Dana-Farber
Cancer Institute

MatchMiner-AI: Artificial intelligence for cancer clinical trials matching

Kenneth L. Kehl, MD, MPH
Ethan Cerami, PhD
Jennifer Altreuter, MD

Dana-Farber Cancer Institute

Introduction to Trial Matching

Using MatchMiner-AI

1 2 3 4

MatchMiner-AI

Clinician Feedback

Introduction to Trial Matching

01

- Historically, less than 10% of adults with cancer enroll in clinical trials.
  - Rates are even lower in community practices and among less advantaged patients
  - Clinical trials often struggle to reach their accrual goals

Background

- Modern cancer trials are highly complex and targeted, often relying on concepts recorded in unstructured clinical data for screening.

➔ *Cancer type/histology/biomarkers*
➔ *Burden of disease*
➔ *Prior treatment*
➔ *Comorbidities*

Background

Patient Medical Information

Clinical Trial Eligibility Criteria

Matching Results

Clinical Trial Matching

MatchMiner-AI

02

MatchMiner

Computational platform to match patients to clinical trials, based on their genomic data.

Open Source, scalable to other Cancer Centers.

MatchMiner

MatchMiner is extremely good at matching patients to trials based on their genomic data.

Not all trials are genomically driven.

~80% of patient data is locked in unstructured clinical notes.

MatchMiner has Limitations

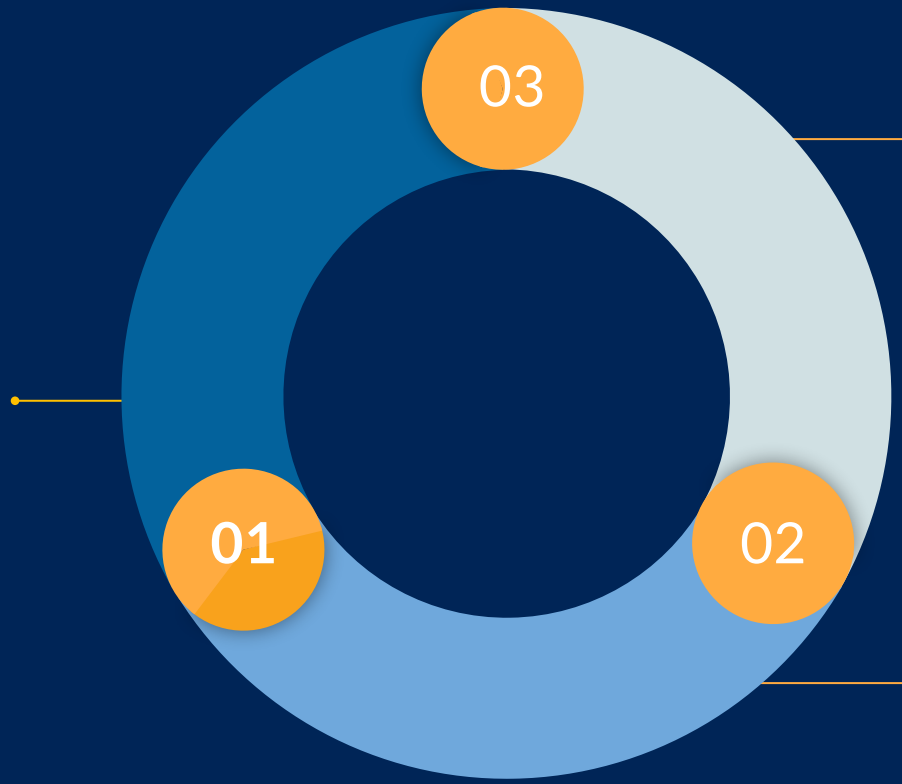
Not all trials are genomically driven.

~80% of patient data is locked in unstructured clinical notes.

MatchMiner has Limitations

Match Patients to all Genomically Driven Trials

MatchMiner Genomics



Trial Intelligence

MatchMiner Dashboard

Match Patients to all Trials

MatchMiner-AI

MatchMiner Suite

New Trial Matching platform that leverages AI to match patients to all clinical trials, based on their full medical record.

MatchMiner-AI

Platform is now live across DFCI (launched February 2026).

MatchMiner-AI

MatchMiner-AI leverages advances in Large Language Models (LLMs) to process and understand clinical text notes for each patient.

Uses patient summaries to match each patient to clinical trials.

How it works

How it works

1 We focus primarily on core clinical criteria.

- Age
- Sex
- Cancer Type/Histology
- Burden of disease
- Prior Treatment
- Key Biomarkers

Given core clinical criteria:

2 Use LLM to generate:

patient summaries and trial target populations.

How it works

3

Use synthetic patient data to train our models.

Historial patient data from 2016-2024.

13,425 enrollments onto 1,534 clinical trials.

How it works

4

Use retrospective patient enrollment data to evaluate our models.

Historical patient data from 2016-2024.

13,425 enrollments onto 1,534 clinical trials.

How it works

5 A three-step trial pre-screening tool.

How it works

5 A three-step trial pre-screening tool.

a Trial Space Model

Input:

Patient Summary
Clinical Trial Summary

How it works

How it works

5 A three-step trial pre-screening tool.

a Trial Space Model

Similarity metric measures: how close is this patient to this specific clinical trial?

5 A three-step trial pre-screening tool.

b Trial Checker Model

Input:

Patient Summary
Clinical Trial Summary

How it works

5 A three-step trial pre-screening tool.

b Trial Checker Model

Ask LLM to reason about whether this patient is a reasonable match for this trial.

