The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM)
ISSN: 2602-3199

The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 2024

Volume 28, Pages 34-46

ICBASET 2024: International Conference on Basic Sciences, Engineering and Technology

Complex Thinking Abilities in the Rapidly Evolving Field of Genomics and Personalized Medicine: Analysis of Actionability on Cancer with ChatGPT and Literature

M.S. Ramírez-Montoya
EGADE Business School

G.A. Sanchez-Zuno
Yale University

R.M. Marques González
Hospital Civil de Guadalajara

F. Casıllas-Muñoz
Tecnologico de Monterrey

Abstract: This study aims to explore complex thinking abilities within the field of genomics and personalized medicine, focusing on the analysis of actionable cancer data using Big Data Analytics (BDA) and ChatGPT. It seeks to understand how these advanced technologies can be harnessed to derive more actionable approaches for students and professionals in genetics, biology, medicine, and related fields. Methods: The research methodology involves a machine learning (ML) analysis to visualize the distribution of genes based on top ten actionability counts, development status, and drug combinations. This includes ChatGPT prompts for visualization of gene distribution and the use of pivot tables for data validation. The study facilitates complex data analysis and decision-making processes in genomics. The findings reveal that BDA and ChatGPT can significantly improve the analysis and interpretation of genomic data. Visualization techniques enabled by these technologies allow for the identification of patterns, correlations, and predictive models. These insights can lead to more accurate diagnoses, personalized treatment plans, and a better understanding of drug combinations and mutations in cancer. This research highlights the essential role of automation and open access in managing and interpreting large volumes of genomic data efficiently. Conclusion: The integration of BDA and ChatGPT into genomics and personalized medicine offers promising avenues for advancing personalized medicine, enhancing clinical decision-making, and fostering research and development in the field of cancer.

Keywords: Genomics, Cancer, Actionable mutations, ChatGPT in medicine

Introduction

Addressing actionable information in an educational context, especially for educators in genetics, biology, medicine, and related fields, involves a multifaceted strategy. These strategies involve embracing the emergence of big data to support informed decision-making (Fischer et al., 2020). The rapidly evolving field of big data, particularly within the medical sciences, has announced a transformative era for education with the urgent necessity to tailor educational experiences to fit the specific needs of individual learners and for the advancement of technologies (Luan et al., 2020). In this sense, Big data is revolutionizing bioinformatics education by necessitating curriculum updates to incorporate skills for managing and analyzing large datasets,

thus preparing students for the data-intensive challenges in biomedical fields (Greene et al., 2016). By facing these challenges, medical institutions will ensure students and professionals are well-equipped for advancements in biomedical research and healthcare.

Complex and Critical Thinking is a Necessity that Goes Far Beyond the Lack of Specialization

Some strategies educators might consider stimulating complex thinking to derivate more information into more actionable approaches in students; Big data analytics (BDA) is enabling complex thinking analytics, with the help of data analysis decomposition to facilitate more accurate diagnoses and personalized treatments (Batko & Ślęzak, 2022). Artificial Intelligence (AI) such as ChatGPT can complement these analytics; notably cannot be used reliably for clinical applications, but with further specialized training and validation, it can support enormously medical contexts to offer reliable and fine-tuned data for such purposes (Li et al., 2024). In this sense, ChatGPT and other large language models (LLMs) are also enhancing drug discoveries, accelerating target identification and optimizing drug design through predictive analytics on pharmacodynamics, pharmacokinetics, and toxicity significantly reducing time and costs associated with bringing new drugs to market (Chakraborty et al., 2023). This evidence reflects the importance of leverage in these technologies for big data interpretation, especially in the biomedical sciences.

The complexity of medical data arises from numerous sources and types, each with unique traits that influence how data is analyzed and applied. This complexity necessitates advanced analytical strategies to effectively understand and utilize the data, ensuring that actionable insights are both accurate and relevant to patient care (Lee & Yoon, 2017).

Identifying actionable somatic mutations (ASM) in cancer involves understanding specific genetic alterations that can guide targeted therapy (they directly influence treatment decisions that is why they are actionable), for example, in non-small cell lung cancer (NSCLC), the presence of specific mutations in the EGFR gene can dictate the effectiveness of EGFR inhibitors like gefitinib and erlotinib (Puiu et al., 2024). The importance of AI in predicting ASM in cancer is highlighted by its ability to accelerate the identification of genetic alterations that can inform targeted therapy, and improve patient outcomes, for example, computational models to pinpoint mutations like the V600E mutation in the BRAF gene can predict responsiveness to specific inhibitors such as vemurafenib (Ostroverkhova et al., 2023). However, elucidating ASM in cancer is another complex challenge due to the immense genetic diversity and heterogeneity of tumors and these complex data requires sophisticated computational tools and deep biological understanding to effectively translate genomic information into clinically actionable treatment plans (Ahmed, 2023). The strategies to cope with these complexities with the help of AI will define the rate of medical advancements.

Actionability in Medicine Based on Big Data: Misleading Opportunities

Paul Wesson and colleagues in 2022 mention the "six Vs" algorithm, which details the main elements and keys related to the concept of big data. These elements are important because they allow organizations to harness the power of data for decision-making, operational efficiency, and equity in various areas (Wesson et al., 2022). One of the most well-known and highly rated actionable programs for big data management is COSMIC, which is used by most precision oncology specialists. This database provides information on mutated sequences observed in cancer patients, as well as possible associations with other events, thus becoming a reliable guide for research evidence, potentially reducing efforts and costs, and setting a standard in cancer genomics (Tempini and Leonelli (2021). Most databases store data inefficiently and only for the minimally required retention period, this is one of the main misleading opportunities for big data analysis, as databases are considered mere data dumps. To prevent this, the essential aspects that data must have to be adequately usable in research are summarized with the acronym FAIR: findable, accessible, interoperable, and reusable (Blatter et al., 2022). One of the great current uses of these big data analyses allows the recognition of patterns in specialized metabolic diagnostics, as well as technique and medical validation, and quality management.

