SAPIENZA
UNIVERSITÀ DI ROMA

TODO Titolo

Faculty of Information Engineering, Computer Science and Statistics
Bachelor's Degree in Computer Science

Alessio Bandiera
ID number 1985878

Advisor
Prof. Ivano Salvo

Academic Year 2023/2024

TODO Titolo
Bachelor's Thesis. Sapienza University of Rome

© 2024 Alessio Bandiera. All rights reserved

This thesis has been typeset by LaTeX and the Sapthesis class.

Author's email: alessio.bandiera02@gmail.com

TODO.

Contents

Chapter 1

Introduction

1.1 Context

1.1.1 Cancer

Cancer is a medical condition characterized by an uncontrolled proliferation of the cells, allowing their infiltration into organs and tissue, altering their functions and structure. The exponential growth is caused by mutations in the cellular DNA, because it contains information that describes how cells should develop and multiply, and errors in these instructions may lead the cell to become cancerous. In the vast majority of cancer types one single aberration is not sufficient for cancer development and multiple mutations are necessary, some already present at birth like in cellular DNA, and others obtained throughout life either by chance or by lifestyle choices. Moreover, for the tumor to proliferate mutations on genes that regulate cell growth are needed [1], in particular **proto-oncogenes** which promote mitosis, and **tumor suppressor genes** which discourage cell growth — a process called **oncoevolution**.

talk about oncogenes in depth?

1.1.2 Cancer treatment

Research towards finding a cure for cancer is in constant development, because of the tumor's lethality and complexity. At present, the following are the main techniques used to remove, control, restrain and delay the effects of cancer [2]:

- **surgery**, which allows the removal of the cancerous region, and it's usually reserved for solid tumors;
- **radiotherapy**, which uses x-rays to destroy tumor cells, and it's aimed towards the cancerous region as much as possible to preserve healthy cells; radiotherapy can increase the risk of developing other tumors because of x-rays, such as leukemia or sarcomas, and it can lead to delayed effects like dementia, amnesia or progressively worsening cognitive difficulties;

- **chemotherapy**, which blocks cellular division — through cytotoxic drugs — of both cancerous and healthy cells, inducing side effects to every rapidly renewing tissue;
- **hormone therapy**, through which the balance of specific hormones gets altered, which can lead to some side effects such as joint pain or osteoporosis;
- **target therapy**, i.e. drugs containing antibodies or inhibitory substances that target the cancer cell and promote its destruction by the immunity system; target therapy may be difficult to develop depending on the structure or the function of the target, it may induce unwanted side effects to various organs, and cancer cells can become increasingly resistant to this type of therapy if they find a way to develop which does not depend on the therapy's target [3] .

check this out

In particular, in recent years target therapy has been the subject of lots of research, because it could be a way to affect only the desired target, helping to reduce the side effects which currently characterize every available cancer cure, possibly limiting the damage to healthy cells [4].

expand target therapy on how it works? if yes, make subsection

1.2 Mutations

1.2.1 Cell signaling and signaling pathways

Cell signaling is the process through which cells interact with themselves, other cells, or their environment; through cell signaling, signals are transduced, and they can be of many types, usually chemical but also pressure, voltage, temperature, or light signals [5]. **Pathways** are a series of actions between molecules inside a cell which lead to a change in the cell or the creation of some product [6]. Pathways have a *direction* in which the actions happen, and the terms *upstream* and *downstream* are used to indicate what happens at their beginning or their end respectively. For cancer study, of particular interest are **signaling pathways**, which allow the transduction of cell signals, because locating and blocking pathways that are responsible for the core functions of cancer growth could terminate the development of the latter.

check this out, also check if what i wrote is actually true, i think i read it somewhere but can't find the source right now; expand on cell signaling? expand of pathways? if yes, make subsections

1.2.2 Passenger and driver mutations

There are two types of mutations in cancer: **passenger** and **driver** mutations. Passenger mutations don't provide direct benefits to tumor growth and development, while driver mutations can directly influence cancer, providing evolutionary advantage and allowing an increase in the number of tumor cells. A **driver gene** is a gene that contains at least one driver mutation, but it can also contain passenger mutations ; a **driver pathway** is made up of at least one driver gene. Driver mutations, genes, and pathways are of great scientific interest since hold an important role in cancer proliferation.

DO I ADD THIS A CITATION?

Driver genes can be classified into 12 signaling pathways, which regulate functions of survival, fate, and genomic maintenance of the cell .

use (and expand) this? same source as prev

1.3 Classifying mutations

1.3.1 Frequency

To classify mutations into the two described categories, it is necessary to check their biological function, to this day a difficult task to accomplish. There are lots of methods that allow predicting the functional impact of mutations through *a priori* knowledge, but usually, these approaches are not able to integrate information adequately across the various types of mutations, and because they are based on known information about already discovered proteins they are less successful for less studied ones. Thanks to the decreasing cost of DNA sequencing, it is now possible to categorize mutations by studying their frequency, i.e. driver mutations are proportionately the most recurrent in patients' genomes. However, this approach often fails because driver mutations vary between cancer patients' samples, even those having the same cancer type, significantly reducing the statistical potential. [7]. It was found that there is little overlap between mutated genes over sample pairs even taken from the same patient [8]. This heterogeneity is mainly explained by the fact that driver mutations are mainly found in genes that are part of cell signaling pathways, thus different patients may harbor mutations in different pathway loci. Consequentially, it is not sufficient to look into single gene frequency, but it must be tested whether group genes are recurrently mutated. Moreover, the study must be done at the pathway level, because it is well known that there may be different mutations inside the same pathway, across multiple samples [7].

