공학석사 학위 논문

폐 선암에서 유해한 승객 돌연변이를 식별하는 모델 개발 및 검증

Development and verification of identification model of deleterious passenger mutation in lung adenocarcinoma data

울 산 대 학 교 대 학 원

의 공 학

권 솔 비

Development and verification
of identification model
# of deleterious passenger mutation
in lung adenocarcinoma data

지 도 교 수   이 재 호

이 논문을 공학석사 학위 논문으로 제출함

2021년 2월

울 산 대 학 교 대 학 원
의 공 학
권 솔 비

권솔비의 공학석사 학위 논문을 인준함

심사위원　이 재 호　(인)
심사위원　이 유 라　(인)
심사위원　박 유 랑　(인)

울 산 대 학 교　대 학 원
2021년　　2월

영문 요약

In cancer, some of somatic mutations are driver mutations but the vast majority are passenger mutations. According to cancer research, there are passenger mutations that influence various gene categories in the coding regions. Most passenger mutations in cancer are neutral for cancer progression or can be damaging to cancer cells. But passenger mutations have been excluded in cancer research. Here we identify passenger mutations that lowers the rate of lung adenocarcinoma cell growth. In the Somatic mutation data of lung adenocarcinoma cells, all mutations except driver mutations were considered as passenger mutations. Then, in the statistical analysis with non-parametric analysis, the impact of passenger mutations was analyzed by comparing IC 50 values according to occur of the passenger mutation in cells. Our approach confirms the impacts of single passenger mutation and multiple passenger mutations at the module level on lung adenocarcinoma cell. The module for passenger mutation was developed through network analysis. From analysis, we find that the IC50 was low in the presence of certain passenger mutation. A survival analysis of lung adenocarcinoma patients was used to verify if these passenger mutation affects patient survival. But passenger mutations are weak influence individually, which limits the ability to accurately identify deleterious passenger mutations in a single passenger mutation approach. We then identified their deleterious effects at multiple passenger mutations at the module level. We identified modules with low IC50 when mutations were present in all the passenger mutations that consist the module. The module was verified through IC50 values measured by different experimental methods. These modules were related to protein binding, ATP binding, and protein ubiquitination, among molecular functions. And they were related to extracellular region, nucleoplasm, and cytosol, among cellular components. In addition, our study has confirmed that the sensitivity of drugs increases when some passenger mutations occur in drug-injected lung adenocarcinoma cells, especially those injected with the antineoplastic and immunomodulating agents. Our results demonstrate the possibility that some passenger mutations can inhibit lung adenocarcinoma cell growth and this study suggests a model for identifying deleterious passenger mutations that inhibit cancer cell growth and increase drug sensitivity.

차 례

표 차례

그림 차례

Introduction

There are mutations in the human genome and most mutations are deleterious mutations with at least a small harmful effect [1]. Cancers arise owing to deleterious mutations. The mutations that are implicated in cancer development are known as 'cancer mutations' and are divided into "driver mutations" and "passenger mutations" [2]. Cancer mutations can affect the abundance and function of gene products that induce cancer growth, but on the contrary, changes in the normal functioning of cells can adversely affect the maintenance of the cells. Passenger mutations occur and accumulate because of genomic instability caused by driver mutation and cancer progression [3, 4]. These passenger mutations can adversely affect cell maintenance and survival. It is known that cancer progression is driven by the accumulation of driver mutations. However, driver mutations reside in a cancer genome alongside tens of thousands of additional mutations termed passenger mutations [5, 6].

Existing evolutionary models of cancer progression neglect the effects of passenger mutations. Passenger mutations observed in cancer biology are generally assumed to be neutral and to not play a role in cancer evolution. However, many passenger mutations fall within protein-coding genes and other functional elements that can have potentially deleterious effects on cancer cells [7]. Recently, studies on carcinogenesis have found that many mild passenger mutations accumulate in cancer and can influence cancer progression [6, 8]. In addition, if passenger mutations arise, then many can be deleterious to cancer cells [9-11], potentially via proteotoxic stress [12], loss of function [13], provoking an immune response [14], or other mechanisms [7]. Because cancer genomes contain hundreds to thousands of accumulated protein-coding passenger mutations, they may individually exert small effects, yet collectively be significant enough to alter the course of cancer progression [7, 15]. These studies indicate that passenger mutations can be distinguished by their influence.

Interest in passenger mutations has recently increased. However, sufficient research has not been completed to prove that passenger mutations can affect cancer progression. Here, we investigated the possibility that passenger mutations can increase the sensitivity of drugs and suppress cell growth. We analyzed the influence of single passenger mutation and multiple passenger mutations with a single-based approach and a module-based approach. And we developed a model to identify deleterious passenger mutations that have a detrimental effect on cancer progression. We developed and verified models using lung adenocarcinoma cells and

patients genome data. By comparing the cell death rate according to exist of the passenger mutation in the model, a passenger mutations that assist cancer cell death and suppress cancer progression could be identified. Additionally, passenger mutations that increase the sensitivity of certain drugs could be identified.