The availability of multimodal datasets could improve our ability to find phenotypic and genotypic disease characteristics, helping to predict dose-response, risk prediction, prognosis, treatment response, and patient outcomes. Recently, new AI/ML algorithms for precision medicine for neurological diseases, cancer, and cardiovascular diseases have shown promising results regarding disease risk prediction, phenotypic prediction, dose-treatment response, with a high degree of precision and accuracy (Blatter et al., 2022). AI, big data, and ML technologies offer alternative pathways, either individually or in combination, that facilitate a better

understanding and management of cancer and its impact on patients. The value of these technologies has been recognized and integrated into programs such as the Cancer Moonshot in the United States and the European Beating Cancer Plan in Europe (Sahu et al., 2022), their integration into clinical practice also poses new requirements for healthcare professionals in terms of skills and preparedness to use them effectively and efficiently. Therefore, there is a growing need for investment and training in oncology to address and overcome some of the challenges associated with cancer control. In Spain, several projects have begun developing tools based on massive data. In Girona, the Savana project is underway, a platform supporting medical decision-making that is already being implemented in some Spanish healthcare centers. This platform is designed to transform Electronic Health Records into large datasets, using Artificial Intelligence (AI) to unlock the clinical value contained in unstructured information. This is bringing a disruptive change in healthcare by enhancing efficiency in clinical management and research.

Method

The methods in these explore complex thinking abilities in the field of genomics and personalized medicine through the analysis of actionable cancer data involve a multifaceted approach integrating Big Data Analytics (BDA) and ChatGPT:

Data Collection and Preprocessing

The COSMIC database, standing for the Catalogue Of Somatic Mutations In Cancer, is an expert-curated platform that compiles a wide variety of somatic mutation mechanisms driving human cancer. The data is publicly available and it compilates different information about cancer genomics. It was accessed to the repository of actionable mutations though this link: https://cancer.sanger.ac.uk/cosmic/download/cosmic, which contains complete actionable data in .tar format. This dataset was transformed into .csv. and we gather genomic data related to genes related with actionability counts, development statuses, drug combinations and other relevant data about drug candidates according to genomic information.

Enhanced Data Visualization Analysis Using ChatGPT

Complex prompts were inputted into ChatGPT to produce sophisticated data visualizations. These prompts were designed to refine the analysis based on initial observations in the spectrum of actionable variables and were introduced to explore deeper insights, particularly focusing on drug combinations, genomic mutations, actionability rank, and development status. The process included asking increasingly complex prompts to ChatGPT for visualization techniques and specific analysis queries. Utilizing Python, data visualizations were validated with pivot tables to explore the distribution patterns within the dataset. Pivot tables resembled the frequencies and relationship in matrices regarding the association of the different variables. This step was crucial to ensure that the insights obtained are accurate or reflective (with approximation) of the current scientific understanding and scientific evidence.

Results

Process of Data Visualization and Validation:

A simple prompt in chatGPT such as this “visualize the distribution of genes based on the top ten actionability counts, development status, and drug combinations” can help to visualize the distribution based on actionability rank, development status, and drug combinations, output a plot like the following:

This information was validated in python :

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

data_full = pd.read_csv(' /content/Actionability_AllData_v11_GRCh37.csv')
```

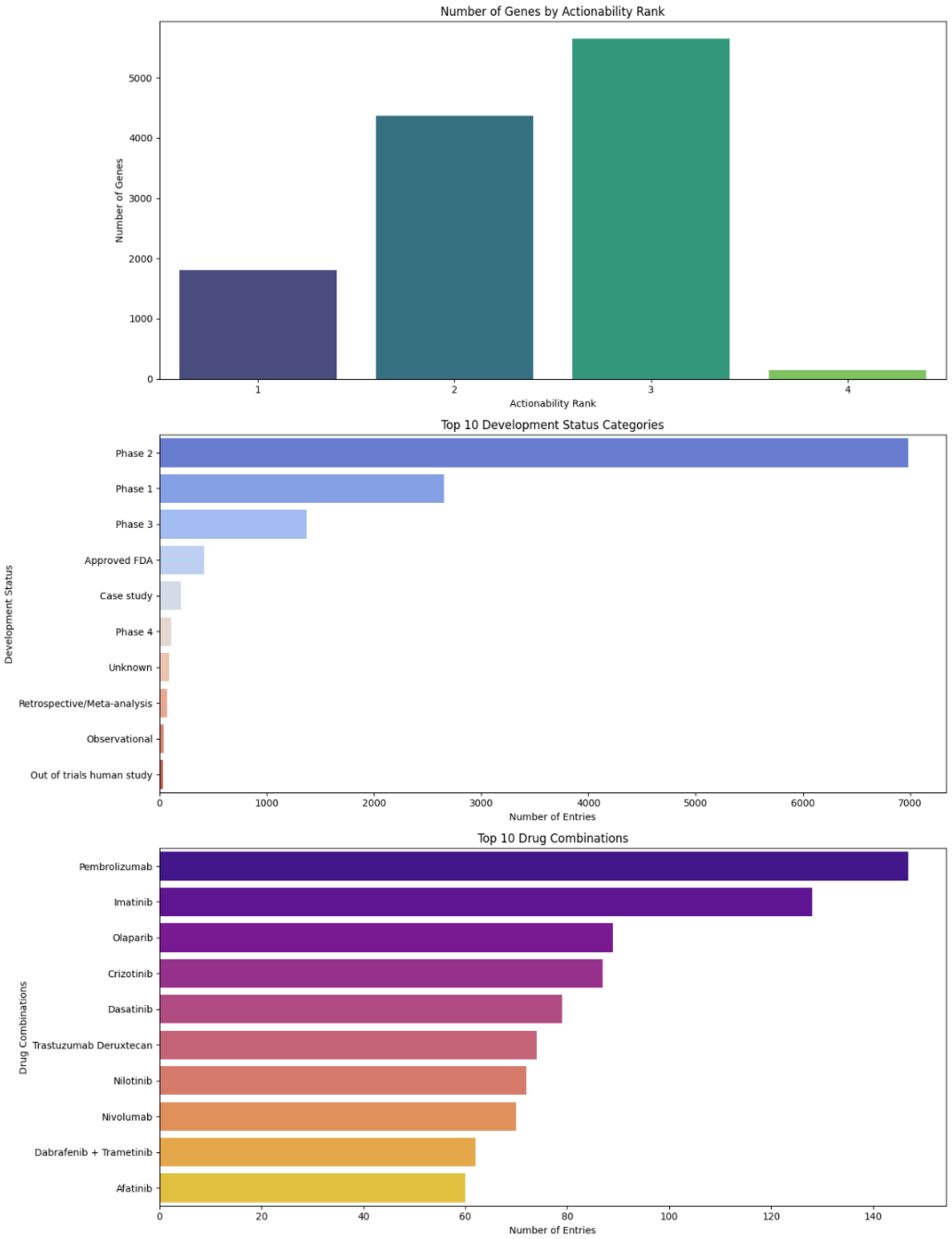

Figure 1.A. Number of drug combinations by actionability rank and development status: Actionability rank: 1: approved, 2: marketed drugs, 3: through clinical trial phases 4: case studies.

