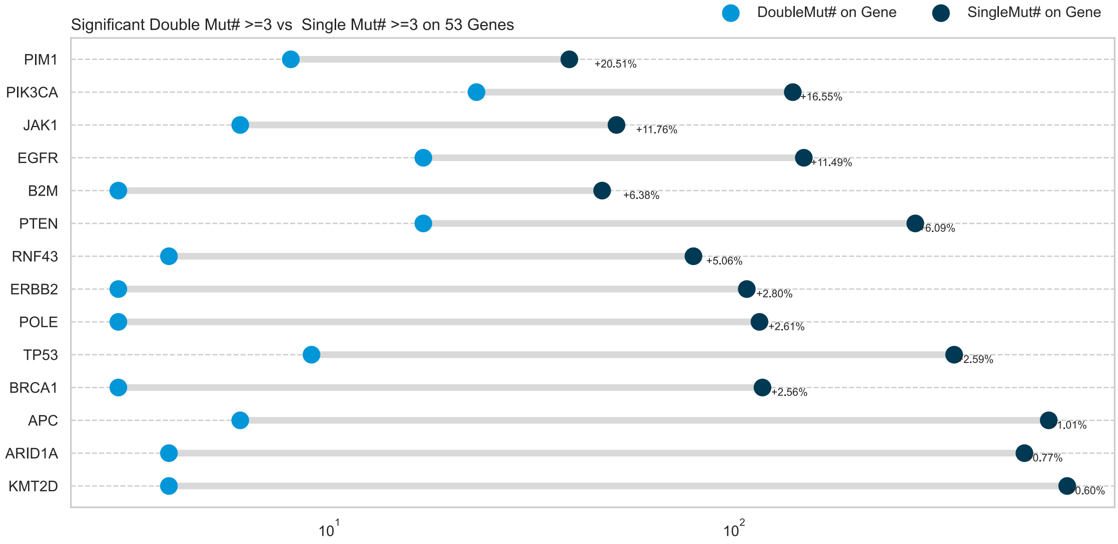
Appendix A - Figure 4: Mutation loads of tumors that that carry at least one significant double mutation on oncogenes (n=468) and tumor suppressor genes (n=307). There are 13 oncogenes and 25 tumor suppressor genes. Mutation loads of tumors with at least one double mutation on a TSG is significantly higher (Mann Whitney-U Test, $p<4x10^{-30}$).



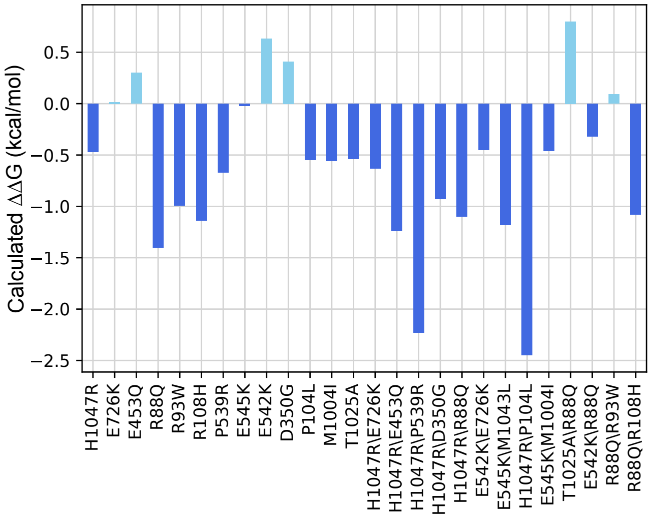
Appendix A - Figure 5: Analysis of the chemical class of (A) TSGs and OGs (B) known driver and latent driver mutations. Charged>Polar and Charged>Charged switches are more dominant among TSGs and OGs, Charged>Polar and Hydrophobic>Hydrophobic switches are more dominant among known driver and latent driver mutations, respectively.



