Indian Institute of Technology Jodhpur

Design and development of an app for cancer outcome prediction

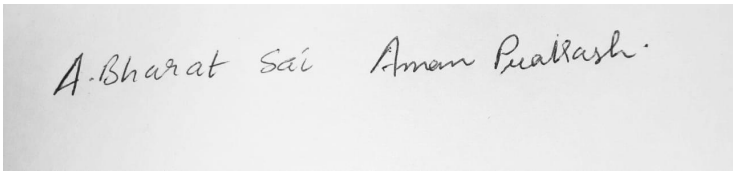
A B.Tech Project Report submitted by

Addepalli Bharat Sai (B19BB002)
Aman Prakash (B19BB005)

Under the Supervision of

Dr. Siddharth Srivastava
Dr. Sushmitha Jha
Department of BSBE
Dr. Sumit Kalra
Department of CSE

On
18/11/2022



Signature of Students

Contents

Introduction

Cancer is when some body cells grow out of control and spread to other areas. Cancer can develop practically anywhere in the billions of cells that make up the human body. Human cells often divide among themselves to create new cells as the body requires them. New cells replace old ones when they die due to aging or damage.

Since genes that determine how our cells behave, mainly how they grow and divide, are altered, cancer is a genetic disease.

Genetic changes that cause cancer can happen because

1. of errors that occur as cells divide.
2. DNA deterioration is brought on by unfavorable environmental elements, including tobacco smoke toxins and the sun's UV radiation.
3. They were inherited from our parents.

Precision medicine is "an emerging strategy for illness treatment and prevention that incorporates individual variability in genes, environment, and lifestyle for each person," according to the Precision Medicine Initiative. With the help of this method, medical professionals and researchers will be able to anticipate with more accuracy which disease-specific treatments and preventative measures will be effective in different populations. The one-size-fits-all strategy, in contrast, develops illness treatment and prevention measures for the average person with little regard for individual differences.

Although the term "precision medicine" is relatively new, the idea has long been present in healthcare. To limit the danger of complications, a person who needs a blood transfusion, for instance, is not given blood from a donor chosen at random but rather from a donor whose blood type matches that of the recipient. Although there are numerous medical fields where precision medicine is used, its application to routine healthcare is still somewhat limited. In the upcoming years, researchers anticipate that this strategy will be applied to numerous healthcare and wellness fields.

Normal or tumor cells that proliferate quickly are killed by traditional chemotherapy. Targeted therapy is a type of precision medicine. It functions by halting or reducing the progression or growth of cancer. This takes place at the cellular level. For cancer cells to live, grow, and spread, particular chemicals, frequently proteins, are required. The cancer-causing genes and the cells themselves typically produce these chemicals.

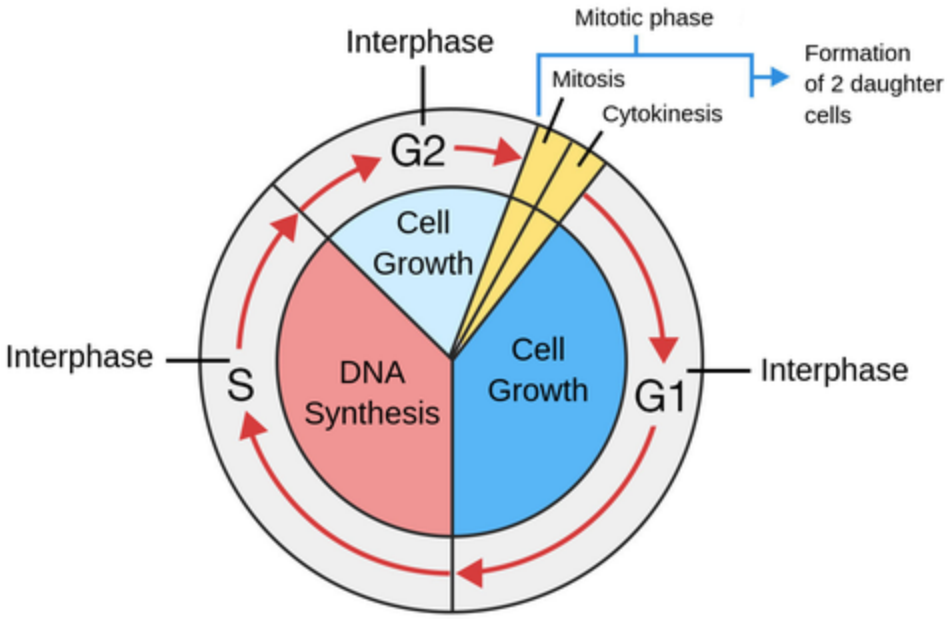


Figure 1 - Cell-division Cycle

These chemicals or the cancer-causing genes that produce them have targeted therapy targets. The medicine may occasionally bind to the molecule intended to affect, preventing it from working. Sometimes the medication will stop the molecule from going where it usually does. Cancer can be prevented or slowed down in its progression by preventing these molecules from performing their normal functions.