```
# Counts for 'Actionability Rank'
actionability_counts = data_full['ACTIONABILITY_RANK'].value_counts().nlargest(10)

# Counts for 'Development Status'
development_status_counts =data_full['DEVELOPMENT_STATUS'].value_counts().nlargest(10)

# Counts for 'Drug Combinations'
drug_combination_counts =data_full['DRUG_COMBINATION'].value_counts().nlargest(10)
actionability_table = actionability_counts.reset_index().rename(columns={'index': 'Actionability Rank',
'ACTIONABILITY_RANK': 'Frequency'})
development_status_table =
Same code for 'DEVELOPMENT_STATUS and 'DRUG_COMBINATION':
```

Table 1. Validation of number of drug combinations by actionability rank and development status

Actionability Rank	**Frequency**
3	5649
2	4368
1	1808
4	147
Development Status	**Frequency**
Phase 2	6986
Phase 1	2653
Phase 3	1371
Approved FDA	415
Case study	197
Phase 4	109
Unknown	92
Retrospective/Meta-analysis	71
Observational	40
Out of trials human study	33
Drug Combinations	**Frequency**
Pembrolizumab	147
Imatinib	128
Olaparib	89
Crizotinib	87
Dasatinib	79
Trastuzumab Deruxtecan	74
Nilotinib	72
Nivolumab	70
Dabrafenib + Trametinib	62
Afatinib	60

In the above Table 1, the actionability rank (also seen in the upper histogram illustrated in Figure 1) is divided in four stages that allow to identify the most advanced development stage reached by drugs that have been tested in patients with the specified mutation and disease. The values range from approved, marketed drugs (most advanced) through clinical trial phases to case studies (COSMIC, 2024). There may be multiple drugs at the same rank of development; individual drugs are not ranked. It can be seen that most of the drugs are positioned in the 3 rank.

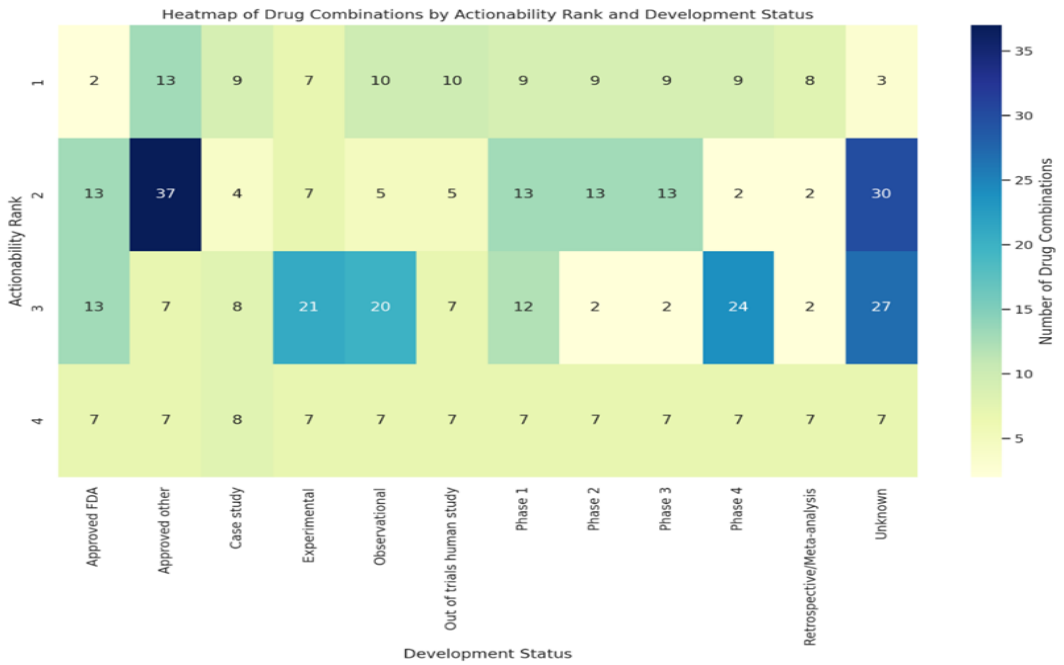

Figure 2. Heatmap of drug combinations by actionability rank and development status: Actionability rank: 1: approved, 2: marketed drugs, 3: through clinical trial phases 4: case studies.

The development status can also be observed (middle figure), most of developments are classified as phase 2 (with 6986 entries), this phase means that these studies are determining the effectiveness of an experimental drug on a particular disease or condition in approximately 100 to 300 volunteers. Finally, In the graph at the bottom it can be seen that Pembrolizumab is the drug with the greatest number of combinations in trials. Once identified these elements, the complexity of the analysis was enhanced by the following prompt: "Visualize the combination of actionability rank and development status, focusing on common drug combinations.

To validate the above table, we perform an approach to count the total number of drug combinations and obtain a pivot table of combinations in python:

```
pivot_table_combination_counts = data_filtered.groupby(['ACTIONABILITY_RANK',
'DEVELOPMENT_STATUS'])['DRUG_COMBINATION'].nunique().unstack(fill_value=0)

# Returning the adjusted pivot table for validation of numbers
pivot_table_combination_counts
```

Table 2. Validation of number of distinct drug combinations by actionability rank and development status

Actionability Rank	Approved FDA	Approved other	Case study	Experimental	Observational	Out of trials human study	Phase 1	Phase 2	Phase 3	Phase 4	Retrospective/ Meta-analysis	Unknown
1	65	2	10	0	3	4	243	533	136	21	12	22
2	79	1	4	0	15	5	665	1654	507	41	8	32
3	30	0	12	1	15	16	1033	1617	259	16	33	24
4	0	0	39	0	0	0	0	0	0	0	0	0
TOTAL	174	3	65	1	33	25	1941	3804	902	78	53	78

As it can be seen, Figure 2 showed wrong information about the name of drug combination, for example, the rank No 4 indicated 7 drug combinations approved by FDA (the data in Tabla 2 says that real number is 0). According to this, the code to plot was adjusted to show the number of drug combinations properly:

```
plt.figure(figsize=(14, 8))
sns.heatmap(pivot_table_combination_counts, annot=True, fmt="d", cmap='YlGnBu', cbar_kws={'label':
'Number of Distinct Drug Combinations'})
plt.title('Heatmap of Distinct Drug Combinations by Actionability Rank and Development Status')
plt.xlabel('Development Status')
plt.ylabel('Actionability Rank')
plt.show()
```

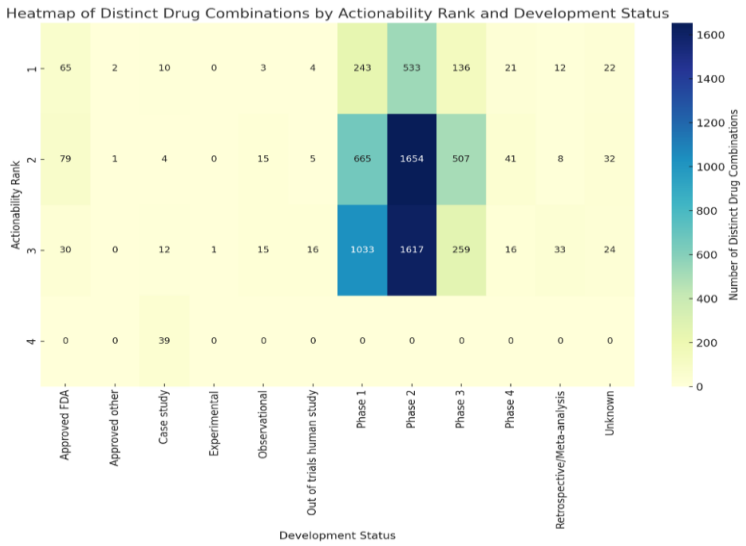

Figure 3. Heatmap of number of distinct drug combinations by actionability rank and development status, according to validation

Above plot does not add a lot of actionable information about the most prominent drugs used according to actionable mutations, so it was asked to chatGPT: "Graph but according to the top 10 drug combination" obtaining the next plot:

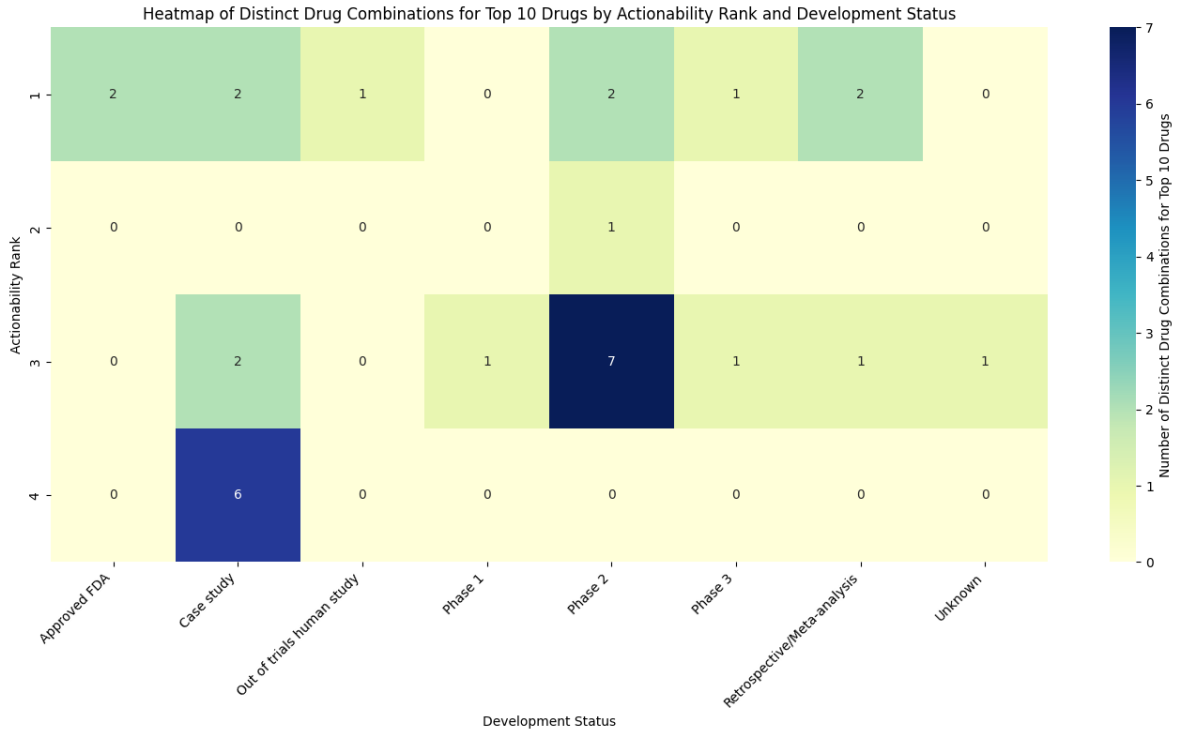

Figure 4. Heatmap of the top 10 drug combinations by actionability rank and development status

However, the matrix did not show the names of the drugs, so it was asked to chatGPT to "improve the above graph showing the name of the drugs" and this graph was obtained:

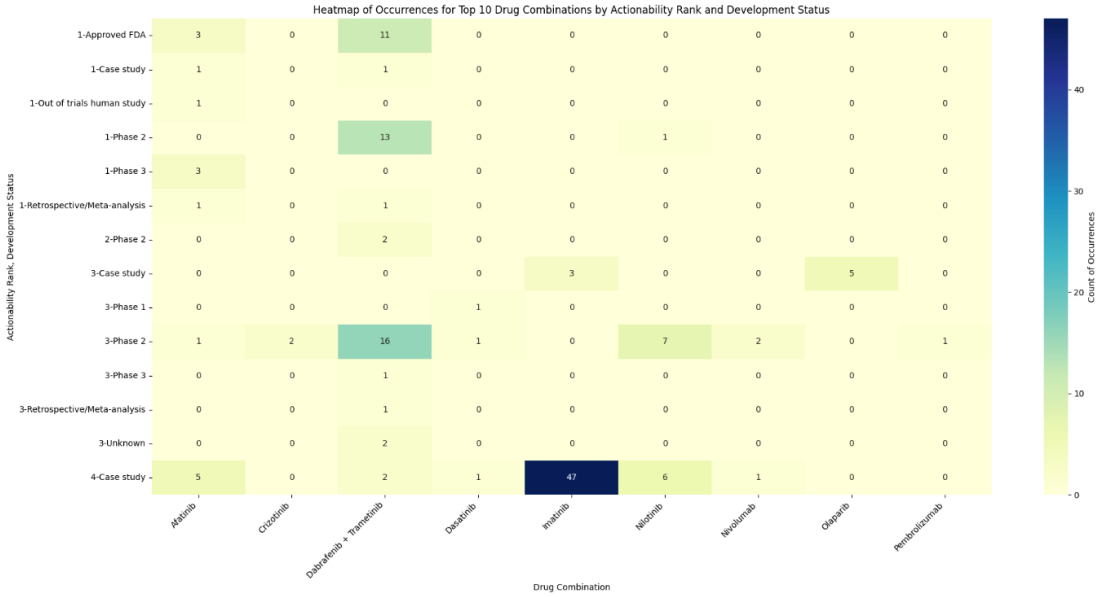

Figure 5. Heatmap of the top 10 drug combinations by actionability rank and development status and by the names of the combinations