ritrova il paper che parlava della frequenza più in dettaglio

1.3.2 Mutual exclusivity and coverage

MISSING THE FIRST SENTENCE, UNDERSTAND MULTI-DENDRIX THING

The majority of techniques developed in recent years for driver mutation recognition use a statistical property which is evident from data of cancer patient: each patient has a relatively small number of mutations, which affect multiple pathways, thus each pathway will contain *1 single driver mutation on average* per sample. This idea of **mutual exclusivity** of driver mutations inside the same pathway, statistically observed in patients' samples, is then axiomatized and used by the research algorithms that look for driver mutations [7]. In addition, mutual exclusivity *does not affect different pathways*, it is a phenomenon that happens only *within* a single pathway. An explanation for this occurrence is not known yet, but the following hypotheses seem promising [9, 10]:

- one idea is that mutually exclusive genes are functionally connected within a common pathway, acting on the same downstream effectors and creating functional redundancy; consequently, they would share the same selective advantage, meaning that the alteration of one mutually exclusive gene would be sufficient to disrupt their shared pathway, removing the selective pressure to alter the others; this explanation is yet not able to justify the phenomenon completely because if this were true, co-alteration of mutually exclusive genes should not lead to any negative effects on the cell;
- a different explanation could be that the co-occurrence of mutually exclusive

alterations is detrimental for cancer survival, thus cells which harbor said co-occurrences would be eliminated from the population; moreover, some mutually exclusive gene pair could be *synthetic lethal*, meaning that the alteration of only one of the genes would be compatible with cell survival but the aberration of both would not .

add example from survey paper?; also, use example? (mail "Risposte (parziali) alle questioni, ERG e SPOP")

In addition, another important property of driver pathways is **coverage**, i.e. driver genes that describe a driver pathway are frequently mutated in many samples. Thus, *a driver pathway is made up of genes mutated in many patients, and its mutations are approximately mutually exclusive between them.* It is also known that pathways having the said characteristics are generally short and made up of few genes on average [7].

1.3.3 *De novo* and *knowledge-based* approaches

Despite not knowing the true explanation for mutual exclusivity yet, and not knowing whether it holds therapeutical potential, it is a very recurrently observed behavior in data, and many believe that it may lead to some discoveries for cancer treatment. The existing approaches can be categorized into two types: *de novo* approaches, which perform the research of mutually exclusive patterns by using only genomic data from patients, and *knowledge-based* methods, which integrate the analysis with external *a priori* information [9]. Evidently, *de novo* approaches may not have enough information because they don't draw on existing databases; on the counter side, the current understanding of gene and protein interactions in humans is still incomplete, and most existing pathway databases fail to accurately represent the specific pathways and interactions present in cancer cells. As a result, *de novo* methods may find new but not necessarily accurate results, while *knowledge-based* approaches' excessive focus on known data sources may constrain opportunities for uncovering new biological insights [7].

Acknowledgements

TODO

Bibliography

[1] Bert Vogelstein and Kenneth W Kinzler. "Cancer genes and the pathways they control". In: *Nature Medicine* 10.8 (July 2004), 789–799. ISSN: 1546-170X. DOI: 10.1038/nm1087. URL: http://dx.doi.org/10.1038/nm1087.

[2] *Cancro: la cura*. URL: https://www.airc.it/cancro/affronta-la-malattia/guida-alle-terapie/cancro-la-cura.

[3] *Targeted therapy for cancer*. May 2022. URL: https://www.cancer.gov/about-cancer/treatment/types/targeted-therapies.

[4] Cleveland Clinic Medical Professional. *Targeted therapy*. May 2024. URL: https://my.clevelandclinic.org/health/treatments/22733-targeted-therapy.

[5] Wikipedia contributors. *Cell signaling*. Aug. 2024. URL: https://en.wikipedia.org/wiki/Cell_signaling.

[6] Nhgri. *Biological Pathways Fact sheet*. Mar. 2019. URL: https://www.genome.gov/about-genomics/fact-sheets/Biological-Pathways-Fact-Sheet.

[7] Mark D. M. Leiserson, Dima Blokh, Roded Sharan, et al. "Simultaneous Identification of Multiple Driver Pathways in Cancer". In: *PLoS Computational Biology* 9.5 (May 2013). Ed. by Niko Beerenwinkel, e1003054. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1003054. URL: http://dx.doi.org/10.1371/journal.pcbi.1003054.

[8] Junfei Zhao, Shihua Zhang, Ling-Yun Wu, et al. "Efficient methods for identifying mutated driver pathways in cancer". In: *Bioinformatics* 28.22 (Sept. 2012), 2940–2947. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/bts564. URL: http://dx.doi.org/10.1093/bioinformatics/bts564.

[9] Yulan Deng, Shangyi Luo, Chunyu Deng, et al. "Identifying mutual exclusivity across cancer genomes: computational approaches to discover genetic interaction and reveal tumor vulnerability". In: *Briefings in Bioinformatics* 20.1 (Aug. 2017), 254–266. ISSN: 1477-4054. DOI: 10.1093/bib/bbx109. URL: http://dx.doi.org/10.1093/bib/bbx109.

[10] Jaroslaw Cisowski and Martin O. Bergo. "What makes oncogenes mutually exclusive?" In: *Small GTPases* 8.3 (July 2016), 187–192. ISSN: 2154-1256. DOI: 10.1080/21541248.2016.1212689. URL: http://dx.doi.org/10.1080/21541248.2016.1212689.