To check for blood or tumor DNA, doctors may use genomic tests. This can reveal which genes are altered or have extra copies, as well as whether any abnormalities of the genes can be treated.

Even when a patient possesses a targetable molecule, the medicine may occasionally stop working after a while. This typically happens when cancer discovers a different way to complete the task that the targeted therapy is intended to halt. Repeated genetic analysis can occasionally reveal how the tumor deceived the medication.

Targeted therapies can be administered orally or intravenously and combined with other medicines like chemotherapy or immunotherapy.

There are two main types of targeted therapy drugs:

1. Small molecule drugs can quickly enter the cells and interfere with the molecules inside. They can also be used to interfere with molecules on the surface of the cell.

2. More substantial and influential outside of cancer cells is monoclonal antibodies. They go after nearby chemicals or on the cancer cell's surface. These are produced by cloned cells that make antibodies that block the desired chemical. Additionally, a harmful substance can be delivered by monoclonal antibodies right to a cancer cell.

Also, the net has some excellent information about everything about cancer available. Which source to believe or not is still a challenge to till date. Cancer is ongoing research every day. A new treatment, or precision medicine, is suggested. The doctor or the health care professional should be up to date to give the best suitable and beneficial treatment to the patient.

Also, many doctors are still learning precision medicine, which is still an area of development. Many will not even want to adapt to the new treatments and still want to pursue their traditional treatment. So, we have to give them ample examples of Precision medicine being preferred over traditional chemotherapy or radiation therapy, which will cause more harm to an individual than the benefits provided.

Hence, we took it as a challenge to consider how to reduce the gap between ongoing research and available treatments very precisely for specific cancer-inflicted individuals. We want to create awareness that everyone should be on the same page. The patients can also access the app and discuss a different approach to the treatments with their healthcare professionals.

Problem Statement

Design and development of an app for cancer outcome prediction and able to suggest different available treatments specific to the pathway the cancer is adapted to proliferate and survive the cell checkpoints.
For this, we have to consider that we all are made of simple A, T, C, and G; any undesirable changes in these are called mutations, thereby giving them an advantage for these abnormal cells to proliferate and invade all our normal bodily functions.

We also know about the process of the central dogma, which is a crucial process of making protein from the genetic material available. The central dogma of molecular biology states that genetic information flows only in one direction, from DNA to RNA, to protein, or RNA directly to protein. Any changes in the DNA may lead to the making of abnormal proteins, creating all the abnormalities associated with the cell and leading to tumors.

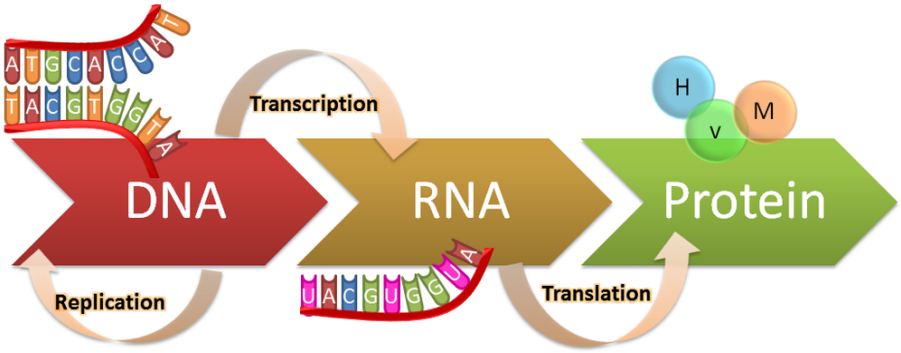


Figure 2 - Central Dogma

Hence, we should search for these abnormal proteins around the cell to detect cancer. These can also be called biomarkers. Testing for biomarkers is a means to find genes, proteins, and other components (also known as tumor markers or biomarkers) that can reveal cancer information. A distinct pattern of biomarkers characterizes each person's cancer. Some biomarkers influence the efficacy of specific cancer therapies. You and your doctor may choose a cancer treatment with biomarker testing.

Now coming to the main question, what are the biomarkers to look for? Every cancer has different biomarkers. We, humans, have between 20,000 and 25,000 genes. What to search for and what not to look for is essential.