How it works

5 A three-step trial pre-screening tool.

c "Boilerplate" Exclusions Model

Input:

Patient "Boilerplate" Summary
Clinical Trial "Boilerplate" Summary

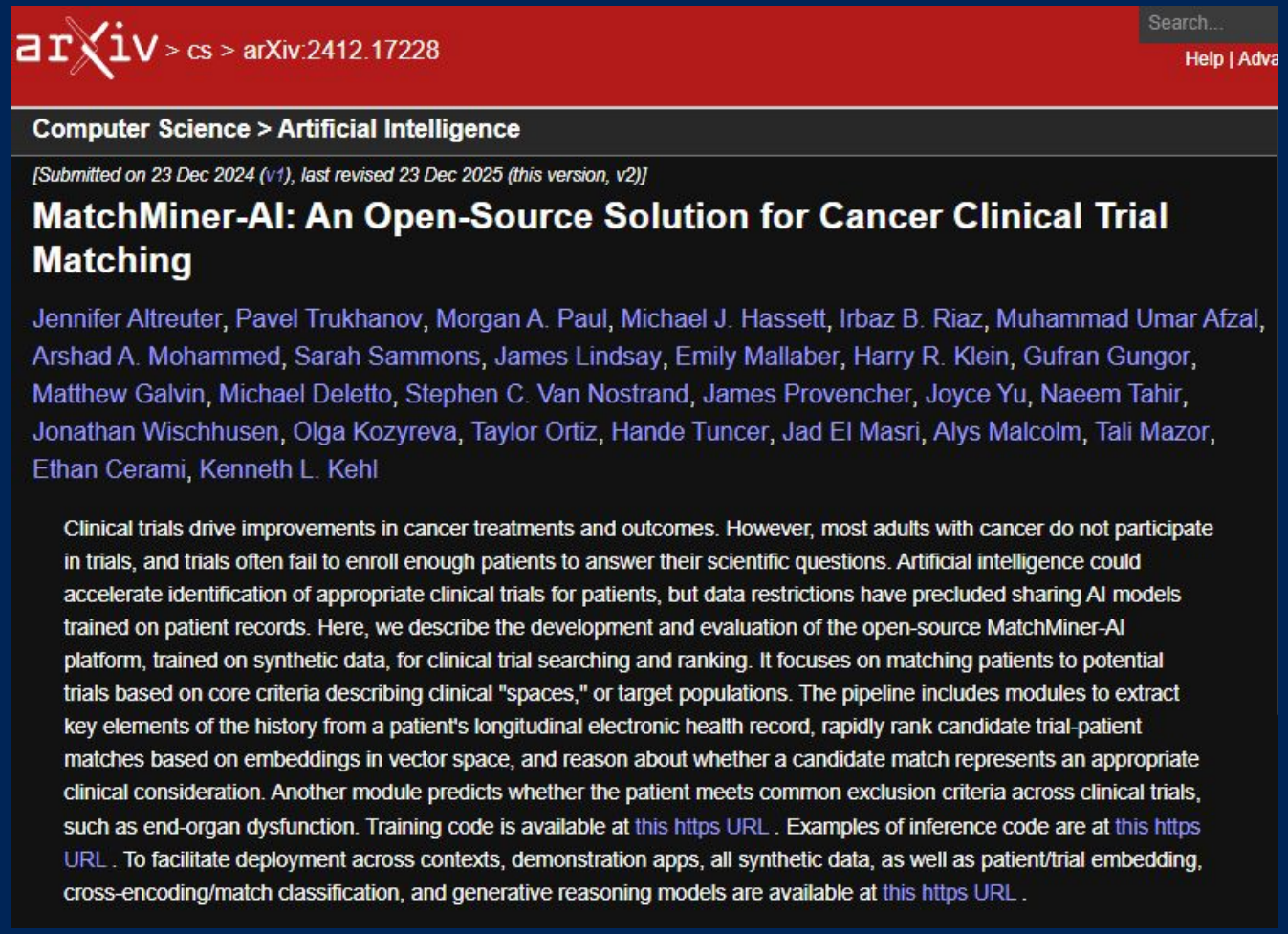
How it works

5 A three-step trial pre-screening tool.

c "Boilerplate" Exclusions Model

Ask LLM to reason about whether this patient matches any broad exclusions for this trial.

How it works



arXiv > cs > arXiv:2412.17228

Search...

Help | Adva

Computer Science > Artificial Intelligence

[Submitted on 23 Dec 2024 (v1), last revised 23 Dec 2025 (this version, v2)]

MatchMiner-AI: An Open-Source Solution for Cancer Clinical Trial Matching

Jennifer Altreuter, Pavel Trukhanov, Morgan A. Paul, Michael J. Hassett, Irbaz B. Riaz, Muhammad Umar Afzal, Arshad A. Mohammed, Sarah Sammons, James Lindsay, Emily Mallaber, Harry R. Klein, Gufran Gungor, Matthew Galvin, Michael Deletto, Stephen C. Van Nostrand, James Provencher, Joyce Yu, Naeem Tahir, Jonathan Wischhusen, Olga Kozyreva, Taylor Ortiz, Hande Tuncer, Jad El Masri, Alys Malcolm, Tali Mazor, Ethan Cerami, Kenneth L. Kehl

Clinical trials drive improvements in cancer treatments and outcomes. However, most adults with cancer do not participate in trials, and trials often fail to enroll enough patients to answer their scientific questions. Artificial intelligence could accelerate identification of appropriate clinical trials for patients, but data restrictions have precluded sharing AI models trained on patient records. Here, we describe the development and evaluation of the open-source MatchMiner-AI platform, trained on synthetic data, for clinical trial searching and ranking. It focuses on matching patients to potential trials based on core criteria describing clinical "spaces," or target populations. The pipeline includes modules to extract key elements of the history from a patient's longitudinal electronic health record, rapidly rank candidate trial-patient matches based on embeddings in vector space, and reason about whether a candidate match represents an appropriate clinical consideration. Another module predicts whether the patient meets common exclusion criteria across clinical trials, such as end-organ dysfunction. Training code is available at this https URL . Examples of inference code are at this https URL . To facilitate deployment across contexts, demonstration apps, all synthetic data, as well as patient/trial embedding, cross-encoding/match classification, and generative reasoning models are available at this https URL .