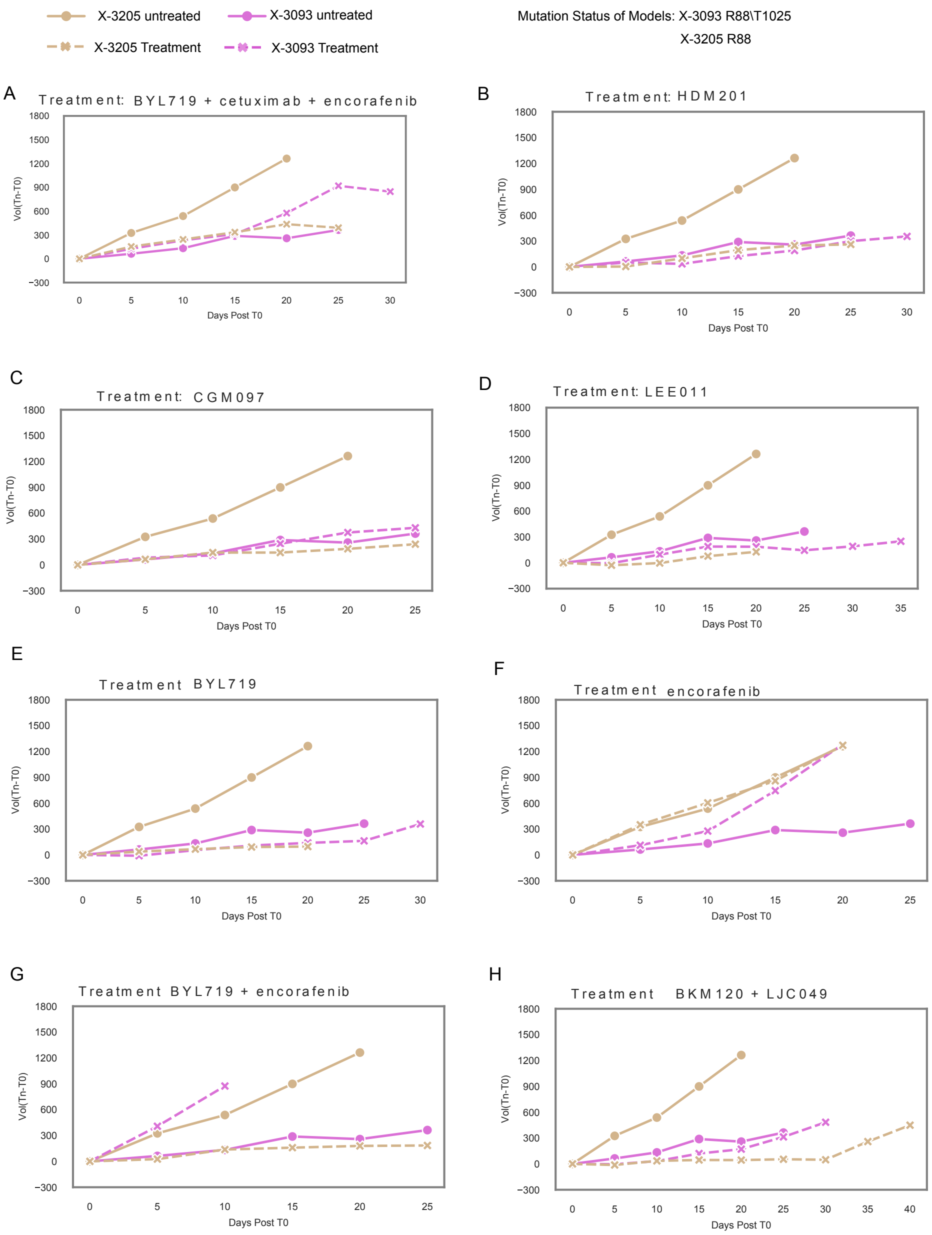
Appendix A - Figure 6: Combinations of missense mutations are highly prominent among double mutations on oncogenes; but for the doublets on tumor suppressor genes there are various combinations where missense+missense and frameshift_del +frameshift_del get the highest shares when the fractions (%) compared.



Appendix A - Figure 7: Number of double and single mutations on each gene that harbors at least one significant double mutation present in at least three double mutant tumors. All the single mutations observed on at least three tumors are included. Double mutation fraction among the single mutations are also noted.



Appendix A - Figure 8: Predicted ΔΔG values for single and double mutations of PIK3CA calculated with Dynamut web server (PDB id: 4OVV).



Appendix A - Figure 9. PIK3CA R88/T1025 mutant xenograft (X-3093, BRCA) volume change compared to single R88 mutant xenograft (X-3205, BRCA) for different drug treatments. x-axis shows treatment days, y-axis shows volume difference Volume(Day=n)-Volume(Day=0). Treatment with the drugs/drug combinations. (A) BYL719+cetuximab+encorafenib combination. (B) HDM201 (Siremadlin). (C) CGM097. (D)LEE011(Ribociclib) (E) BYL719 (Alpelisib) (F) Encorafenib (G) BYL719+Encorafenib (H) BKM120+LJC049

APPENDIX B

ADDITIONAL METHODS OF CHAPTER 3

Variant allele frequency of mutation constituents

We created a dataset of 2917 distinct tumor-mutation pairs and associated VAF values based on the VAF values of each double mutation component on 1308 double mutant tumors and 295 double mutation constituents. Using the first quartile (Q1=0.21), median (0.30), and third quartile (Q3=0.40) values, the data was divided into four groups, each covering 25% of the total data. The values in each group encompass the VAF values 0.125-0.21, 0.21-0.30, 0.30-0.40, and 0.40-0.95. Appendix A-Figure 1 shows the histogram of VAF values for the mutations forming doublets among the double mutant tumors.