Method

1. Study design

We selected the passenger mutations of lung adenocarcinoma (LUAD) using somatic mutation data obtained from Genomics of Drug Sensitivity in Cancer (GDSC) project. And using IC50 of the drugs provided by GDSC, we were able to identify the influence of passenger mutation on LAUD cells. The rate of death between cells was compared using IC50 values to confirm different effects on LUAD cell growth based on the presence or absence of passenger mutations in the cells. Passengers are individually weaker than drivers. So we identified the effect of passenger mutations on LUAD cells by two approaches to develop a model for identifying deleterious passenger mutations that suppress LUAD cell growth. The first was a single-based approach, which identified the effect of one passenger mutation on LUAD cells. The deleterious passenger mutations were identified by comparing the IC50 values of groups divided by the presence or absence of specific passenger mutation in the LUAD cells. The other was a module-based approach, which identified the collective effect of passenger mutations on LUAD cells. The moduleshave been developed into genes that frequently occurs passenger mutations in LUAD cells. We formed a group of cells that have mutated in all the genes that make up the module and a group of cells that have little mutation in the genes that make up the module. And IC50 values between the groups were then compared and the passenger mutations were identified.

2. Database

We used the data from GDSC consisting of hundreds drugs and a panel of 1,000 cancer cell lines [16]. We assessed the whole-exome DNA sequencing and the half maximal inhibitory concentration values (IC50) of LUAD cell in GDSC. A lower value of IC50 indicates a better sensitivity of a cell line to a given drug and the faster the rate of cell death. The GDSC has two datasets for IC50 values: GDSC1 (the old drug response dataset) and GDSC2 (the new drug response dataset). GDSC2 is new using improved experimental procedures. In GDSC1, two institutes conducted experiments with five tools, of which IC50 values of three tools were used for analysis. In GDSC2, only one institute conducted experiments with three tools, of which IC50 values of one tool were used for analysis. Mutation Annotation Format files from GDSC containing only somatic mutations detected in cancer samples were downloaded.

3. Cancer gene

There are many studies on driver gene that induce cancer cell development and aid cancer progression. Driver genehas already proven their impact in cancer from many studies. Catalogue of Somatic Mutations in Cancer (COSMIC) provides 45 drivers identified from the human LUAD genome [17]. Also, GDSC provided information on 279 driver genes [16]. We were able to remove the driver mutation from somatic mutations using the driver gene list provided by GDSC and COSMIC.

4. Statistical analysis

4.1 Non-parametric analysis

Non-parametric statistical methods were used to analyze the data from LUAD cells for each drug, to assess whether any significant differences between cell groups. Wilcoxon two-tailed paired rank sum test was performed to compare IC50 between cell groups by gene. This test is a non-parametric analysis based on the ranks of the dependent variable (IC50), compared to the independent variable (drug). Through test, it can be confirmed that the cell group containing a specific passenger mutation has a low IC50 value. The gene pairs with P-value < 0.05 were selected for the following analysis. This is an exploratory study, no adjustments have been made to multiple comparisons [18, 19].

4.2 Survival analysis

A competing risk analysis by use of Cox proportional hazards regression was done to assess associations with the risk of each gene with passenger mutation-specific mortality, using proportional hazard ratios (HR) and p value [20]. In addition, log-rank test and Kaplan-Meier were used. Kaplan-Meier analysis assessed survival characteristics of the patient groups by gene with passenger mutation [21, 22]. Log-rank test determined statistical significance on Kaplan-Meier analysis. Statistical differences with a p value <0.05 were considered significant.

5. Identification of passenger mutation interaction and significant modules

Passenger mutations mutated from over 75% of the 53 LUAD cells collected from GDSC2 were selected and a modules were developed with these passenger mutations. The modules of passenger mutations were developed in Gephi software.

Gephi 0.9.2 is a software of manipulation and exploration of the module, working with complex data set. Gephi offers a built-in function to filter edges from the module based on edge weight. The edge weight was calculated as the number of simultaneous mutations in the passenger mutations in 53 LUAD cells. We used this feature to focus on significant co-occurrence attribute pairs. And modules were constructed with the following threshold values, p=0.7.

6. Pathway enrichment analysis

Pathway enrichment analysis was performed to determine the signaling pathways in which passenger mutations are active. The biological function of the DEGs was investigated by a GO terms [23]. GO terms can bedivided into the following three parts: (1) Biological Process; (2) Cellular Component; and (3) Molecular Function. Kyoto Encyclopedia of Genes and Genomes (KEGG) is a database resource that integrates genomic, chemical and systemic functional information [24]. We chose KEGG and GO because they are widely used as a reference knowledge base for integration and interpretation of large-scale datasets generated by genome sequencing and other high-throughput experimental technologies. In the present study, the GO and KEGG pathway enrichment analysis were executed using R package with a P value less than 0.05.

7. Drug categories

The DrugBank database stores drug data with corresponding drug target and treatment indication information. The DrugBank includes 1,692 approved/experimental drugs. Drug Bank was used to classify the drugs into categories and to collect information on the drugs.

Result

1. Sample cohort

A total of 62 LUAD cells were collected from the GDSC1 and 53 LUAD cells were collected from the GDSC2 (Table 1). A total of 572,316 Single nucleotide polymorphism (SNP) in 24,442 genes and 482,035 Single nucleotide polymorphism (SNP) in 19,388 genes. There were IC50 values for 297 drugs in 62 cells and 136 drugs in 53 cells. The cells had a mean of 127 driver genes and 4,914 gene with mutation and 128 driver genes and 4,778 genes with passenger mutation.