How it works

Using MatchMiner-AI

03

Two Access Methods

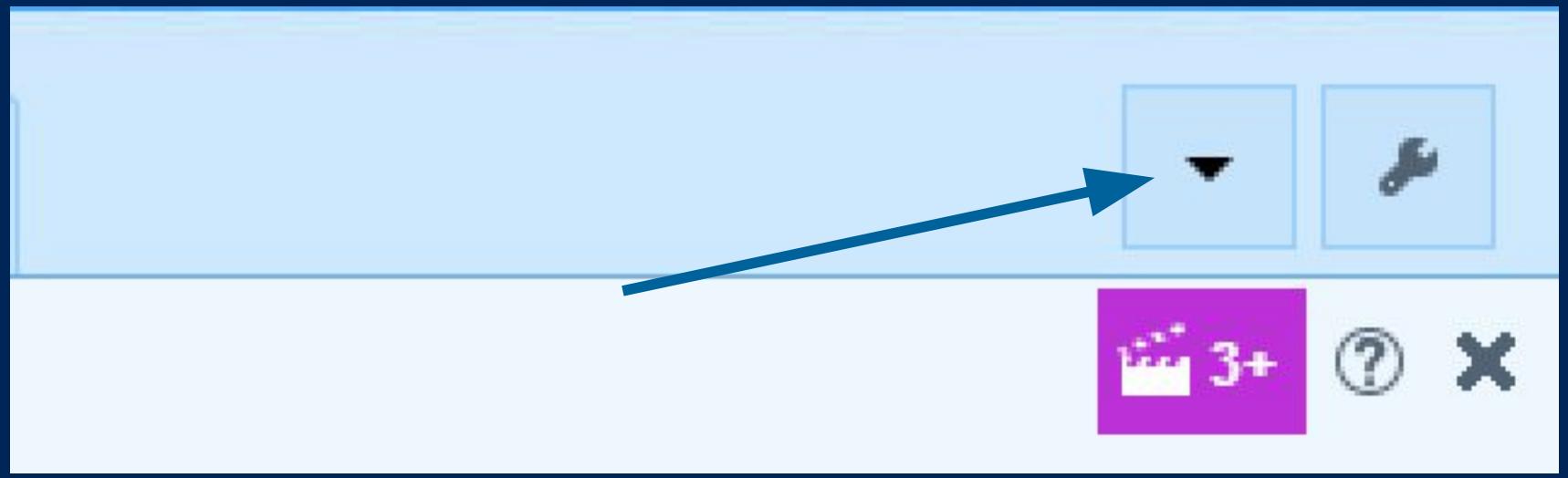
1. Access via EPIC.
2. Access via Website.

1.