In the above graph, the names of the combinations are now illustrated, the heatmap now visualizes the distribution of the top 10 drug combinations by actionability rank and development status. This graph highlights how these specific drugs are distributed across different stages of development and ranks of actionability,

providing insight into the concentration of the research and clinical interests in these combinations. It is observed, for example, that Imatinib has 47 concurrences just in the rank No.4 of actionability in case studies.

To validate the graph, we created in python a new pivot table thinking as ChatGPT and we filtered occurrences rather than unique counts, and put drug names to the columns:

```
pivot_table_top_10_detailed = data_filtered_top_10.pivot_table(index=['ACTIONABILITY_RANK',
'DEVELOPMENT_STATUS'],
                                                               columns='DRUG_COMBINATION',
                                                               aggfunc='size',
                                                               fill_value=0)
pivot_table_top_10_simple = pivot_table_top_10_detailed.reset_index()
pivot_table_top_10_simple:
```

Table 3. Validation by the top 10 drug combinations by actionability rank and development status and by the names of the combinations.

DRUG_COMBINATION	ACTIONABILITY_RANK	DEVELOPMENT_STATUS	Afatinib	Crizotinib	Dabrafenib + Trametinib	Dasatinib	Imatinib	Nilotinib	Nivolumab	Olaparib	Pembrolizumab
0	1	Approved FDA	3	0	11	0	0	0	0	0	0
1	1	Case study	1	0	1	0	0	0	0	0	0
2	1	Out of trials human study	1	0	0	0	0	0	0	0	0
3	1	Phase 2	0	0	13	0	0	1	0	0	0
4	1	Phase 3	3	0	0	0	0	0	0	0	0
5	1	Retrospective/ Meta-analysis	1	0	1	0	0	0	0	0	0
6	2	Phase 2	0	0	2	0	0	0	0	0	0
7	3	Case study	0	0	0	0	3	0	0	5	0
8	3	Phase 1	0	0	0	1	0	0	0	0	0
9	3	Phase 2	1	2	16	1	0	7	2	0	1
10	3	Phase 3	0	0	1	0	0	0	0	0	0
11	3	Retrospective/ Meta-analysis	0	0	1	0	0	0	0	0	0
12	3	Unknown	0	0	2	0	0	0	0	0	0
13	4	Case study	5	0	2	1	47	6	1	0	0

Later, we were interested in identifying the most common genomic mutations and combinations, we generated the following prompt “code the 10 most common genomic mutations and drug combinations” and this was the plot of ChatGPT:

Figure 6. Heatmap of genomic mutation vs drug combinations.

To validate the Figure 6, we generated a pivot table:

```
import pandas as pd
from collections import Counter
# Counting occurrences of genomic mutations and drug combinations
genomic_mutation_counts = Counter(data['GENOMIC_MUTATION_ID'])
drug_combination_counts = Counter(data['DRUG_COMBINATION'])
# Selection of the top 10 most common genomic mutations and drug combinations
top_genomic_mutations = [mutation for mutation, count in genomic_mutation_counts.most_common(10)]
top_drug_combinations = [drug for drug, count in drug_combination_counts.most_common(10)]
# Filtering the dataset for entries with the top genomic mutations and drug combinations
filtered_data = data[(data['GENOMIC_MUTATION_ID'].isin(top_genomic_mutations)) &
                     (data['DRUG_COMBINATION'].isin(top_drug_combinations))]
# Create a pivot table for the heatmap
pivot_table = filtered_data.pivot_table(index='GENOMIC_MUTATION_ID',
columns='DRUG_COMBINATION',
                                        aggfunc='size', fill_value=0)
```

Table 4. Validation of the 10 most common genomic mutations and drug combinations.

GENOMIC_MUTATION_ID	Afatinib	Dabrafenib + Trametinib	Dasatinib	Imatinib	Nilotinib	Nivolumab	Pembrolizumab
51765492	2	0	0	0	0	0	0
55386424	0	0	1	1	1	0	0
56056643 + 56059110	0	27	0	0	0	0	0
56056643 + 56059110 + 56057713	0	8	0	0	0	0	0
56056643 + 56059110 + 56057713 + 56058419 + 56115685 +	0	9	0	0	0	1	1
56288520 + 56059623 + 56293156 + 56065204 + 56075762 +	0	0	0	0	1	0	0
56076564 + 56080151 + 56124012 + 56152883 + 56228237	0	0	0	0	1	0	0
59323790	0	0	0	0	1	0	0

Discussion

Critical Thinking and Problem Solving in Medicine

Teaching students to critically interpret genomic data and other big data from biomedicine is the future of personalized medicine. Medical datasets are available for free use and sometimes it is difficult to interpret the data. Nowadays, OpenAI's ChatGPT Code Interpreter can handle data and carry out different coding operations, in our case, we played with bottom and less complex prompts to understand a complex dataset. The first step was by asking the number of drug combinations by actionability rank and development. In this case the plots were amazingly clear, showing the histograms of the 10 top combinations whose data were validated with a pivot table and with a raw analysis by searching for the reflected numbers. The precision and reliability of this graphs were 100%. These data were searched in the literature and no graphic as such was found. Just few articles talk about this specific dataset inside the COSMIC database, for example Sondka and cols talk about COSMIC curated information on 988 actionable variants across 445 genes., informing data on 11,121 clinical trials and 5,174 treatment combinations but they did not explore data combinations (Sondka et al., 2024) Boland and cols., have found in COSMIC that 30% of a total of 500 patients had mutations in potentially actionable genes, highlighting the significant potential for genotype-driven clinical trials as they represent a substantial portion of population with advanced cancer, underscoring the clinical utility of this information, furthermore they found that actionable mutations are not confined to the most common cancers (such as KRAS mutations in pancreas) so their approach went even deeper (Boland et al., 2015) Other previous studies, have focused on ICGC/TCGA Pan-Cancer Analysis datasets (also incorporated in COSMIC) to give valuable insights such as signals of positive selection to identify driver genes, mutational patterns on driver genes, functional networks such as the 13 candidate cancer genes within their functional interaction context (ATR, STAG2, PIK3CG, MED13, NCOA3, PIK3CB, FOXA1, CDKN1B, CDKN1A, MED17, FOXA2, FHEB, PRKCD) (Tamborero et al., 2013)

The potential of systematic and context-specific evaluation of drug combinations to enhance the precision and effectiveness of cancer treatments is imperative for new drug combinations discoveries. In our study, the Figure 3 (remember that Figure 2 was wrong and later on was validated) shows the combinations by actionability rank and development status. This graph updates the number of drug combinations reported in COSMIC and demonstrates that most of them are representatives of 2 or 3 actionability rank and participating in phase 1, 2 and 3 of development status. This information, have profound implications for clinical trials and cancer treatment development, for example, Jaaks and cols., studied 2,025 two-drug combinations across 125 molecularly characterized cancer cell lines, identified effective combinations, particularly those showing high synergy in cancers with unmet clinical needs, showing a rational efforts to develop combination treatments (Jaaks et al., 2022). Other works, such as Yang and cols, use this data to predict drug synergies, the importance of computational approaches, and the challenges faced in these predictions, however they focused in MdrDB, a database that integrates data from seven publicly available datasets, which is the largest database of its kind (Yang et al., 2023). Other computational models have been used to predict synergies focusing on co-targeting functionally proximal genes to enhance the efficacy of drug combinations finding some interesting combinations such as CDK4/6 Inhibition with Estrogen Receptor Therapy in breast cancer or BRAF and MEK1/2 Inhibitors in melanoma (Nair et al., 2023). All these this combination of experimental screening and computational analysis provides comprehensive methods for identifying and validating effective drug combinations in cancer treatment-