Table 1. Sample information identifiers of the study population

Features	GDSC1	GDSC2
Cell lines, n	62	53
Gender, n (%)		
Female	24 (38.71)	22 (41.51)
Male	31 (50.00)	27 (50.94)
Unknown	7 (11.29)	4 (7.55)
Age (Mean ±SD)	47.96 ± 18.73	50.84 ± 14.59
Drug, n (%)		
1046	78 (24.92)	0
1045	191 (61.02)	0
1043	0	18(11.54)
1025	7 (2.24)	0
1019	36 (11.50)	136 (87.18)
2018	0	2 (1.28)
1009	1 (0.32)	0
IC 50 (Mean ±SD)	2.22 ± 2.61	3.18 ± 2.76
Cancer mutation		
Driver mutation (Mean ±SD)	252.1 ± 122.28	251.8 ± 118.30
Passenger mutation (Mean ±SD)	9,203 ± 3,233.18	8,843 ± 2946.11
Tumor mutation burden (Mean ±SD)	15.67 ± 11.70	16.148 ± 12.51

2. A single-based approach for distinguishing deleterious passenger mutations

2.1 Identification of deleterious passenger mutations

We compared the rate of cell death between cell groups using IC50 from drugs provided by the GDSC1 and GDSC2. Tools that experimented with less than 20 of the 5 tools in GDSC1 and 3 tools in GDSC2 were excluded from the analysis. Non-parametric method was used to analyze the impact of 19,095 genes with passenger mutation selected from GDSC2 LUAD cell somatic mutation data and 24,150 genes with passenger mutation selected from GDSC1 LUAD cell somatic mutation data. We added an option to wilcox to identify genes with passenger mutation in which the cell group with passenger mutation was significantly lower in IC50 than the cell group without passenger mutation. In GDSC1, 19,201, 16,419, and 8,947 genes with passenger mutation showed significant results. In GDSC2, 11,001 genes with passenger mutation showed significant results (Table 2).

Table 2. Results of wilcox on single-based approach

Features	GDSC1			GDSC2
	1046	1045	1019	1046
Cell lines, n	62	59	62	53
Genes with passenger mutation, n	24,150	21,802	24,150	19,095
Drug, n	78	191	36	136
Wilcox (one-sided test : "less")				
P-value <0.05, n	71,823	160,017	30,957	65,752
Genes with passenger mutation, n	19,201	16,419	8,947	11,001
Drug, n	78	191	36	136

Passenger mutations that inhibit cancer cell growth were estimated using three IC50 value datasets measured by three tools provided by the GDSC1 and one IC50 value dataset measured by one tool provided by the GDSC2. In a total of four datasets, different numbers of passenger mutations that inhibit cancer cell growth were estimated. A total of 5,073 genes with passenger mutation were estimated to overlap in all four datasets (Figure 1). These 5,073 genes with passenger mutation had the effect of suppressing cancer cell growth in LUAD cells, and we identified them as putative genes with deleterious passenger mutation.
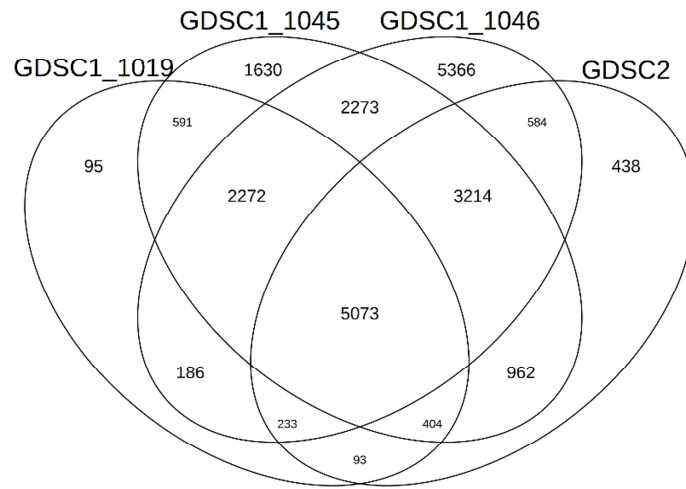
Figure 1. Venn diagrams indicating the total number of putative genes with deleterious passenger mutation in LUAD that inhibit cancer cell growth retrieved by 4 different IC50 datasets.

2.2 Validation

Validation analysis was performed to confirm whether putative genes with deleterious passenger mutation identified using the LUAD cell's IC50 data can also be identified in the LUAD patient. TCGA project profiles the genomics of about 10,000 patient samples across over 30 cancer types. Restricted access whole-genome sequences and whole-exome sequences were obtained from the TCGA data portal. We assessed the whole-exome DNA sequencing and survival data for LUAD patients from TCGA. A total of 230 LUAD patients primary cancer samples were collected from the TCGA (Table 3). All 15,028 genes with mutation detected within 230 LUAD patient primary samples and we used only single nucleotide polymorphism mutations.

Table 3. Sample information identifiers of the validation study population

Features	Lung adenocarcinoma
Patient, n	230
Gender, n (%)	
Female	131 (56.96)
Male	99 (43.04)
Cancer specific death, n (%)	68 (29.57)
Cancer mutation	
Driver mutation (Mean ±SD)	10.23 ± 8.09
Passenger mutation (Mean ±SD)	283.70 ± 286.83
Tumor mutation burden (Mean ±SD)	0.0029 ± 0.0044

Of the 5,073 putative genes with deleterious passenger mutations selected from LUAD cells, 3,461 putative genes with deleterious passenger mutations were present in 230 LUAD patients. For each of the 3,461 putative genes with deleterious passenger mutations, 230 LUAD patients were divided into case with a passenger mutation and control without a passenger mutation. Survival analysis was conducted for the patient groups of 3,461 putative genes with deleterious passenger mutations.

