Version 1.0

Authors: Marco Huberts and Ayat Abourashed

Last updated: March 28, 2025

0 Abstract

Cancer remains one of the greatest global health challenges, with standard "one-size-fits-all" treatments often leading to suboptimal outcomes, serious side effects, and escalating healthcare costs. Recent advances in gene expression profiling and artificial intelligence (AI) offer a promising solution by enabling treatments tailored to the genetic and molecular features of a patient's tumor. To address this critical need, we introduce weCURA, a precision oncology platform designed to predict individual patient responses to cancer therapies—initially focusing on pancreatic and brain cancers, which have high mortality rates and limited treatment options.

The AI-driven predictions consider both gene expression data and treatment responses, allowing clinicians to identify which treatments are most likely to be effective for specific tumor gene expression profiles. This iterative model is refined in near real-time through ongoing data contributions, bridging the gap between rapid laboratory discoveries and clinical application. By offering multiple revenue streams—including licensing, subscription-based access, and predictive analytics services—weCURA aims to remain sustainable while accelerating personalized care.

Ultimately, weCURA's personalized predictive framework has the potential to transform cancer treatment by minimizing trial-and-error approaches, reducing unnecessary side effects, and improving patient outcomes. Its dynamic adaptability ensures that as new research and biomarkers emerge, the model evolves to maintain clinical relevance, supporting a new era of truly patient-tailored oncology.

Table of Contents

1 Introduction

1.1 Current cancer treatments don't solve the problem

One of the biggest global health challenges remains cancer. With almost 20 million new cases of cancer and over 9 million deaths due to cancer in 2022, new strategies to fight this disease are necessary (1). Current cancer treatments include surgery, radiation, and systemic therapies like chemotherapy, targeted therapy, hormonal therapy, and immunotherapy. These treatments are typically selected based on certain patient factors like sex, cancer stage, and tumor grade (2–4). However, this approach often falls short in delivering optimal outcomes for every individual. The process of identifying effective cancer treatments can be long and tedious, taking months or even years before a therapy proves successful in fighting the disease. Compounding this challenge is the issue of treatment resistance, where cancer cells adapt and no longer respond to a previously effective therapy (5). As a result, many cancer patients must undergo a range of treatments in the hope that at least one will yield a therapeutic benefit.

Currently, the standard approach to cancer treatment is population-based, meaning that treatment recommendations are largely derived from clinical studies involving large groups of patients (6). Thus, these treatment decisions are guided by factors like the average responses of patients with similar cancer types, stages, and other demographic or clinical factors (4). While this method has led to significant advancements, it fails to account for the unique biological and genetic differences between individuals. This "one-size-fits-all" approach often leads to suboptimal outcomes for patients whose cancers do not conform to the patterns observed in population studies (7). For example, two patients diagnosed with the same type and stage of lung cancer may receive the same standard chemotherapy regimen based on clinical trial data. However, one patient's tumor may harbor an EGFR mutation that makes it highly responsive to EGFR-targeted therapy, while the other patient's cancer may have a KRAS mutation that renders it resistant to EGFR inhibitors. In this case, without molecular testing, the patient with the EGFR mutation might unnecessarily undergo a trial of chemotherapy, delaying access to more effective targeted treatment (8,9). A shift toward personalized medicine, which tailors treatments based on the specific characteristics of each patient's cancer, is therefore essential to improve efficacy and reduce the burden of cancer worldwide.

1.2 Solution: Personalizing cancer treatments

One approach to developing personalized treatments is to identify biomarkers, such as genetic mutations, gene expression levels, or proteins, that can predict a patient's sensitivity to the therapy of interest (10). Several studies have shown that especially gene expression levels of cancers have a huge impact on therapy outcome (11,12). However, there are too few effective tools that can help clinicians predict therapy outcomes per individual patient.

In today's rapidly evolving technological landscape, harnessing biomarker data and artificial intelligence (AI) is critical for addressing the clear need for more effective predictive tools. Researchers are developing novel methods to identify biomarkers that indicate whether a patient will respond to a particular therapy. These approaches show promise not only for targeted

therapies—designed to attack specific cancer cells—but also for immunotherapies, which boost the immune system's ability to recognize and attack cancerous tissue.

At PoSciDonDAO, we combine gene expression data linked to treatment sensitivity with AI to create weCURA: an advanced tool that predicts patient responses to cancer treatments based on each tumor's unique gene expression profile.