Materials and Methods

Initially, we needed guidance about where to begin and how to proceed. We have drawn a Figma on how our app's basic structure will be presented. It was our initial work, we changed our design various times, but the basic structure was the same. When we started to develop the app, we felt some internal decisions needed to make, so we modified the design according to our requirements. link

The whole genome can be considered a time and resource waste. So, we need some limited biomarkers to check. For this purpose, we can use the KEGG organisms database [1] for metabolic pathways. The KEGG pathway representation focuses on the network of gene products, mostly proteins, including functional RNAs.

For Example, let us see the pancreatic cancer pathway of cancer progression. It looks like a detailed or complicated description of every gene and protein involved in pancreatic cancer-occurring cells.
We can observe from this pathway that mainly specific genes are responsible for cancer. Ongoing research also tells us about what genes to look at.

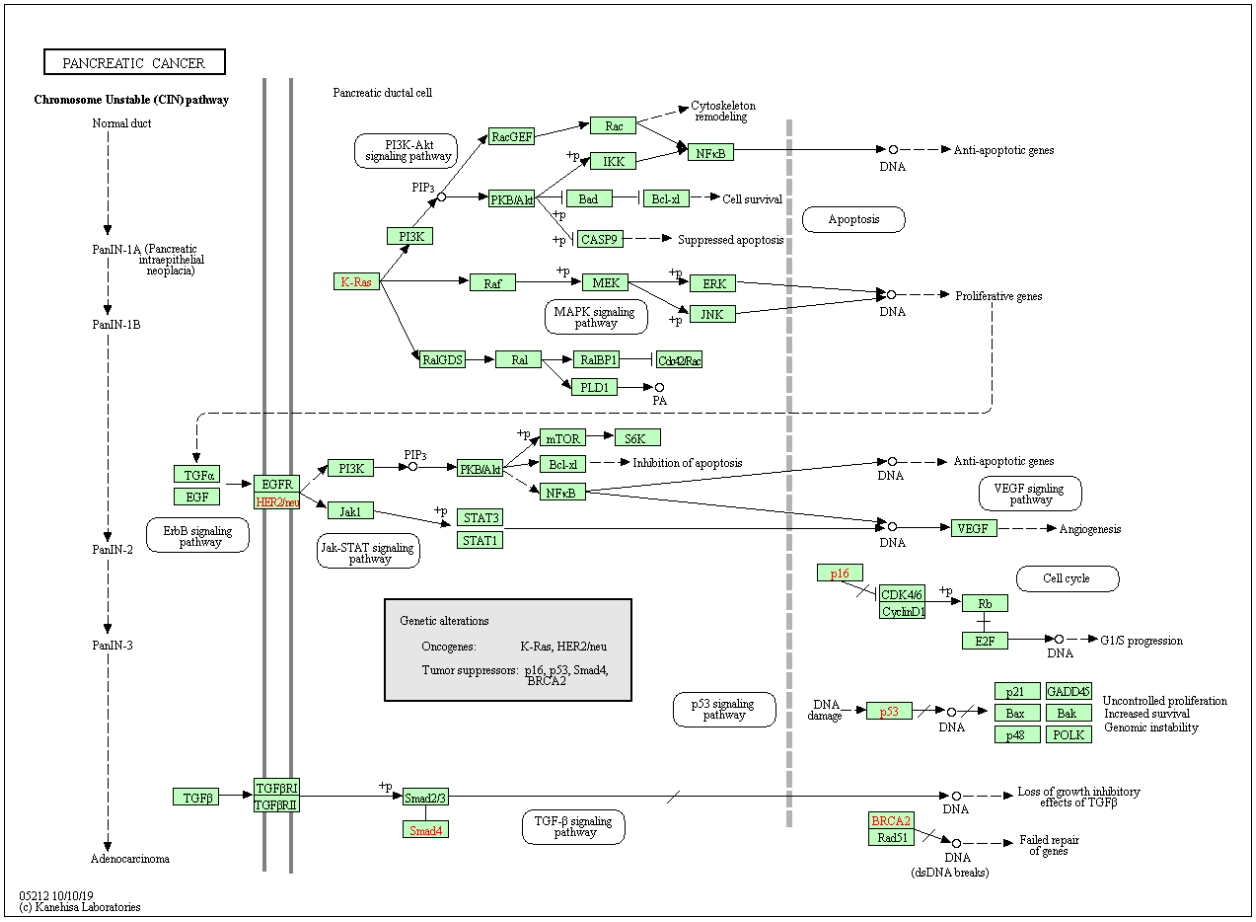


Figure 3 - Kegg Pathway Pancreatic Cancer

Now, to get this information, we explored relevant criteria. We learned about the National Comprehensive Cancer Network [2], a non-profit organization that deals explicitly with cancer awareness and is proven to provide the best available guidelines for patients and healthcare professionals to provide the best treatment.