The overall unmatched survival analyses were performed, including all LUAD patients, to compare groups. It was confirmed that the case in gene DNAH3 with passenger mutation, had a lower risk of death than control (Hazard ratio = 0.33, P value = 0.03) (Table 4, Figure 2).

Table 4. Survival result for DNAH3 in LUAD patients

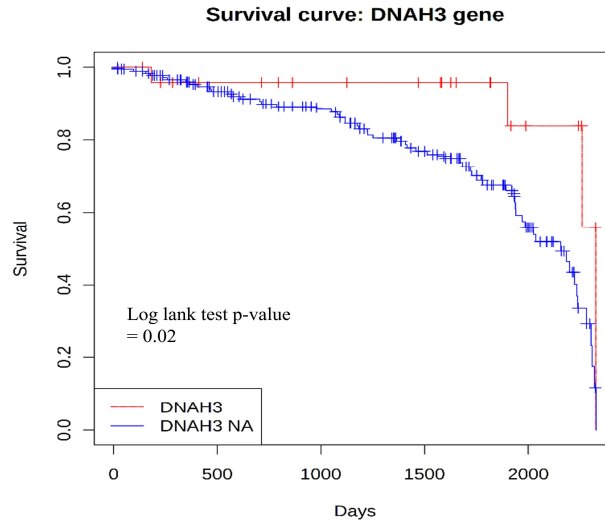| Gene	TCGA LUAD patients		Cox	
	Case (Alive/Dead)	Control (Alive/Dead)	p.value	*Hazard ratio (95% CI)
DNAH3	27 (23/4)	203 (139/64)	0.0325	0.327 (0.118-0.911)

Survival curve: DNAH3 gene

Figure 2. Kaplan Meier plot about DNAH3

3. A modules-based approach for distinguishing deleterious passenger mutations

3.1 Development of networks and identification of deleterious passenger mutations modules

We developed modules and analyzed whether the IC50 value changes according to exist of passenger mutations that compose the module. The module was developed in Gephi software to identify the interaction among genes with passenger mutation. These modules were derived from a dataset comprising gene pairs of 53 LUAD were collected from the GDSC2. The module was developed using passenger mutations mutated in more than 75% of 53 LUAD cells. There were 19,095 genes with passenger mutations in 53 cells, of which 1,044 genes with passenger mutations were selected. Pathway enrichment analysis was performed on the selected genes.

We detected a total of seventy-one modules. LUAD cells in which mutations are all present in the genes constituting the modules is classified as a case, and LUAD cells in which two or fewer passenger mutations are present is classified as a control. Of the seventy-one modules, only twenty-five modules could be grouped. We compared the IC50 values between groups using wilcox to identify the modules

10

that affect suppressing cancer cell growth. And drugs with significantly lower IC50 values in case were identified (Table 5).

Among the twenty-five modules, there were significant lower IC50 values in case in more than one drugs in fourteen modules. Of these, three modules showed results with significantly lower IC50 values at case in more than 10% of drugs, module 31, module 61, and module 62. The module 31 consisting of four genes, ATF7IP, ENSG00000236954, RAD52, and TG (Table 6). The IC50 values for the 17 (11.18%) drugs were significantly lower in the case.

The module 61 consisting five genes, CUL9, DNAH1, RAD52, ZNF813, and WAC (Table 6). The IC50 values for the 26 (17.22%) drugs were significantly lower in the case. Of the five genes contained in the module 61, CUL9 and WAC were related to protein binding in the molecular function of GO term and protein ubiquitination in the biological process of GO term. CUL9 and DNAH1 were related to ATP binding in the molecular function of GO term (Table 7).

The module 62 consisting of seven genes, ACTR3B, ANAPC1, COL21A1, COL24A1, FAF2, ECPAS, and LRRC16A (Table 6). The IC50 values for the 31 (20.67%) drugs were significantly lower in the case. Of the seven genes contained in the module 62, COL24A1, FAF2, ECPAS, and LRRC16A were related to protein binding in the molecular function of GO term. COL21A1, COL24A1, and FAF2 were related to extracellular region in the cellular component of GO term. ANAPC1, ECPAS, and LRRC16A were related to nucleoplasm in the cellular component of GO term. ANAPC1, COL21A1, and LRRC16A were related to cytosol in the cellular component of GO term (Table 7).

In 83 drugs, the IC50 value was significantly lower when passenger mutations were present in all the genes constituting the module. We divided the drugs into categories by reference to the DrugBank. We selected the top five categories that contain the most drugs: (1) Antineoplastic Agents (33.73%); (2) Antineoplastic and Immunomodulating agents (30.12%); (3) Enzyme Inhibitors (21.69%); (4) Immunosuppressive Agents (21.69%); and (5) Myelosuppressive Agents (20.48%) (Table 8). The 83 drugs included drugs used to treat other cancers, as well as lung cancer.

Table 5. Results of wilcox on module-based approach

Module id	Drug, n (%)
Module 0	5 (3.29)
Module 5	2 (1.32)
Module 7	2 (1.32)
Module 12	1 (0.66)
Module 14	3 (1.97)
Module 15	13 (8.55)
Module 31	17 (11.18)
Module 39	7 (4.61)
Module 56	1 (0.66)
Module 60	3 (1.99)
Module 61	26 (17.22)
Module 62	31 (20.67)
Module 64	2 (1.32)
Module 70	5 (5.1)

Table 6. List of member genes in the passenger mutation module associated with inhibiting LUAD cell growth.