Annotation of double mutations

To find out spatial closeness of same gene double mutations we use 3DHotspots[16] which identifies statistically significant mutations clustering in 3D protein structures. There are 943 clusters of 504 different genes. If two mutated residues that are containing a dual mutation belong to the same cluster, we consider this same gene dual mutation components are in close proximity. We used Interactome Insider to identify if the components of either same gene or different gene double are located in the same interface [185]. Besides the experimental data in PDB and predicted data in Interactome3D, it also contains the predicted interfaces with their in-house method. We used EnrichR to find the pathway annotation of the genes having co-occurring mutations[186] .

Alterations in chemical properties of amino acids

In order to classify alterations with respect to chemical classes of amino acids before and after mutations, we prepared a file containing unique rows as follows "patient barcode| gene | residue number | wild type amino acid | mutant amino acid". We excluded the cases where the final amino acid is a stop codon. The 9 categories we evaluated in our analysis are Polar-Hydrophobic, Charged-Polar, Hydrophobic-Hydrophobic, Hydrophobic-Polar, Hydrophobic-Charged, Polar-Charged, Polar-Polar, Charged-Hydrophobic, Charged-Charged.

In total 188 missense mutations from 863 samples were analyzed. Positions that are mutated to multiple amino acids are counted multiple times.

We collected wild type and mutant residue information for each double mutant tumor and the double mutation constituent it comprises, as well as the chemical classifications of these residues. Considering that some mutations contribute to more than one double mutation, we ensured that each tumor and mutation was recorded only once. We excluded the cases where the mutant residue is a stop codon. Among the double mutant tumors on TSGs and OGs, there were 795 and 1170 unique tumor and mutation records, respectively.

PIK3CA stability analysis via Dynamut tool

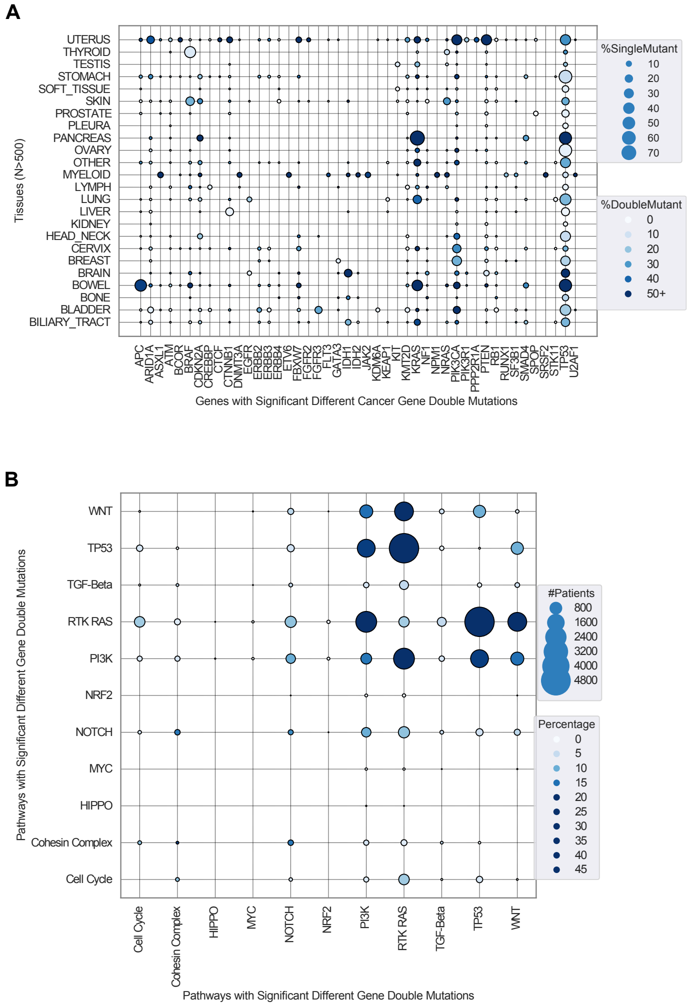
The mechanisms of activation of PI3Kα by some of the driver mutations have been recently worked out [16,40,170]. Unsurprisingly, considering their diverse mechanisms of action no clear trend is observed in the calculated folding free energy (ΔΔG) upon double or single mutation with DynaMut [171] (Appendix A-Figure 11). If the components of double mutations act via distinct mechanisms, the additivity of their activation potential is high; otherwise the additivity is low as in the E545/E542 example where the mutations execute the same mechanism of action. Using the inactive state (PDB id: 4OVV) we calculated the folding free energy (ΔΔG) upon mutation using DynaMut [171] to assess the impact of single and double mutations on PIK3CA stability. Unsurprisingly, considering their diverse mechanism of action no clear trend is observed (Appendix A-Figure 11). For example, H1047R is a strong driver that promotes interaction with the membrane. It destabilization impact is minor (ΔΔG ≈ -0.5 kcal/mol). The impact of weak drivers R88Q and R93W is somewhat stronger (ΔΔG ≈ -1.5 kcal/mol and ΔΔG ≈ -1 kcal/mol, respectively). The effect of allosteric mutation D539R is also minor (ΔΔG ≈ -0.6 kcal/mol). Another strong driver E542K (ΔΔG ≈ 0.7 kcal/mol), stabilizes the protein like the weak drivers D350G (ΔΔG ≈ 0.5 kcal/mol) and E453Q (ΔΔG ≈ 0.3 kcal/mol). The most prominent stability changes occur when the strong driver H1047R cooperates with the allosteric mutation P539R (ΔΔG ≈ -2.3 kcal/mol) and the minor mutation P104L (ΔΔG ≈ -2.5 kcal/mol). These two dual mutations H1047R/P539R and H1047R/P104L destabilize the protein as do T1025A/R88Q (ΔΔG ≈ 0.7 kcal/mol) while T1025A and R88Q have a destabilizing effect.