Tumor/somatic molecular profiling is recommended for patients with locally advanced/metastatic disease who are candidates for anti-cancer therapy to identify uncommon mutations. Consider specifically testing for potentially actionable somatic findings including, but not limited to: fusions (*ALK, NRG1, NTRK, ROS1, FGFR2, RET*), mutations (*BRAF, BRCA1/2, KRAS, PALB2*), amplifications (*HER2*), microsatellite instability (MSI), and/or mismatch repair (MMR) deficiency.

Figure 4 - FromNCCN, ESMO, AICR, and Some Research Papers

According to NCCN, ESMO, AICR, and Some Research Papers, the biomarkers for pancreatic adenocarcinoma is to look the fusions (ALK, NRG1, NTRK, ROS1, FGFR2, RET), mutations (BRAF, BRCA1/2, KRAS, PALB2), amplifications (HER2) and microsatellite instability (MSI), or mismatch repair (MMR) deficiency.

Fusion - A gene created by combining pieces from two distinct genes. Fusion genes and the resulting fusion proteins can be created in a lab or arise naturally in the body when a portion of one chromosome travels to another. Some types of cancer may emerge due to the fusion proteins created by this alteration. For instance, certain kinds of leukemia have the BCR-ABL fusion gene and protein.
Several additional malignancies, such as soft tissue sarcoma, prostate, breast, lung, bladder, colon, rectum cancers, and CNS tumors, may also contain fusion genes and proteins. Cancer detection and treatment are being explored with fusion genes and proteins.

Mutation - Any alteration to a cell's DNA sequence. Errors in cell division can result in mutations, as can exposure to environmental DNA-damaging substances. Mutations may be harmful, beneficial, or ineffective. Mutations in other types of cells are not passed on to offspring; however, if they do so in the cells that produce eggs or sperm, they may be. Cancer and other diseases may result from specific mutations.

Gene Amplifications - A rise in the gene's copy number.
Additionally, there can be an increase in the RNA and protein produced by that gene.
In cancer cells, gene amplification is prevalent, and some amplified genes may cause cancer cells to proliferate or develop therapeutic resistance. Genes can also be increased in number in a lab setting for research.

MSI - A change occurs in specific cells (such as cancer cells) in which the number of repeated DNA bases in a microsatellite (a short, repeated DNA sequence) differs from when the microsatellite was inherited. Microsatellite instability may be caused by mistakes not corrected when DNA is copied in a cell. It is often found in colorectal, gastric, and endometrial cancer, but it may also be found in many other types of cancer. Knowing whether cancer has microsatellite instability may help plan the best treatment. Also called MSI.

MMR - Describes cells that have mutations in specific genes that correct mistakes made when DNA is copied in a cell. Mismatch repair (MMR) deficient cells usually have many DNA mutations, which may lead to cancer. MMR deficiency is most common in colorectal, other gastrointestinal, and endometrial cancers. However, it may also be found in breast, prostate, bladder, and thyroid cancers.

MMR deficiency may also be found in an inherited disorder called Lynch syndrome. Knowing if a tumor is MMR deficient may help plan treatment or predict how well the tumor will respond to treatment.

So, with these available genes and types of alterations, we explored cbioportal.org [3], which is a Cancer Genomics portal that provides visualization, analysis, and download of large-scale cancer genomics data sets. We took all the Pan-Cancer studies and queried with the suggested genes from the NCCN guidelines. Then in the mutations sections, we can find the most likely or known oncogenic and suggest the OncoKB Therapeutic Levels of Evidence [4] targeted therapies available.