ChatGPT also has the power to decompose complex data as it was asked to "Graph but according to the top 10 drug combination" obtaining Figure 4. As this matrix did not show the names of the drugs, it was asked to "improve the above graph showing the name of the drugs" which brings interesting information (Figure 5) about the top 10 drug combinations (with their names) by actionability rank and development status. This plot was validated in Table 3 with 100 matching. Finally, it was also possible to identify the most common genomic mutations and drug combinations (Figure 6) which was also validated with a pivot table ın Table 4. The data obtained in both figures is essential for understanding the distribution and prevalence of treatment strategies for various genetic mutations (according to the data set). According to Sondka, understanding the distribution and prevalence of treatment strategies for various genetic mutations is crucial for advancing personalized medicine, enhancing the effectiveness of treatments, and ultimately improving patient care in oncology and beyond. This approach not only helps in treating diseases but also in preventing them and improving the quality of life for patients (Sondka et al., 2024)

Implications for Education and Research in Medical Science

In education, especially medical and genetics education, the integration of Big Data analytics is transforming core curriculums. Educators are now incorporating data science and ML into courses to equip students with the skills necessary to handle large datasets and complex information. This shift is aimed at preparing students for the challenges they will face in a data-driven healthcare environment, ensuring they can leverage these tools for improved patient care and research outcomes.

In the realm of personalized medicine, identifying actionable information—such as specific biomarkers for diseases—allows for more precise treatment strategies. For instance, knowing particular genetic mutations in cancer can guide the use of targeted therapies, thereby improving treatment efficacy and patient outcomes. This is crucial in cancer treatment, where somatic mutations can influence the effectiveness of drugs. These insights help in advancing personalized medicine by making treatment more responsive to individual genetic profiles. Through the effective use of AI like ChatGPT and ML platforms, both sectors can benefit from more precise and actionable insights, driving forward the capabilities of professionals to make informed decisions and innovate within their respective fields.

Conclusıon

To mitigate the risks associated with potential pitfalls in data-driven medicine, interdisciplinary collaboration among data scientists, healthcare professionals, ethicists, and policymakers is required. Additionally, robust quality assurance mechanisms, transparent reporting standards, and ongoing evaluation of data sources and algorithms are needed to ensure the reliability and validity of insights derived from big data analytics in healthcare. Treatment planning, risk prediction, and diagnostic approach are essential components of artificial intelligence in precision medicine. Although data management and replication issues pose challenges for AI-based precision medicine, it has been shown to enhance healthcare and reduce medical errors. Electronic health records and imaging analysis have been proven to present opportunities for AI-based precision medicine.

A future is envisioned in which data integration is achieved, enhancing its effectiveness, and bolstering the confidence of both researchers and the general population, thereby promoting its usage and contributing to its enhancement. However, the Big Data revolution has not yet reached its full potential; significant challenges still need to be overcome. Among them, the importance of collecting high-quality data and the necessity of efficient processes to translate large volumes of complex data into meaningful insights stand out. It is crucial to recognize that the mere availability of data does not ensure its validity. Additionally, issues such as data interoperability, as well as the legal and ethical implications associated with the use of personal data, must be addressed. Despite these challenges, it is evident that this technology, although still in its early stages, is at the beginning of a long and promising journey.

In summary, the ability to find and use actionable information in complex databases like COSMIC is also essential to advance in personalized medicine, improve patient care, and foster research and development in the field of cancer. Automation and open access are key elements that enable efficient management of large volumes of data, promote equality in access to information, and facilitate collaboration and rapid advancement in health research and education. The use of ChatGPT is a powerful tool to enhance medical research and to obtain clues or fingerprints of actionable data that can be extrapolated to the clinic. In our case, to bring new information about current combinations for cancer in the market, to give a landscape of actionability and to identify a pool of biomarkers that, if handled properly, can guide personalized treatment decisions. The validation of the data is the most important tool when these generative tools are introduced to the research, added to the use of ethical guidelines and protocols to avoid biases and comparison with previous and current data found in literature.

Scientific Ethics Declaration

The authors declare that the scientific ethical and legal responsibility of this article published in EPSTEM journal belongs to the authors.

Acknowledgements or Notes

* This article was presented as oral presentation at the International Conference on Basic Sciences, Engineering and Technology (www.icbaset.net) held in Alanya/Turkey on May 02-05, 2024.