Using MatchMiner-AI

1 Access via EPIC.

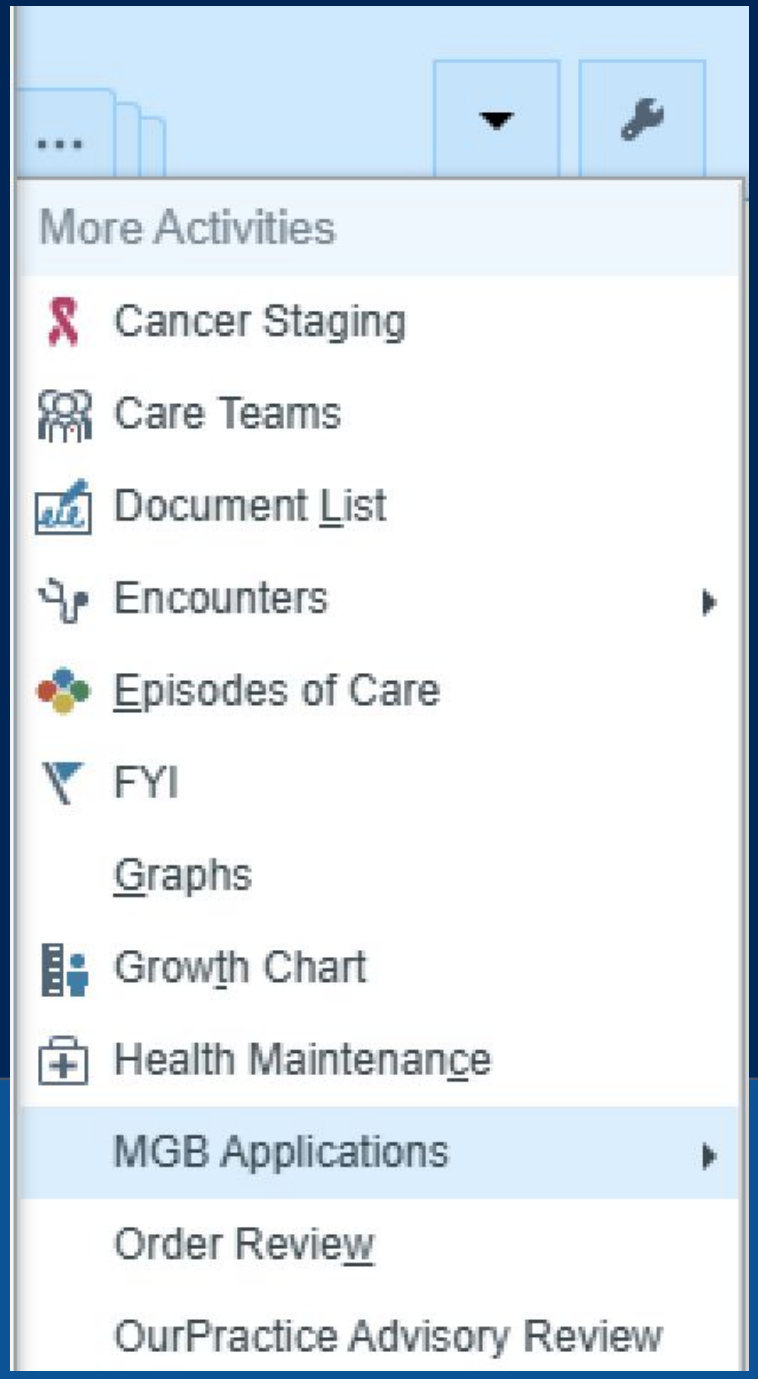
Patient View



Using MatchMiner-AI

1 Access via EPIC.

→ MGB Applications

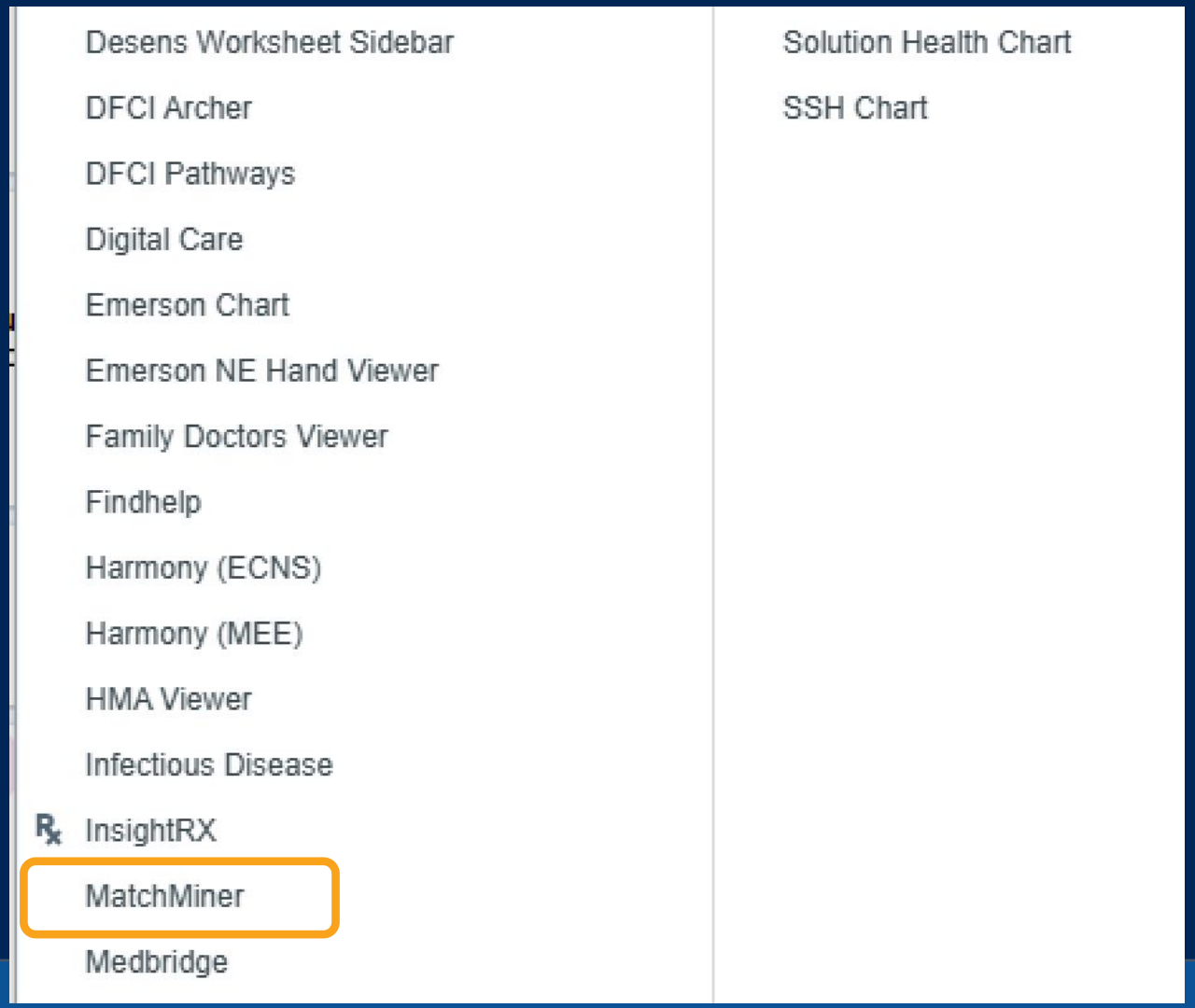


Using MatchMiner-AI