2 What is weCURA?

weCURA is a precision oncology tool predicting patient responses to specific cancer treatments for skin, pancreatic and brain cancer based on patients' genetic makeup. These three cancer types are particularly suitable for this AI model due to their high mortality rates, limited treatment options, and heterogeneity in tumor biology. Pancreatic cancer, often diagnosed at advanced stages, has one of the lowest survival rates among all cancers, emphasizing the urgent need for personalized therapeutic strategies (13–15). Similarly, brain cancers, such as glioblastoma, are notoriously aggressive and resistant to standard treatments, with survival outcomes that vary depending on the genetic background of the patient (16,17). Skin cancers, including melanoma, frequently exhibit distinct genetic mutations such as BRAF or NRAS, playing an important role in variable treatment responses and highlighting the importance of genetic profiling for effective therapy selection.(18) By leveraging AI to analyze the genetic makeup of patients with these cancers, weCURA aims to identify tailored treatment regimens, improving efficacy and reducing the trial-and-error approach often associated with cancer therapy. This approach has the potential to significantly enhance outcomes for patients facing these challenging diagnoses.

3 Data Collection for weCURA

To develop the AI model, it is essential to gather *in vitro* data from preferably near-patient 3-dimensional models such as spheroids and organoids alongside the corresponding drug treatment responses to enable accurate predictions. The data collection process is streamlined into two distinct pipelines (Figure 1).

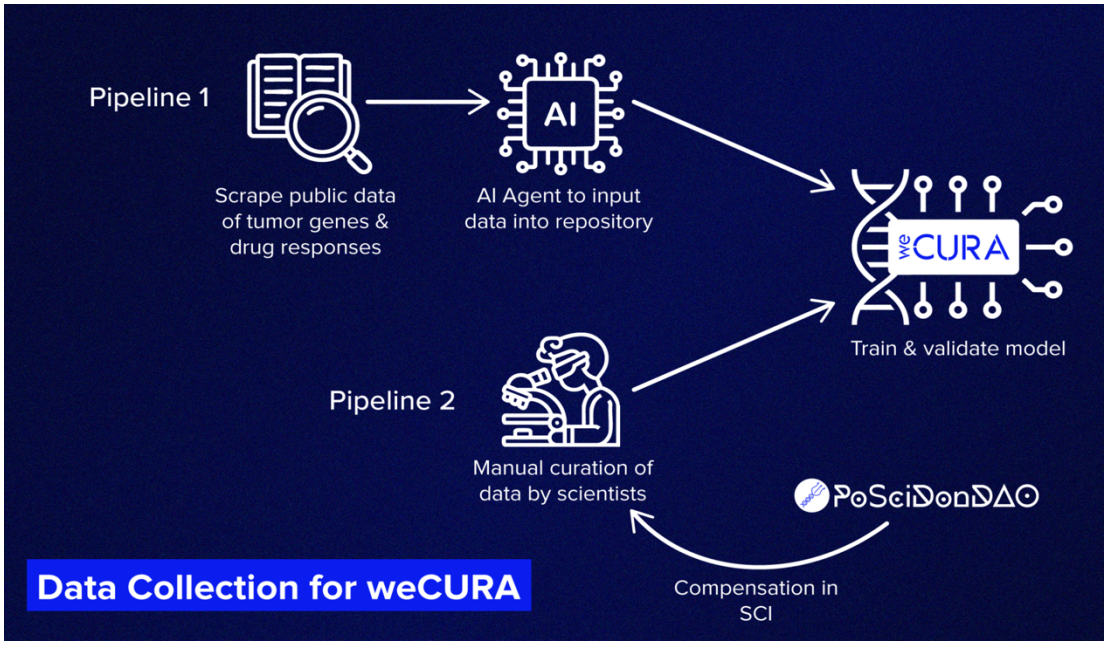

Figure 1. Pipelines for data collection for weCURA

3.1 Pipeline 1

The first approach for data collection involves mining gene expression levels and drug-response profiles from reputable published databases, such as Gene Expression Omnibus, DepMap and the Genomics of Drug Sensitivity in Cancer. These databases provide a wealth of high-quality, publicly available data essential for building the AI model. After retrieving the data, an AI agent will process and organize it systematically, ensuring seamless integration into the weCURA repository for further analysis and use in model training.