Module ID	Gene	Description	Molecular function
Module 31	ATF7IP	Activating transcription factor 7-interacting protein 1	ATPase activity, transcription corepressor activity
	ENSG00000236954	Unknown	Unknown
	SCN9A	Sodium channel protein type 9 subunit alpha	voltage-gated sodium channel activity
	TG	Thyroglobulin	hormone activity, identical protein binding
Module 61	CUL9	Cullin-9	ubiquitin protein ligase binding
	DNAH1	dynein axonemal heavy chain 1	ATPase activity, microtubule motor activity
	FAM104B	family with sequence similarity 104 member B	Unknown
	ZNF813	Zinc finger protein 813	nucleic acid binding
	WAC	WW domain-containing adapter protein with coiled-coil	chromatin binding, RNA polymerase II complex binding
Module 62	ACTR3B	Actin-related protein 3B	actin binding, ATP binding
	ANAPC1	Anaphase-promoting complex subunit 1	molecular adaptor activity
	COL21A1	Collagen alpha-1(XXI) chain	extracellular matrix structural constituent conferring tensile strength
	COL24A1	Collagen alpha-1(XXIV) chain	extracellular matrix structural constituent
	FAF2	FAS-associated factor 2	lipase binding, lipase inhibitor activity, ubiquitin binding, ubiquitin protein ligase binding
	ECPAS	Proteasome adapter and scaffold protein ECM29	molecular adaptor activity, proteasome binding
	LRRC16A	F-actin-uncapping protein LRRC16A	protein-containing complex binding

Table 7. Categorization of member genes in the passenger mutation module by reference to Gene Ontology.

Module ID	Category	Term	Definition	Gene name
Module 61	Molecular function	protein binding	Interacting selectively and non-covalently with any protein or protein complex (a complex of two or more proteins that may include other nonprotein molecules).	CUL9, WAC
	Molecular Function	ATP binding	Interacting selectively and non-covalently with ATP, adenosine 5'-triphosphate, a universally important coenzyme and enzyme regulator.	CUL9, DNAH1
	Biological Process	protein ubiquitination	The process in which one or more ubiquitin groups are added to a protein.	CUL9, WAC
Module 62	Molecular Function	protein binding	Interacting selectively and non-covalently with any protein or protein complex (a complex of two or more proteins that may include other nonprotein molecules).	COL24A1, FAF2, ECPAS, LRRC16A
	Cellular Component	extracellular region	The space external to the outermost structure of a cell. For cells without external protective or external encapsulating structures this refers to space outside of the plasma membrane. This term covers the host cell environment outside an intracellular parasite.	COL21A1, COL24A1, FAF2
	Cellular Component	nucleoplasm	That part of the nuclear content other than the chromosomes or the nucleolus.	ANAPC1, ECPAS, LRRC16A
	Cellular Component	cytosol	The part of the cytoplasm that does not contain organelles but which does contain other particulate matter, such as protein complexes.	ANAPC1, COL21A1, LRRC16A

Table 8. Categorization of drugs by reference to DrugBank

DrugBank category	Drug, n (%)	ATC Classification	Description
Antineoplastic Agents	28 (33.73)	Antineoplastic and immunomodulating agents	Substances that inhibit or prevent the proliferation of neoplasms.
Antineoplastic and Immunomodulating agents	25 (30.12)	Antineoplastic and immunomodulating agents	Not Available
Enzyme Inhibitors	18 (21.69)	Not Available	Compounds or agents that combine with an enzyme in such a manner as to prevent the normal substrate-enzyme combination and the catalytic reaction.
Immunosuppressive Agents	18 (21.69)	Antineoplastic and immunomodulating agents	Agents that suppress immune function by one of several mechanisms of action. Classical cytotoxic immunosuppressants act by inhibiting DNA synthesis.
Myelosuppressive Agents	17 (20.48)	Not Available	Not Available

3.2 Validation

Validation analysis was performed to confirm whether modules identified using the cell's IC50 data in GDSC2 can also be identified in GDSC1. GDSC1 had somatic mutation data of 62 LUAD cells and IC50 value dataset for 313 drugs. Groups were divided into cases and controls using the module developed from GDSC2, and IC50 values between groups were compared using wilcox. IC50 values were significantly lower in case for one or more drugs in nineteen modules (Table 9). Of these, five modules showed results with significantly lower IC50 values at case in more than 10% of drugs, module 31, module 39, module 61, module 62, and module 70.

Table 9. Validation results of wilcox on module-based approach

Module id	Drug, n(%)
Module 0	12 (3.95)
Module 5	21 (6.91)
Module 7	2 (0.76)
Module 12	1 (1.35)
Module 14	4 (1.32)
Module 15	21 (6.91)
Module 19	10 (3.77)
Module 31	102 (33.55)
Module 35	17 (6.42)
Module 39	32 (10.53)
Module 50	5 (1.89)
Module 56	1 (0.38)
Module 57	1 (0.33)
Module 58	1 (0.33)
Module 61	64 (24.15)
Module 62	86 (28.29)
Module 64	2 (0.75)
Module 66	25 (8.22)
Module 70	51 (17.29)

Discussion

We have developed and verified a model for identifying deleterious passenger mutation that suppresses LUAD cell growth and increases drug sensitivity. Passenger mutations in LUAD identified in cancer belong to protein-coding genes and other functional elements. So, passenger mutations can also affect cells. Our model used somatic mutation data and IC50 values for hundreds of drugs. The faster the cell dies, the lower the IC50 value is measured. Therefore, the lower the IC50 value, the more cancer cell growth is inhibited and the sensitivity of the drug increases. This model shows the dynamic relationship between the occurrence of passenger mutations and the rate of cell growth.