1 Access via EPIC.

MGB Applications

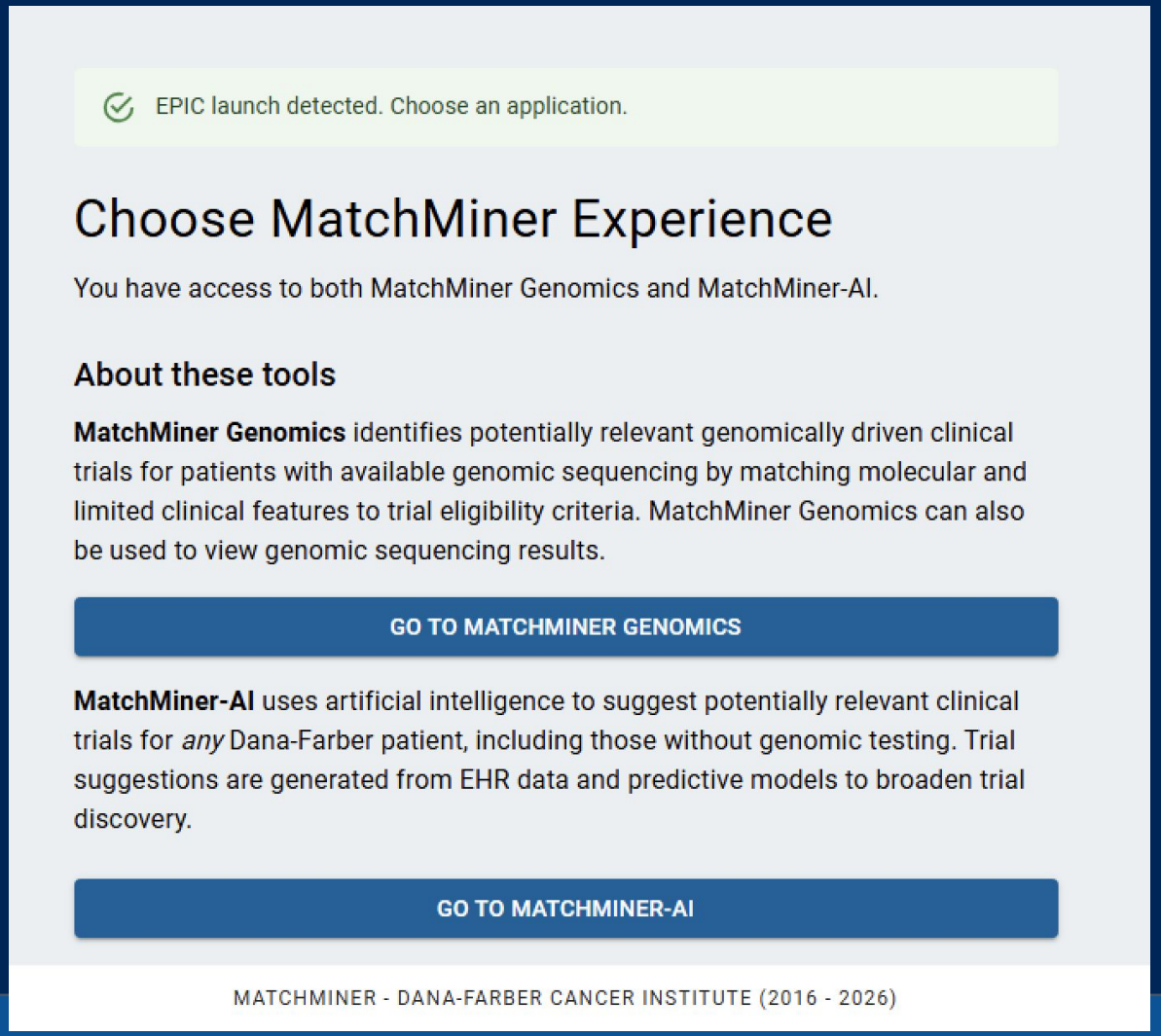
MatchMiner



Using MatchMiner-AI

1 Access via EPIC.

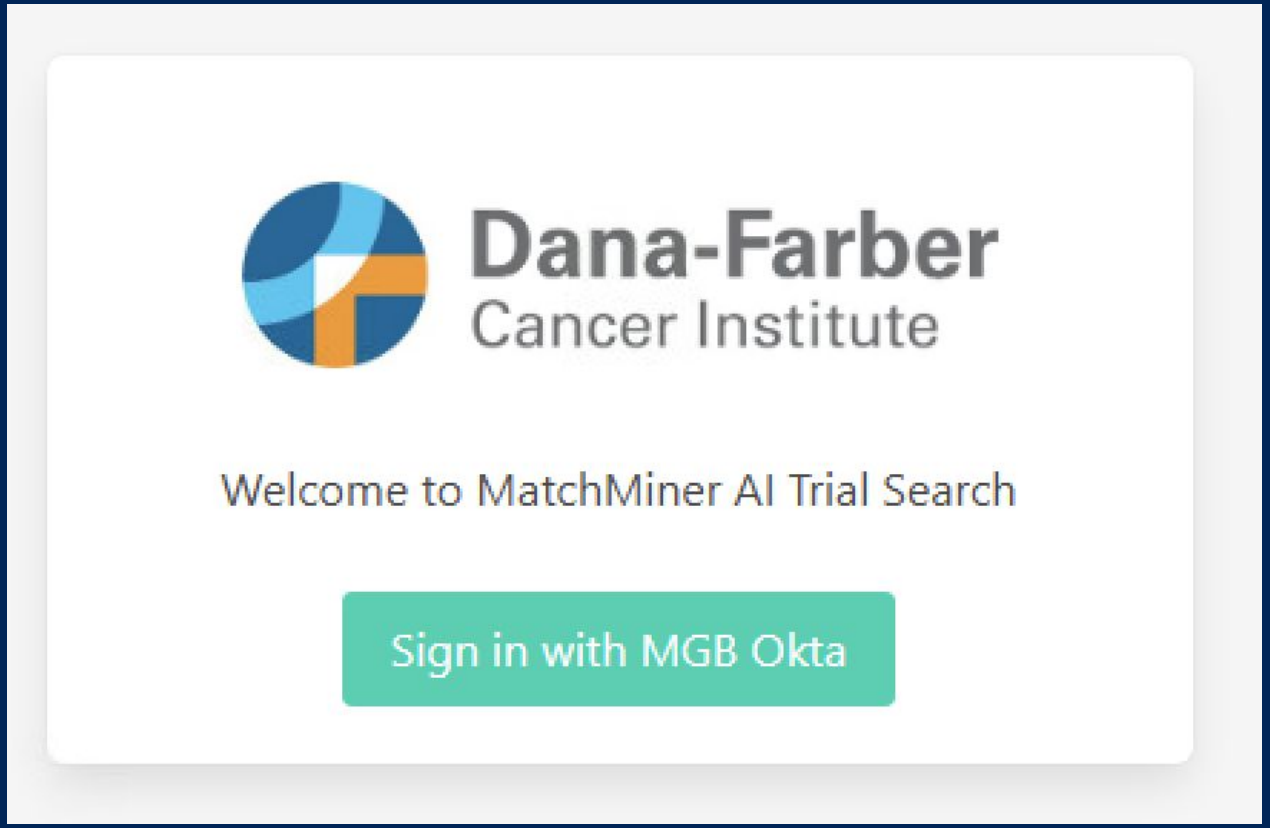
Dual Access Users (MatchMiner Genomics + MatchMiner-AI)