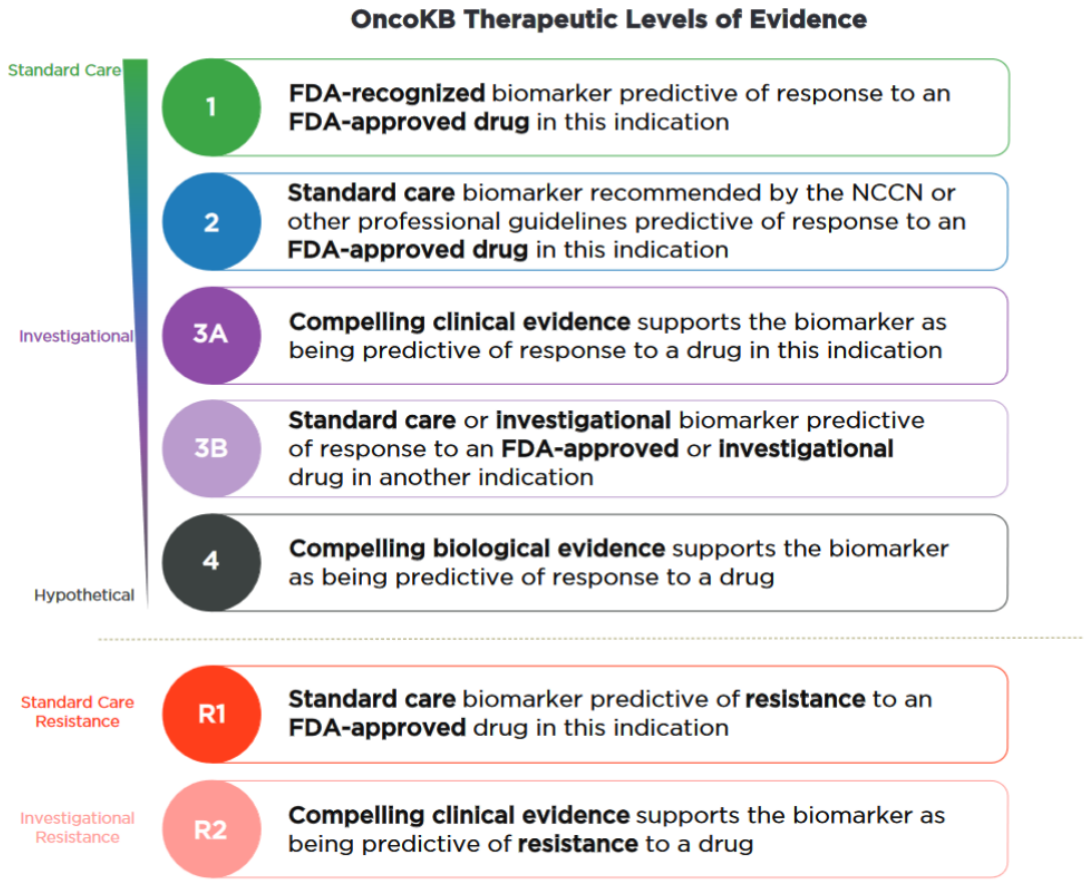


Figure 5 - OncoKB Therapeutic Levels of Evidence

Hence, this completes the circle of our cancer application model. Now, let us see what we can expect in our app.

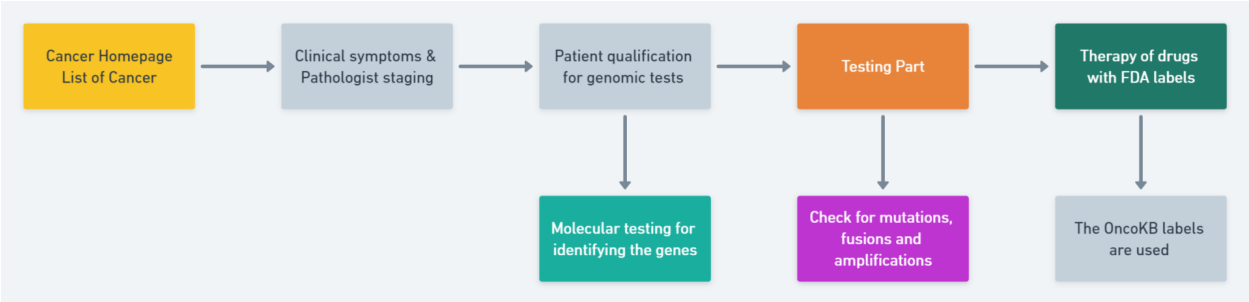


Figure 6 - Whimsical Flowchart of the app - [link]