First, a single-based approach was used to study whether deleterious passenger mutations could be identified by the influence of a single passenger mutations. We identified groups of genes with passenger mutation with significantly lower IC50 values in four IC50 value datasets measured using different drugs. A total of 5,073 genes with passenger mutation were identified as overlap in four IC50 sets. And we tried to verify the effects of these passenger mutations by measuring whether the risk of death in LUAD patients varies according to the presence of these passenger mutations. Verification analysis has shown that the survival rate of LUAD patients only increases when passenger mutation occurs in DNAH gene. The single-based approach could confirm that some passenger mutations in some drugs inhibit LUAD cell growth, but there were limitations in verifying deleterious passenger mutations with weak influence. So, we have proposed a network-based approach to the problem. We selected passenger mutations that occurs frequently in LUAD cells and developed modules. The growth of LUAD cells in more than one drug was inhibited when passenger mutations occurred in all genes contained the module in 14 modules. We used IC50 values measured in more drugs from other experimental ways to carry out the verification of the effects on LUAD cell growth according to the presence of passenger mutations in the module. Of the 14 modules, 13 modules were verified.

A module-based approach was more appropriate to measure the effect of passenger mutations and identify deleterious passenger mutations than a single-based approach. Through a module-based approach, we were able to identify three modules (module 31, module 61, and module 62) that inhibit the growth of LUAD cells in the drugs over 10%. We estimated that these three modules had the collective influence of deleterious passenger mutations, and we identified a passenger mutation occurring in

the genes that consist these modules as deleterious passenger mutations. Module 31 consisted of a small number of genes, so it was difficult to find common traits among genes.

In module 61, CUL9 and WAC were involved in protein binding and protein ubiquitination, among molecular functions and biological processes. CUL9 and DNAH1 were involved in ATP binding, among molecular functions. Protein binding can affect the efficiency of the drug. The degree of drug attached to the protein determines the efficiency of the drug [25-27]. Thus, changing protein binding so that the drug can be well attached can increase the sensitivity of the drug, and cell growth can be inhibited. ATP is the energy needed to carry out the various life activities in living cells and is essential for the cell to survive [28-31]. If mutations change the ATP binding and do not produce ATP, cells can death.

In module 62, COL24A1, FAF2, KIAA0368, and LRRC16A were involved in protein binding. COL21A1, COL24A1, and FAF2 were involved in extracellular region, among cellular components. ANAPC1, KIAA0368, and LRRC16A were involved in nucleoplasm, among cellular components. ANAPC1, COL21A1, and LRRC16A were involved in cytosol, among cellular components. Changes in protein binding due to mutations can increase the sensitivity of the drug and inhibit the growth of cancer cells. Extracellular regions have been associated with communicating anti-cancer drug-resistance [32]. Mutation in in the expression levels of extracellular matrix components and related genes changed the expression level and lowered the resistance of the drug. Some studies have shown that COL21A1 can affect drug resistance by preventing the penetration of the drug into cancer cells and increased resistance to apoptosis in ovarian and breast cancer [33, 34]. Cytosol is the intra-cellular fluid that is present inside the cell and is involved in the transport of molecules within the cell [35, 36]. Within the Cytosol, the proapoptotic factor is carried and activated to assist induce the destruction of the nucleus and the apoptosis of the cells [37]. Moreover, many important cellular processes and reactions, especially of metabolic character, occur in the cytosol itself [36, 38].

In the module-based approach, the growth of LUAD cells was inhibited in a total of 83 drugs. 30 drugs (36.14%) of the 83 drugs were classified antineoplastic and immunomodulating agents that inhibited cancer growth and stimulates or suppresses the immune system. 5 drugs (6.02%) of the 83 drugs were being used to treat lung cancer. Among the 83 drugs, 14 drugs for the treatment of breast cancer was the most common (16.87%). Other than that, drugs for the treatment of solid cancer, solid cancers, gastric cancer, colorectal cancer, pancreatic cancer, kidney cancer, ovarian cancer, liver cancer, skin cancer, cervical cancer, melanoma, lymphatic,

neoplasms, diabetes, atherosclerosis, and coronary artery disease were included. Through a modular basic approach, deleterious passenger mutations identified in LUAD cells can increase the sensitivity of drug to other cancers. This shows that there can be common deleterious passenger mutations in various cancers. And the influence of deleterious passenger mutations could increase drug sensitivity and inhibit cell growth in other cancer cells as well as LUAD cells. These results suggest that the progression of various cancers can be inhibited by inducing deleterious passenger mutations in cancer cells.

Despite these findings, passenger mutations are widely known as cancer-neutral genes. This shows that there is a lack of research on passenger mutations, and that it is an area to be studied in the future. We have found that deleterious passenger mutations by two approaches can affect LUAD cell growth and drug sensitivity, and model using the module basic approach was suitable for identifying deleterious passenger mutations. Measuring the influence of passenger mutations in cancer and classifying passenger mutations by role will help you understand cancer. And most of all, distinguishing the deleterious passenger mutations from the passenger mutations can help predict, diagnose and treat cancer, especially in the field of personalized medicine and therapy. Passenger mutations are need more focus and research in future cancer researches.