Using MatchMiner-AI

1 Access via EPIC.

MatchMiner-AI



Using MatchMiner-AI

2 Access via Website.

1.

https://ksg.dfci.harvard.edu/ai_trial_search

⚠️You must be on site or on VPN to access the website

Using MatchMiner-AI

Site Navigation

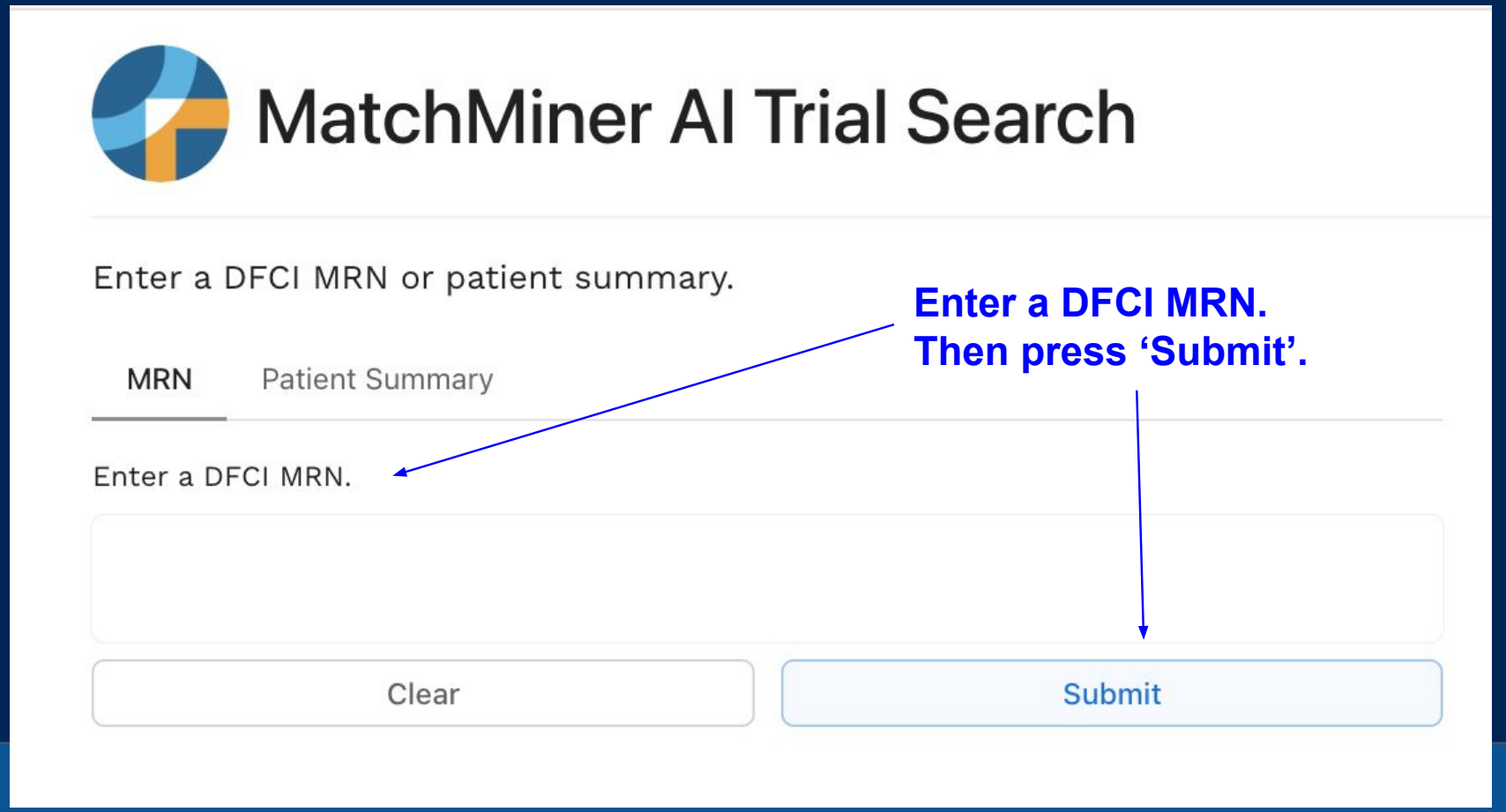
1. Patient Summary
2. Trial Suggestions and Details
3. Filtering Results