Results and Discussion

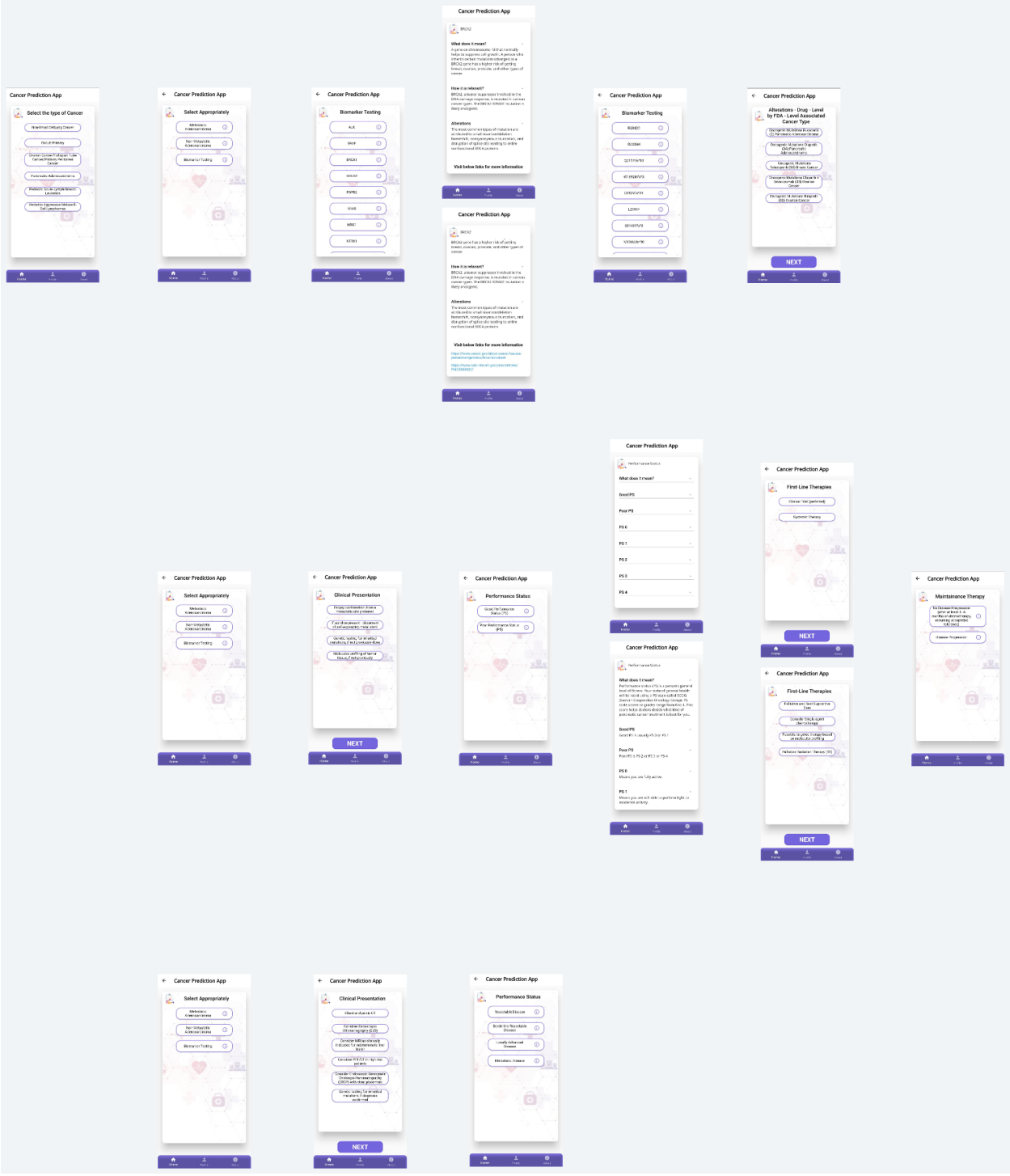


Figure 7 - Whimsical, the flow of app - [link]