3.2 Pipeline 2

For an alternative method for data collection, we will collaborate with scientists who can contribute the necessary data. These researchers will manually curate the information and submit it through our online submission system, ensuring its direct integration into the weCURA repository. To acknowledge their valuable contributions, participating scientists will be rewarded with SCI tokens, providing both recognition and incentive for their efforts.

4 weCURA Model Overview

Once the initial phase of data collection is complete—though this will remain an ongoing effort as new research emerges—we will focus on building, training, and validating the AI model to predict drug responses for pancreatic and brain cancers. weCURA leverages a gradient-boosted tree model (XGBoost), specifically selected for its robust performance on gene expression data, interpretability, and effectiveness even with small to moderate dataset sizes. Validation is a critical step in ensuring the model's accuracy and reliability, made possible through continuous contributions from scientists conducting cutting-edge research. These researchers will provide curated datasets, which will not only expand the weCURA repository but also enable iterative testing and refinement of the model.

To address the variability in data availability, we will develop two distinct model pipelines: one trained exclusively on genetic information and another incorporating both genetic data and gene expression levels. The first model will predict whether a patient with a specific tumor gene is likely to respond to a given treatment, while the second model will estimate the degree of sensitivity a patient may exhibit to that treatment. As new datasets become available, these predictions can be validated against real-world experimental outcomes, ensuring the model remains accurate and clinically relevant.

This iterative validation process allows the AI model to adapt and evolve in near real-time, incorporating the latest scientific discoveries, such as novel biomarkers or new drug therapies. By comparing predictions against experimental results provided by participating scientists, discrepancies can guide adjustments to the model, enhancing its predictive power. This dynamic approach ensures the AI model grows into a robust, versatile, and reliable tool for precision oncology, capable of tailoring treatments to individual patients as the field advances.

4.1 Technical Model Construction Steps

Building the weCURA model involves a multi-stage pipeline that transforms raw genetic and gene-expression data into actionable predictions of drug response for pancreatic and brain cancers. Below is an overview of the core steps:

1. **Data Ingestion & Preprocessing**
    - **Data Consolidation**: Integrate datasets from both Pipeline 1 (AI agent trained on public repositories) and Pipeline 2 (curated data from scientists).
    - **Normalization RNA**: Standardize gene expression measurements, such as converting raw read counts to TPM (Transcripts Per Million). This ensures comparability across different studies and sequencing platforms.
    - **Normalization Sensitivity**: Standardize sensitivity data, such as determining the area under the curve (AUC) of the dose-response curve or the half maximum effective concentration (EC_{50}).
    - **Data Labeling**: Align each sample's gene expression profile with quantitative sensitivity metrics for each drug tested.
    - **Quality Checks**: Perform consistency checks to remove duplicated or contradictory entries. When possible, employ statistical harmonization to mitigate differences in assay protocols and measurement scales (e.g., normalizing EC_{50} ranges across studies).

2. **Model Architecture & Pipeline Setup**
    - **Single-Modality Model:** Implement gradient-boosted decision trees (XGBoost) that outputs a binary label—responsive vs. non-responsive—based solely on gene expression profiles.
    - **Hyperparameter Tuning**: Use automated techniques like Bayesian Optimization to find optimal learning rates and tree depth, number of boosting

rounds, and other parameters to maximize predictive performance.

3. **Training & Cross-Validation**
    - **Data Splitting**: Split the combined dataset into training, validation, and test sets.
    - **Model Training:** We will train the gradient-boosted tree model to convergence by minimizing a suitable loss function (e.g., binary log-loss) using built-in optimization routines. Monitor metrics such as accuracy, sensitivity, and specificity to gauge predictive performance.
    - **Regularization & Generalization**: To limit overfitting, we will apply techniques such as early stopping, limiting tree depth, setting minimum child weight, and using L1/L2 regularization on leaf weights. Gradient boosting frameworks like XGBoost inherently manage some overfitting through their ensemble nature and configurable regularization parameters.

4. **Validation with Real-World Data**
    - **External Datasets**: Continually incorporate new *in vitro* or *ex vivo* data points from participating scientists or new public datasets—preferably data from 3D organoid models or patient-derived xenografts (PDXs).
    - **Performance Monitoring**: Compare predicted drug responses against experimentally measured outcomes (e.g., EC_{50} values from viability assays). Use metrics like mean squared error (for regression-based sensitivity estimates) or confusion matrices (for binary classification) to assess model accuracy.