Using MatchMiner-AI

1 Patient Summary

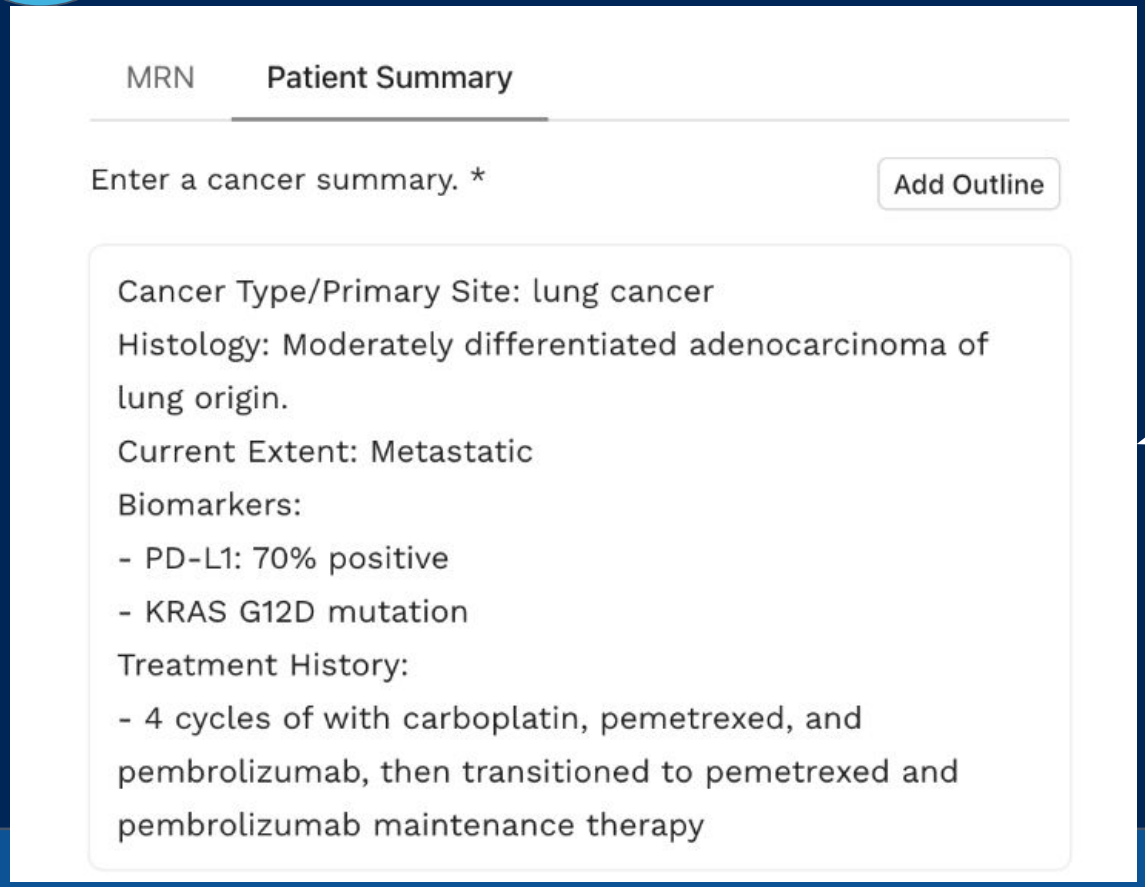
Website Access

EPIC will take you directly to the patient summary



Using MatchMiner-AI

1 Patient Summary



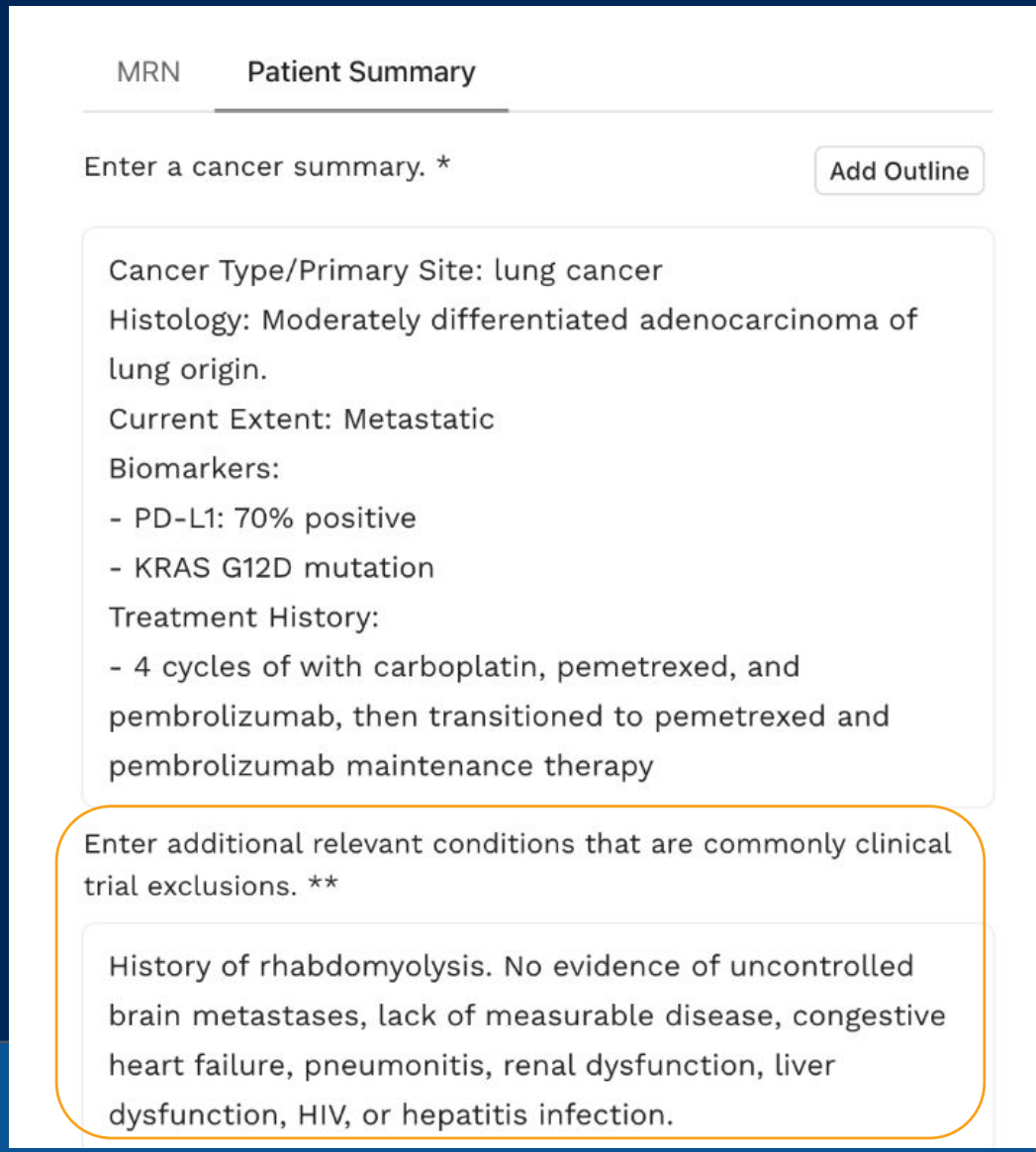
Scroll to read whole summary.