Mutational Signatures

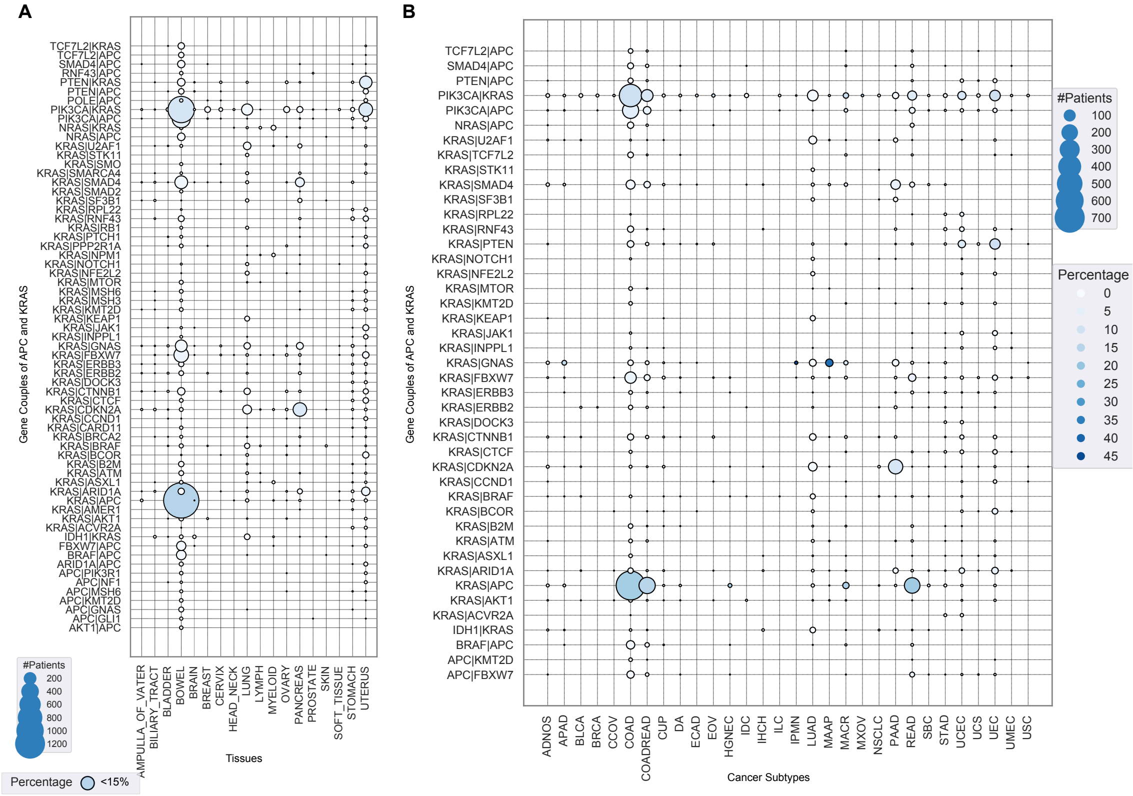
For example, E542/E726 and E542/E545 are of context T[G>A]A in 20 and 10 records. Similarly, the context T[G>A]A forms the doublets E545/E726 and E545/M1004 in 20 and 4 records, respectively. E542 and E545 are strong driver mutations, while E726 and M1004 are strong and weak latent driver mutations. C[A>G]T forms H1047/H1048 doublet in 5 records, and C[G>A]A forms R88/R357 doublet in 7 records. Further examples include S97/L25 doublet in PIM1 with the context G[C>G]T (3 records) and S431/R232 doublet in PTPRD with the context T[C>T]G (3 records) [159].