Limitation

Our study has several limitations. Firstly, the number of LUAD cells in the GDSC was small with 63 cells, it was difficult to find cells with the same passenger mutation. So, only the influence of some passenger mutations in the LUAD cell could be identified. Secondly, there were fewer passenger mutations in LUAD patients in TCGA than in LUAD cells in GDSC. Therefore, all passenger mutations who are deleterious to the LUAD cell growth identified from LUAD cells could not be verified, and only a part of them was verified and analyzed. So, another analysis was performed using module-based approach.

Conclusion

This study proposes a model to identify deleterious passenger mutations by analyzing somatic mutation data and IC50 values by drugs using module-based approach and statistical method. And we confirmed that deleterious passenger mutations can change the direction of the LUAD cell growth and increase drug sensitivity when they exist collectively. And the drug sensitivity of drugs that treat other cancers as well as LUAD has also increased. This suggests that other cancers also have deleterious passenger mutations that inhibit cancer cell growth and increase drug sensitivity. These tentative conclusions await further refinement and correction in the light of further research of other cancers.

Reference

1. Eyre-Walker, A. and P.D. Keightley, *High genomic deleterious mutation rates in hominids.* Nature, 1999. 397(6717): p. 344-347.

2. Futreal, P.A., et al., *A census of human cancer genes.* Nature reviews cancer, 2004. 4(3): p. 177-183.

3. Greenman, C., et al., *Patterns of somatic mutation in human cancer genomes.* Nature, 2007. 446(7132): p. 153.

4. Ji, X., et al., *Distinguishing between cancer driver and passenger gene alteration candidates via cross-species comparison: a pilot study.* BMC cancer, 2010. 10(1): p. 426.

5. Lawrence, M.S., et al., *Discovery and saturation analysis of cancer genes across 21 tumour types.* Nature, 2014. 505(7484): p. 495-501.

6. McFarland, C.D., et al., *A genetic model of metastatic evolution: Driver and passenger mutations affect metastatic fitness.* 2011, AACR.

7. McFarland, C.D., et al., *Impact of deleterious passenger mutations on cancer progression.* Proceedings of the National Academy of Sciences, 2013. 110(8): p. 2910-2915.

8. Kryukov, G.V., L.A. Pennacchio, and S.R. Sunyaev, *Most rare missense alleles are deleterious in humans: implications for complex disease and association studies.* The American Journal of Human Genetics, 2007. 80(4): p. 727-739.

9. Moran, N.A., *Accelerated evolution and Muller's rachet in endosymbiotic bacteria.* Proceedings of the National Academy of Sciences, 1996. 93(7): p. 2873-2878.

10. Walczak, A.M., et al., *The structure of genealogies in the presence of purifying selection: a fitness-class coalescent.* Genetics, 2012. 190(2): p. 753-779.

11. Lynch, M., et al., *The mutational meltdown in asexual populations.* Journal of Heredity, 1993. 84(5): p. 339-344.

12. Neznanov, N., et al., *Proteotoxic stress targeted therapy (PSTT): induction of protein misfolding enhances the antitumor effect of the proteasome inhibitor bortezomib.* Oncotarget, 2011. 2(3): p. 209.

13. Boyko, A.R., et al., *Assessing the evolutionary impact of amino acid mutations in the human genome.* PLoS genetics, 2008. 4(5).

14. Andersen, M.H., et al., *Cytotoxic T cells.* Journal of Investigative Dermatology, 2006. 126(1): p. 32-41.

15. McFarland, C.D., et al., *The damaging effect of passenger mutations on cancer progression.* Cancer research, 2017. 77(18): p. 4763-4772.

16. Yang, W., et al., *Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells.* Nucleic acids research, 2012. 41(D1): p. D955-D961.

17. Sondka, Z., et al., *The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers.* Nature Reviews Cancer, 2018: p. 1.

18. Althouse, A.D., *Adjust for multiple comparisons? It's not that simple.* The Annals of thoracic surgery, 2016. 101(5): p. 1644-1645.

19. Jafari, M. and N. Ansari-Pour, *Why, when and how to adjust your P values?* Cell Journal (Yakhteh), 2019. 20(4): p. 604.

20. Schoenfeld, D., *Partial residuals for the proportional hazards regression model.* Biometrika, 1982. 69(1): p. 239-241.

21. Goel, M.K., P. Khanna, and J. Kishore, *Understanding survival analysis: Kaplan-Meier estimate.* International journal of Ayurveda research, 2010. 1(4): p. 274.

22. Efron, B., *Logistic regression, survival analysis, and the Kaplan-Meier curve.* Journal of the American statistical Association, 1988. 83(402): p. 414-425.

23. Botstein, D., et al., *Gene Ontology: tool for the unification of biology.* Nat genet, 2000. 25(1): p. 25-29.

24. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes.* Nucleic acids research, 2000. 28(1): p. 27-30.

25. Oravcova, J., B. Bo, and W. Lindner, *Drug-protein binding studies new trends in analytical and experimental methodology.* Journal of Chromatography B: Biomedical Sciences and Applications, 1996. 677(1): p. 1-28.

26. Benet, L.Z. and B.A. Hoener, *Changes in plasma protein binding have little clinical relevance.* Clinical Pharmacology & Therapeutics, 2002. 71(3): p. 115-121.