5. **Model Refinement & Deployment**
    - **Continuous Learning**: Integrate new curated datasets into the training pipeline, re-running hyperparameter tuning or ensemble updates to capture emerging biomarkers or drug-response patterns.
    - **Discrepancy Analysis**: Investigate cases where predictions and experimental / clinical results differ; these insights inform both biological follow-ups and model improvements.
    - **Deployment & API**: The final model will be deployed as a containerized microservice with a RESTful API, enabling secure and versioned drug-response predictions. SHAP will be integrated for model explainability, with frontend support for input submission and result visualization.

By following these steps—ranging from extensive data preprocessing and model design to continuous validation and learning—the weCURA platform can become a cutting-edge, reliable resource for precision oncology, evolving in simultaneously with new and up to date research insights.

5 Business Model

Generating revenue from the weCURA model can be achieved through multiple avenues. Here's an outline of potential strategies:

1. Licensing Agreements

- **Pharmaceutical Companies:** License the model to pharmaceutical companies for use in drug discovery, development, and clinical trial optimization. The model's ability to predict drug responses can help identify promising compounds, reduce development costs, and streamline trial designs.
- **Clinical Labs and Hospitals:** Offer licenses to clinical laboratories and hospitals to integrate the model into their diagnostic workflows. This could support oncologists in making personalized treatment recommendations for their patients.

2. Subscription-Based Access

- **Platform as a Service (PaaS):** Develop a subscription-based platform where researchers, clinicians, and biotech companies can access the model's predictions and datasets. Tiers can vary based on access level, prediction volume, or additional features such as real-time updates and advanced analytics.

3. Data Partnerships

- Partner with organizations to exchange access to datasets in return for a licensing fee or profit-sharing arrangement. Companies or research institutions contributing data can gain tailored predictions while expanding the model's repository, creating a mutually beneficial relationship.

4. Custom Model Development

- Offer tailored versions of the weCURA model for specific clients, such as companies focusing on rare cancers or niche therapeutic areas. These custom models could include client-specific genetic markers, datasets, or proprietary drug compounds.

5. Predictive Analytics Services

- Provide direct predictive analytics as a service, charging clients for reports or insights on their datasets without requiring them to license the entire model. This approach is particularly attractive for smaller biotech startups or academic researchers.

6. Integration with Drug Development Pipelines

- Collaborate with biotech firms to integrate the model into their drug development pipelines. Revenue could be generated through milestone payments, such as success fees for identifying drug candidates, or royalties based on the commercial success of drugs informed by the model.

8. Training and Certification

- Create training programs or certifications for clinicians and researchers on using the weCURA model effectively. Charge for enrollment in these programs, either as workshops or online courses.

9. Public-Private Partnerships

- Collaborate with healthcare providers, governments, or insurance companies to integrate the model into national or regional cancer treatment programs. Revenue could come from long-term contracts or shared cost savings achieved through personalized treatment optimization.

10. AI-Driven Drug Repurposing

- Use the model to identify new therapeutic applications for existing drugs (drug repurposing). Revenue can be generated by licensing these insights to pharmaceutical companies or through co-development agreements.

By diversifying revenue streams across licensing, service-based models, and strategic collaborations, weCURA can maximize its commercial potential while driving advancements in precision oncology.

6 Clinical Relevance

The clinical relevance of the weCURA model lies in its transformative potential to bridge the gap between laboratory discoveries and personalized patient care, placing the patient at the heart of innovation. By predicting drug responses with high precision, the model equips clinicians with invaluable insights to make informed, evidence-based decisions about the most effective treatments for individual patients. This not only reduces the trial-and-error approach often associated with cancer treatment but also minimizes the risk of unnecessary side effects and multiple trial rounds of expensive treatments, improving the overall quality of life for patients.

This capability becomes even more critical in skin, pancreatic and brain cancers, where therapeutic options are often limited, prognosis can be dire, and time is of the essence. For these patients, tailored treatment strategies are not just a luxury—they are a lifeline. By leveraging advanced predictive analytics, the weCURA model provides hope where traditional approaches may fall short, helping identify therapies that are most likely to succeed based on a patient's unique genetic and molecular profile.