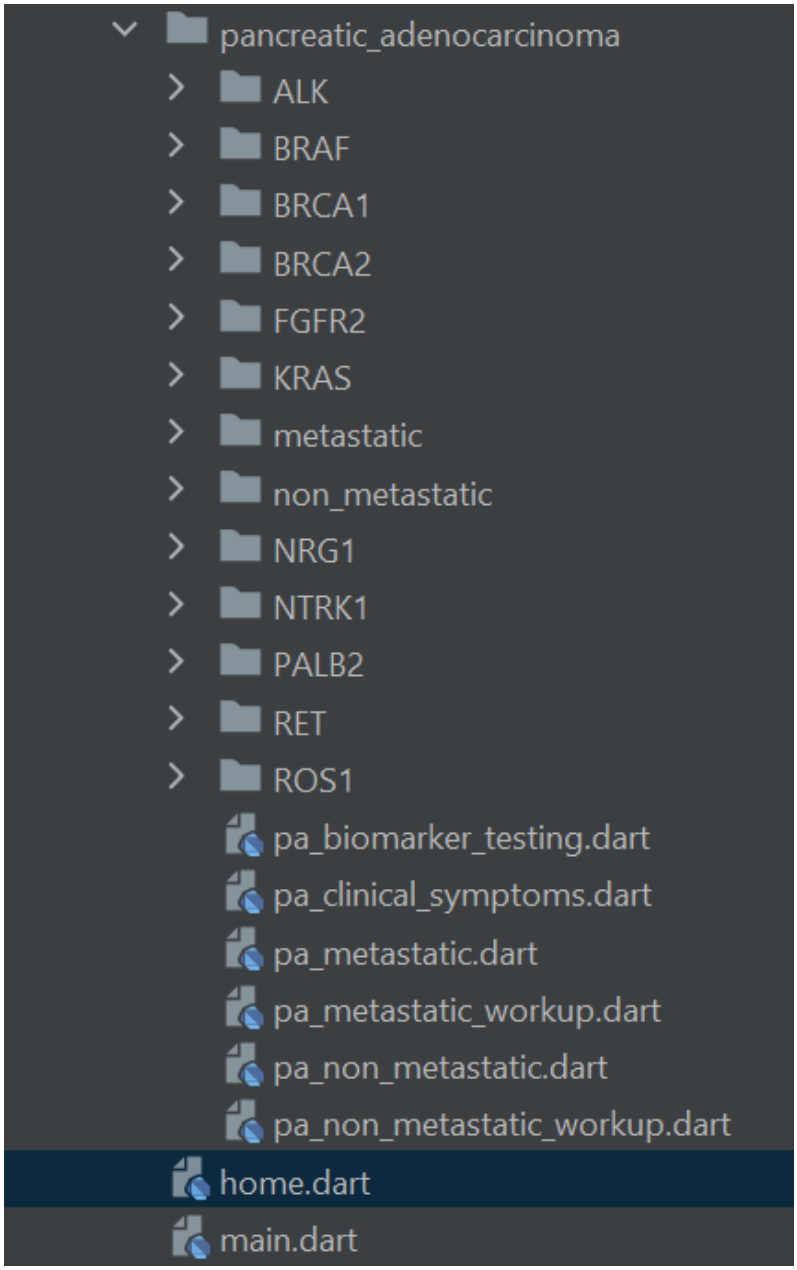


Figure 8 - App Code structure - Github link

Figure - 6 shows us the flow of our app in the case of pancreatic adenocarcinoma, whether the individual clinical prognosis was metastatic or non-metastatic or the patient is willing to opt for any targeting medicine for their benefit.

Metastatic, then further decisions must be made by the health care professions based on the patient's performance status and suggest their respective treatments.

Non-metastatic, on the other hand, specifically for pancreatic adenocarcinoma, is further subclassified based on resectable, borderline-resectable, locally advanced, or metastatic and suggests the health care professional their respective available options on how to proceed or suggest some available treatments.