APPENDIX C

ADDITIONAL FIGURES OF CHAPTER 4



Appendix C - Figure 1: (A) Different gene dual mutation tendency. Node size represents the percentage of the Gene-mutant tumors and the node color represents percentage of different gene dual mutant tumors in the corresponding tissue where the gene on the x-axisis one of the constituents.**(B)** Different gene couples pathway prevalence.



Appendix C - Figure 2: (A) Presence of different gene dual mutations across different cancer tissues when one of the constituents is APC or KRAS. **(B)** Presence of different gene dual mutations across different cancer subtypes when one of the constituents is APC or KRAS.

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name	:	Yavuz, Bengi Ruken
Nationality	:	Turkish (TC)

EDUCATION

Degree	**Department**	**Institution**	**Year of Graduation**
MS	Mathematics	Bilkent Uni.	2013
BS	Mathematics	METU	2010

WORK EXPERIENCE

Enrollment	**Place**	**Years**
Guest Researcher	National Cancer Institute, USA	2019-2022
Teaching Assistant	Informatics Institute, METU	2018 – Present
Lecturer (part-time)	Bilkent University	2016– 2018
Research Assistant	Bilkent University	2010 – 2016

PUBLICATIONS

1. Yavuz, B.R., Arici, M.K., Demirel, H.C. et al. Neurodevelopmental disorders and cancer networks share pathways; but differ in mechanisms, signaling strength, and outcome (preprint). BioRxiv (2023).

2. Nussinov, R., Yavuz, B.R., Arici, M.K. et al. Neurodevelopmental disorders, like cancer, are connected to impaired chromatin remodelers, PI3K/mTOR, and PAK1-regulated MAPK. Biophys Rev 15, 163–181 (2023). https://doi.org/10.1007/s12551-023-01054-9.

3. Yavuz B.R., Tsai C.J., Nussinov R., Tuncbag N. Pan-cancer clinical impact of latent drivers from double mutations, Communications Biology 6, 202 (2023). https://doi.org/10.1038/s42003-023-04519-5.

4. Nussinov R., Zhang M., Maloney R., Tsai C.J., Yavuz B.R., Tuncbag N., Jang H. Targeting the mechanism of activation and the rewired network: New concepts in drug design, Medicinal Research Reviews (2021). DOI: 10.1002/med.21863.

THESIS PERMISSION FORM

ENSTİTÜ / INSTITUTE

Fen Bilimleri Enstitüsü / Graduate School of Natural and Applied Sciences ☐
Sosyal Bilimler Enstitüsü / Graduate School of Social Sciences ☐
Uygulamalı Matematik Enstitüsü / Graduate School of Applied Mathematics ☐
Enformatik Enstitüsü / Graduate School of Informatics ☒
Deniz Bilimleri Enstitüsü / Graduate School of Marine Sciences ☐

YAZARIN / AUTHOR

Soyadı / Surname : Yavuz
Adı / Name : Bengi Ruken
Bölümü / Department : Tıp Bilişimi

TEZİN ADI / TITLE OF THE THESIS (**İngilizce /** English) : A Comprehensive Analysis of Mutations in Cancers and Neurodevelopmental Disorders

TEZİN TÜRÜ / DEGREE: **Yüksek Lisans /** Master ☐ **Doktora /** PhD ☒

1. **Tezin tamamı dünya çapında erişime açılacaktır. /** Release the entire work immediately for access worldwide. ☒

2. **Tez iki yıl süreyle erişime kapalı olacaktır. /** Secure the entire work for patent and/or proprietary purposes for a period of two year. * ☐

3. **Tez altı ay süreyle erişime kapalı olacaktır. /** Secure the entire work for period of six months. * ☐

** Enstitü Yönetim Kurulu Kararının basılı kopyası tezle birlikte kütüphaneye teslim edilecektir.*

A copy of the Decision of the Institute Administrative Committee will be delivered to the library together with the printed thesis.

Yazarın imzası / Signature .... **Tarih /** Date : 05.06.2023