Furthermore, by incorporating real-world experimental outcomes and iterating continuously with emerging data, the weCURA model evolves alongside the latest scientific advancements. This dynamic adaptability ensures that patients benefit from cutting-edge discoveries, such as novel biomarkers or breakthrough drugs, as soon as they become available. Ultimately, the model serves as a vital tool in guiding clinical decisions, expediting the translation of research into meaningful, impactful outcomes that directly benefit patients. By empowering clinicians with actionable predictions, the weCURA model has the potential to extend survival, enhance quality of life, and offer renewed hope to patients and their families.

7 References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians. 2024;74(3):229–63.

2. Dede Z, Tumer K, Kan T, Yucel B. Current Advances and Future Prospects in Cancer Immunotherapeutics. Medeniyet Medical Journal. 2023;38(1).

3. Liu B, Zhou H, Tan L, Siu KTH, Guan XY. Exploring treatment options in cancer: tumor treatment strategies. Signal transduction and targeted therapy. 2024;9(1):175.

4. Anand U, Dey A, Chandel AKS, Sanyal R, Mishra A, Pandey DK, et al. Cancer chemotherapy and beyond: Current status, drug candidates, associated risks and progress in targeted therapeutics. Genes & diseases. 2023;10(4):1367–401.

5. Mansoori B, Mohammadi A, Davudian S, Shirjang S, Baradaran B. The different mechanisms of cancer drug resistance: a brief review. Advanced pharmaceutical bulletin. 2017;7(3):339.

6. National Comprehensive Cancer Network. NCCN clinical practice guidelines in oncology (NCCN Guidelines®). (No Title). 2017;

7. Krzyszczyk P, Acevedo A, Davidoff EJ, Timmins LM, Marrero-Berrios I, Patel M, et al. The growing role of precision and personalized medicine for cancer treatment. Technology. 2018;6(03n04):79–100.

8. Shin DH, Jo JY, Choi M, Kim KH, Bae YK, Kim SS. Oncogenic KRAS mutation confers chemoresistance by upregulating SIRT1 in non-small cell lung cancer. Experimental & Molecular Medicine. 2023;55(10):2220–37.

9. American Lung Association. EGFR and Lung Cancer [Internet]. 2025 [cited 2025 Mar 17]. Available from: https://www.lung.org/lung-health-diseases/lung-disease-lookup/lung-cancer/symptoms-diagnosis/biomarker-testing/egfr

10. Zhou Y, Tao L, Qiu J, Xu J, Yang X, Zhang Y, et al. Tumor biomarkers for diagnosis, prognosis and targeted therapy. Signal transduction and targeted therapy. 2024;9(1):132.

11. Giampietri C, Scatozza F, Crecca E, Vigiano Benedetti V, Natali PG, Facchiano A. Analysis of gene expression levels and their impact on survival in 31 cancer-types patients identifies novel prognostic markers and suggests unexplored immunotherapy treatment options in a wide range of malignancies. J Transl Med. 2022 Oct 12;20:467.

12. Chawla S, Rockstroh A, Lehman M, Ratther E, Jain A, Anand A, et al. Gene expression based inference of cancer drug sensitivity. Nat Commun. 2022 Sep 27;13(1):5680.

13. Survival Rates for Pancreatic Cancer [Internet]. [cited 2025 Mar 17]. Available from: https://www.cancer.org/cancer/types/pancreatic-cancer/detection-diagnosis-staging/survival-rates.html

14. Li Q, Feng Z, Miao R, Liu X, Liu C, Liu Z. Prognosis and survival analysis of patients with pancreatic cancer: retrospective experience of a single institution. World Journal of Surgical Oncology. 2022 Jan 7;20(1):11.

15. Bengtsson A, Andersson R, Ansari D. The actual 5-year survivors of pancreatic ductal adenocarcinoma based on real-world data. Sci Rep. 2020 Oct 2;10(1):16425.

16. Stupp R, Hegi ME, Mason WP, van den Bent MJ, Taphoorn MJB, Janzer RC, et al. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. Lancet Oncol. 2009 May;10(5):459–66.

17. Pouyan A, Ghorbanlo M, Eslami M, Jahanshahi M, Ziaei E, Salami A, et al. Glioblastoma multiforme: insights into pathogenesis, key signaling pathways, and therapeutic strategies. Molecular Cancer. 2025 Feb 26;24(1):58.

18. Banerji U, Affolter A, Judson I, Marais R, Workman P. BRAF and NRAS mutations in melanoma: potential relationships to clinical response to HSP90 inhibitors. Molecular Cancer Therapeutics. 2008 Apr 1;7(4):737–9.