27. Vayá, I., et al., *Photoactive assemblies of organic compounds and biomolecules: drug–protein supramolecular systems.* Chemical Society Reviews, 2014. 43(12): p. 4102-4122.

28. Richter, K., et al., *Intrinsic inhibition of the Hsp90 ATPase activity.* Journal of Biological Chemistry, 2006. 281(16): p. 11301-11311.

29. Hinton, A., S. Bond, and M. Forgac, *V-ATPase functions in normal and disease processes.* Pflügers archiv-European journal of physiology, 2009. 457(3): p. 589-598.

30. Serrano, R., *Structure and function of plasma membrane ATPase.* Annual review of plant biology, 1989. 40(1): p. 61-94.

31. Stewart, A.G., et al., *Rotary ATPases—dynamic molecular machines.* Current opinion in structural biology, 2014. 25: p. 40-48.

32. Mc Namee, N. and L. O'Driscoll, *Extracellular vesicles and anti-cancer drug resistance.* Biochimica et Biophysica Acta (BBA)-Reviews on Cancer, 2018. 1870(2): p. 123-136.

33. Januchowski, R., et al., *Extracellular matrix proteins expression profiling in chemoresistant variants of the A2780 ovarian cancer cell line.* BioMed research international, 2014. 2014.

34. Huang, P., et al., *lncRNA profile study reveals the mRNAs and lncRNAs associated with docetaxel resistance in breast cancer cells.* Scientific reports, 2018. 8(1): p. 1-15.

35. Clegg, J.S., *Properties and metabolism of the aqueous cytoplasm and its boundaries.* American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 1984. 246(2): p. R133-R151.

36. Luby-Phelps, K., *Cytoarchitecture and physical properties of cytoplasm: volume, viscosity, diffusion, intracellular surface area*, in *International review of cytology*. 1999, Elsevier. p. 189-221.

37. Ferrando-May, E., *Nucleocytoplasmic transport in apoptosis.* Cell Death & Differentiation, 2005. 12(10): p. 1263-1276.

38. Carcamo, W.C., et al., *Molecular cell biology and immunobiology of mammalian rod/ring structures*, in *International review of cell and molecular biology*. 2014, Elsevier. p. 35-74.

국문 요약

암은 흔히 정상세포의 유전자에 돌연변이가 생겨 발생하는 질병으로 잘 알려져 있다. 이러한 돌연변이에 대한 이해는 암 발생의 근본적인 과정을 이해하는 데 도움을 주고, 개인 맞춤형 암 치료법을 개발할 수 있는 정보를 제공한다.

암 돌연변이는 암 성장을 유도 할 수도 있지만 반대로 세포의 정상적인 기능을 변화시켜 세포 생존에 악영향을 미칠 수 있다. 암 성장을 유도하는 운전자 돌연변이로 인해 발생한 유전적 불안정성은 승객 돌연변이를 발생시키고 축적시킨다. 최근 연구에 따르면, 이러한 승객 돌연변이가 세포의 기능 유지와 생존에 악영향을 미칠 수 있다고 한다. 하지만 승객 돌연변이는 암 진행에 중립적인 역할로 알려지면서 암 연구에서 제외되어 연구가 진행되고 있다.

본 연구에서는 승객 돌연변이들 중 일부가 암 세포의 생존에 해로운 영향을 줄 수 있다는 것을 검증하였다. 또한 암세포의 생존에 해로운 영향을 미치는 승객 돌연변이를 식별하기 위한 모델을 제한하고 검증하였다.

폐 선암 세포의 체세포 돌연변이 데이터에서 운전자 돌연변이를 제외한 모든 돌연변이는 승객 돌연변이로 분류하였다. 세포 내 승객 돌연변이의 발생 유무에 따라 세포그룹을 나누고, 비모수 통계검정 방법을 사용하여 각 세포에서 측정된 IC50 값을 비교하여 승객 돌연변이가 암 세포의 죽음에 미치는 영향력을 측정하였다. 본 연구에서는 단일 승객 돌연변이와 네트워크 분석으로 식별된 다중 승객 돌연변이가 암 세포의 죽음에 미치는 영향력을 평가하였다.

결과적으로, 승객 돌연변이는 개별적으로는 세포기능에 약한 영향을 미치기 때문에 단일 승객 돌연변이에 대한 영향력을 검증하는데 한계가 있었다. 하지만 네트워크 분석을 통해 식별된 승객 돌연변이 모듈이 세포기능에 미치는 영향력은 검증 할 수 있었다. 단백질 결합과 ATP 결합에 연관된 유전자에서 발행한 승객 돌연변이들이 포함되어 있는 승객 돌연변이 모듈은 암 세포의 사멸 속도를 높이는데 도움을 주었다. 이러한 결과를 통해 우리는 일부 승객 돌연변이가 암 세포의 사멸을 촉진시킬 수 있다는 것을 증명하였다.

승객 돌연변이는 암 진행에 있어 중립적인 역할로 알려지면서 암 연구에서 제외가 되었지만, 앞으로는 승객 돌연변이를 포함한 암 연구를 진행해야 한다. 승객 돌연변이가 암 세포에 미치는 영향력을 측정하고 승객 돌연변이를 역할별로 분류하는 것은 암을 이해하는 데 도움이 될 것이다. 또한 암 진행에 해로운 승객 돌연변이를 식별함으로써, 암을 예측하고, 암 환자 맞춤형 치료를 개발하는 데 도움이 될 수 있을 것이다.