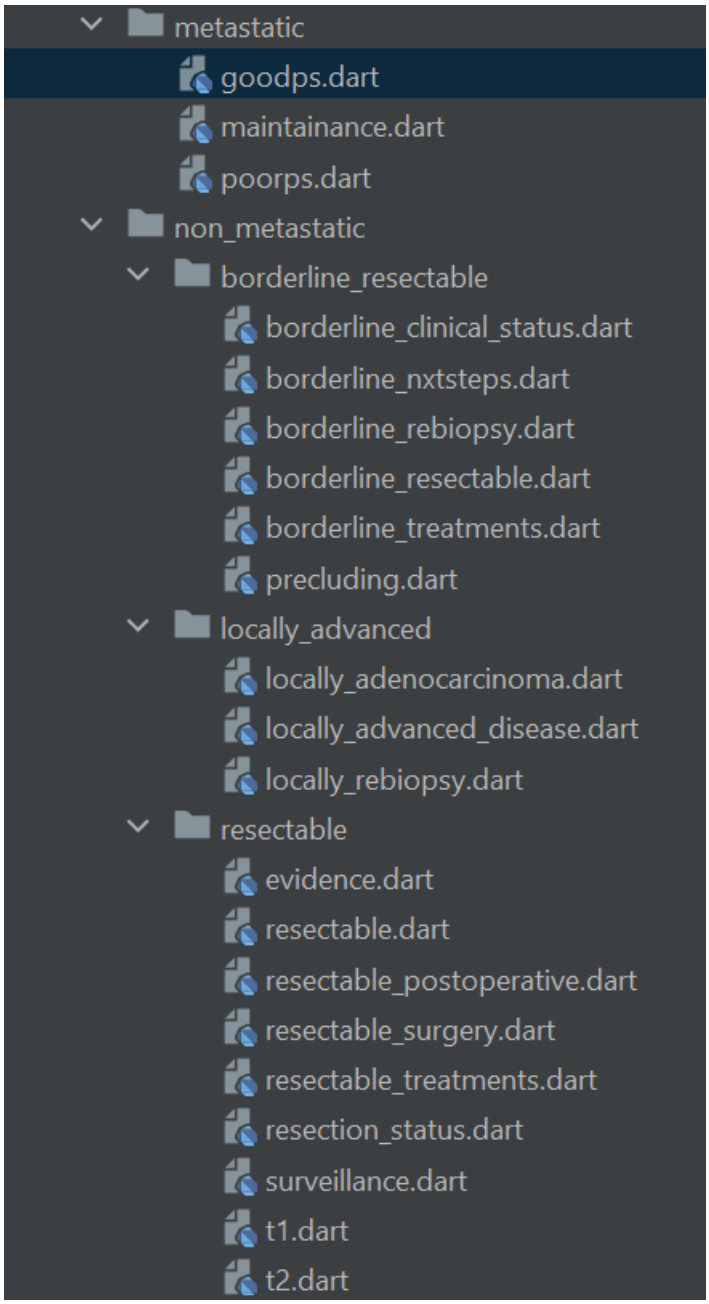


Figure 9 - App Code structure - Github link

Biomarker testing lists various genes involved in tumor formation. The next page lists their respective fusions or mutations, or amplification. It suggests the respective available and

working treatments which are tested or are being tested by the clinical trials professionals and suggests the patients accordingly.

The listing above is so simple to read, but all these were not even available 50 years back. There is thorough research going on as we speak that will benefit future generations. Modernizing the available information (long pdf) in the app (simple so that everyone can access it quickly) is a very crucial step we are taking.

Conclusion and Future Aspects

Our app will have a wide range of applications in the healthcare industry when made fully available to the public. However, we have worked only on single cancer, namely pancreatic cancer, which should not limit us from recognizing the importance of this soon. It is just a brick in building the bridge. Our other colleagues have worked on NSCLC. Our design model will benefit many when all cancer models are ready. However, as stated before, cancer is ongoing research. Currently, standard new treatment and diagnostic biomarkers will be updated, so there is a requirement for a team of individuals to collect this information and update the app, respectively.

Our app will help the healthcare community by being a one-stop solution for many doctors who have to do much research to come up with the proper biomarker testing and the right therapeutic drugs to suggest.

Also, the cbioportal data majorly originated from the American population, which we adapt as per our requirement, however in India, such initiatives to collect the data and give them some basic functionalities to analyze the data are required more than ever, with this information and reports might benefit some other person.

References

[1] *Pancreatic cancer - Homo Sapiens (human). KEGG PATHWAY: Pancreatic cancer - Homo sapiens (human)*. Available at: https://www.genome.jp/pathway/hsa05212 (Accessed: November 15, 2022).

[2] *Pancreatic Adenocarcinoma - National Comprehensive Cancer Network* (*NCCN*). Available at: https://www.nccn.org/professionals/physician_gls/pdf/pancreatic.pdf (Accessed: November 15, 2022).

[3] Cerami et al. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. Cancer Discovery. May 2012 2; 401. PubMed. *cBioPortal* Available at: https://www.cbioportal.org/ (Accessed: November 15, 2022).

[4] *OncoKB Therapeutic Levels of Evidence*. Available at: https://www.oncokb.org/content/files/levelOfEvidence/V2/LevelsOfEvidence.pdf (Accessed: November 15, 2022).

[5] *Work better, faster, together. Whimsical*. Available at: https://whimsical.com/ (Accessed: November 15, 2022).

[6] *NCI Dictionary of Cancer Terms.* Available at: https://www.cancer.gov/publications/dictionaries/cancer-terms/ (Accessed: November 15, 2022).

[7] *European Society for Medical Oncology.* Available at: https://www.esmo.org/ (Accessed: November 15, 2022).

[8] *American Institute for Cancer Research.* Available at: https://www.aicr.org/ (Accessed: November 15, 2022).

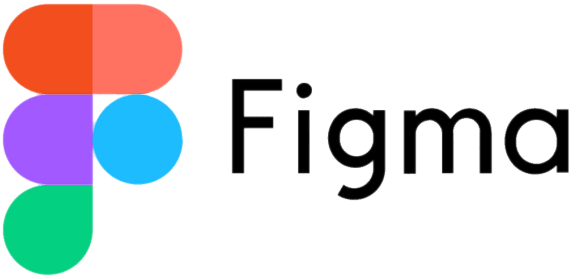
Appendix

1) Technologies Used,


cBioPortal
for Cancer Genomics


NCCN
National
Comprehensive
Cancer
Network®


Whimsical


Figma