* The authors would like to thank the financial support from Tecnologico de Monterrey through the "Challenge-Based Research Funding Program 2022". Project ID # I001 - IFE001 - C1-T1 - E. Also, financial and technical support from Writing Lab, Institute for the Future of Education, Tecnologico de Monterrey, Mexico.

References

Batko, K., & Ślęzak, A. (2022). The use of big data analytics in healthcare. *Journal of Big Data*, *9*(1), 3.

Blatter, T. U., Witte, H., Nakas, C. T., & Leichtle, A. B. (2022). Big data in laboratory medicine faır quality for AI?. *Diagnostics*, *12*(8), 1923.

Boland, G. M., Piha-Paul, S. A., Subbiah, V., Routbort, M., Herbrich, S. M., Baggerly, K., Patel, K. P., Brusco, L., Horombe, C., Naing, A., Fu, S., Hong, D. S., Janku, F., Johnson, A., Broaddus, R., Luthra, R., Shaw, K., Mendelsohn, J., Mills, G. B., & Meric-Bernstam, F. (2015). Clinical next generation sequencing to identify actionable aberrations in a phase I program. *Oncotarget*, *6*(24), 20099-20110.

Chakraborty, C., Bhattacharya, M., & Lee, S.-S. (2023). Artificial intelligence enabled ChatGPT and large language models in drug target discovery, drug discovery, and development. *Molecular Therapy - Nucleic Acids*, *33*, 866-868.

COSMIC. (n.d.). *COSMIC Actionability | Product Information*. Recuperado 12 de abril de 2024, de Retrieved from https://cancer.sanger.ac.uk/actionability/help

Fischer, C., Pardos, Z. A., Baker, R. S., Williams, J. J., Smyth, P., Yu, R., Slater, S., Baker, R., & Warschauer, M. (2020). Mining big data in education: affordances and challenges. *Review of Research in Education*, *44*(1), 130-160.

Greene, A. C., Giffin, K. A., Greene, C. S., & Moore, J. H. (2016). Adapting bioinformatics curricula for big data. *Briefings in Bioinformatics*, *17*(1), 43-50.

Jaaks, P., Coker, E. A., Vis, D. J., Edwards, O., Carpenter, E. F., Leto, S. M., Dwane, L., Sassi, F., Lightfoot, H., Barthorpe, S., van der Meer, D., Yang, W., Beck, A., Mironenko, T., Hall, C., Hall, J., Mali, I., Richardson, L., Tolley, C., &Garnett, M. J. (2022). Effective drug combinations in breast, colon and pancreatic cancer cells. *Nature*, *603*(7899), 166-173.

Lee, C. H., & Yoon, H.-J. (2017). Medical big data: Promise and challenges. *Kidney Research and Clinical Practice*, *36*(1), 3-11.

Li, J., Dada, A., Puladi, B., Kleesiek, J., & Egger, J. (2024). ChatGPT in healthcare: A taxonomy and systematic review. *Computer Methods and Programs in Biomedicine*, *245*, 108013.

Luan, H., Geczy, P., Lai, H., Gobert, J., Yang, S. J. H., Ogata, H., Baltes, J., Guerra, R., Li, P., & Tsai, C.-C. (2020). Challenges and future directions of big data and artificial ıntelligence in education. *Frontiers in Psychology*, *11*.

Nair, N. U., Greninger, P., Zhang, X., Friedman, A. A., Amzallag, A., Cortez, E., Sahu, A. D., Lee, J. S., Dastur, A., Egan, R. K., Murchie, E., Ceribelli, M., Crowther, G. S., Beck, E., McClanaghan, J., Klump-Thomas, C., Boisvert, J. L., Damon, L. J., Wilson, K. M., & Benes, C. H. (2023). A landscape of response to drug combinations in non-small cell lung cancer. *Nature Communications*, *14*(1), 3830.

Ostroverkhova, D., Przytycka, T. M., & Panchenko, A. R. (2023). Cancer driver mutations: Predictions and reality. *Trends in Molecular Medicine*, *29*(7), 554-566.

Puiu, A., Gómez Tapia, C., Weiss, M. E. R., Singh, V., Kamen, A., & Siebert, M. (2024). Prediction uncertainty estimates elucidate the limitation of current NSCLC subtype classification in representing mutational heterogeneity. *Scientific Reports*, *14*(1), 6779.

Sahu, M., Gupta, R., Ambasta, R. K., & Kumar, P. (2022). Artificial intelligence and machine learning in precision medicine: A paradigm shift in big data analysis. Progress in Molecular Biology and Translational Science, *190*(1), 57-100.

Sondka, Z., Dhir, N. B., Carvalho-Silva, D., Jupe, S., Madhumita, McLaren, K., Starkey, M., Ward, S., Wilding, J., Ahmed, M., Argasinska, J., Beare, D., Chawla, M. S., Duke, S., Fasanella, I., Neogi, A. G., Haller, S., Hetenyi, B., Hodges, L., & Teague, J. (2024). COSMIC: A curated database of somatic variants and clinical data for cancer. *Nucleic Acids Research*, *52*(D1), D1210-D1217.

Tamborero, D., Gonzalez-Perez, A., Perez-Llamas, C., Deu-Pons, J., Kandoth, C., Reimand, J., Lawrence, M. S., Getz, G., Bader, G. D., Ding, L., & Lopez-Bigas, N. (2013). Comprehensive identification of mutational cancer driver genes across 12 tumor types. *Scientific Reports*, *3*, 2650.

Tempini, N., & Leonelli, S. (2021). Actionable data for precision oncology: Framing trustworthy evidence for exploratory research and clinical diagnostics. *Social Science & Medicine, 272*, 113760.

Yang, Z., Ye, Z., Qiu, J., Feng, R., Li, D., Hsieh, C., Allcock, J., & Zhang, S. (2023). A mutation-induced drug resistance database (MdrDB). *Communications Chemistry*, *6*(1), 1-9.

Wesson, P., Hswen, Y., Valdes, G., Stojanovski, K., & Handley, M. A. 2022. Risks and opportunities to ensure equity in the application of big data research in public health. *Annu.Rev.Public Health, 43,* 59-78.

Author Information

M.S. Ramírez-Montoya
EGADE Business School, Institute for the Future of Education, Tecnologico de Monterrey,
Monterrey, Mexico
Contact e-mail: *solramirez@tec.mx*

G.A. Sanchez-Zuno
School of Medicine. Internal Medicine Department (USA),
Yale University,
New Haven, CT, U.S.A.

R.M. Marques González
Hospital Civil de Guadalajara, Medical Genetics Department, Mexico

F. Casıllas-Muñoz
Institute for the Future of Education, Tecnologico de Monterrey, Mexico City, Mexico

To cite this article:

Ramírez-Montoya, M.S., Sanchez-Zuno, G.A., Marques González, R. M., & Casillas-Muñoz, F. (2024). Complex thinking abilities in the rapidly evolving field of genomics and personalized medicine: Analysis of actionability on cancer with ChatGPT and literature. The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), *28,* 34-46.