You can also provide feedback on the AI summary.

Using MatchMiner-AI

1 Patient Summary

AI extracted common trial exclusions for the patient

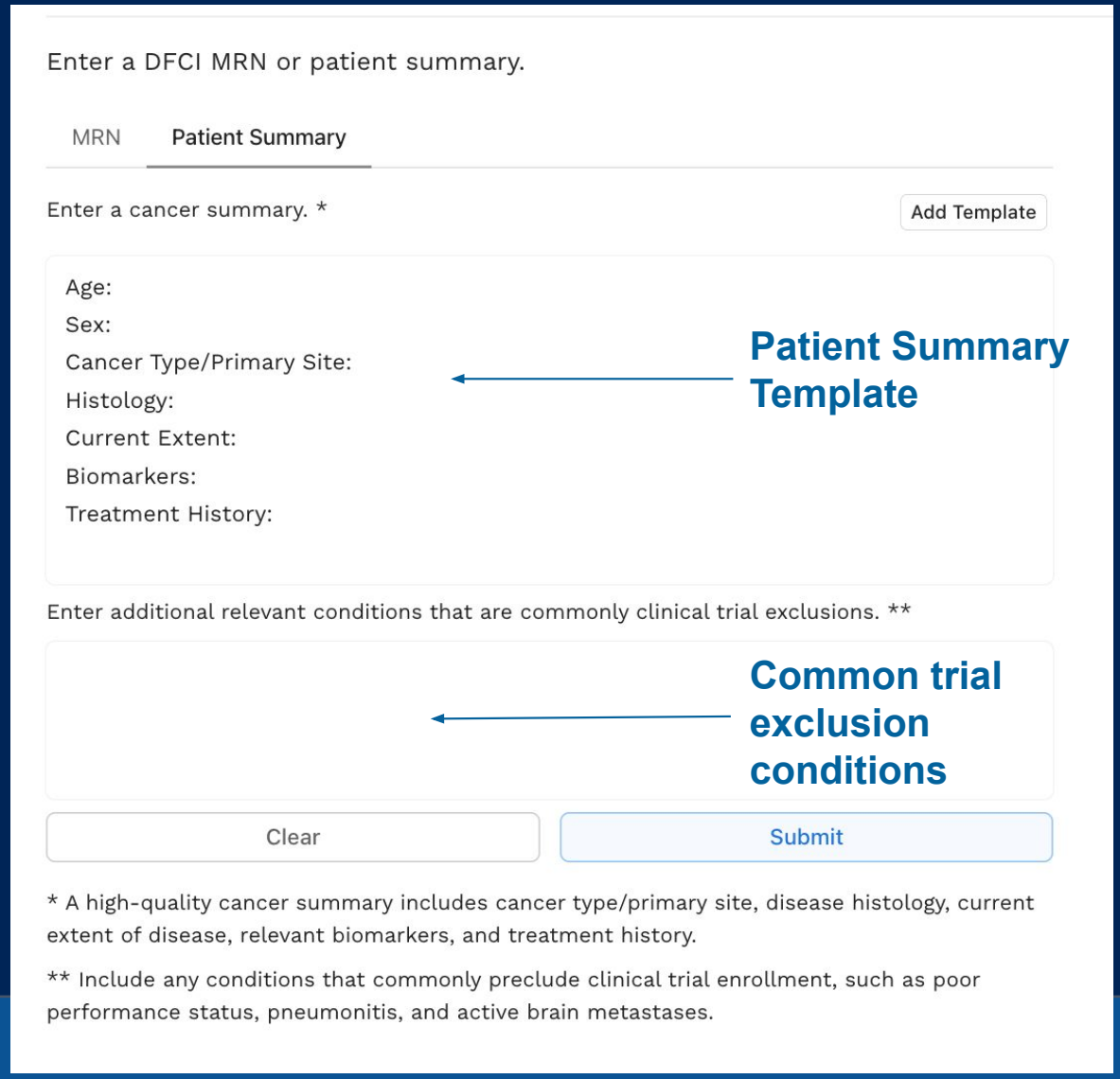


Using MatchMiner-AI

1 Patient Summary

Modify a summary

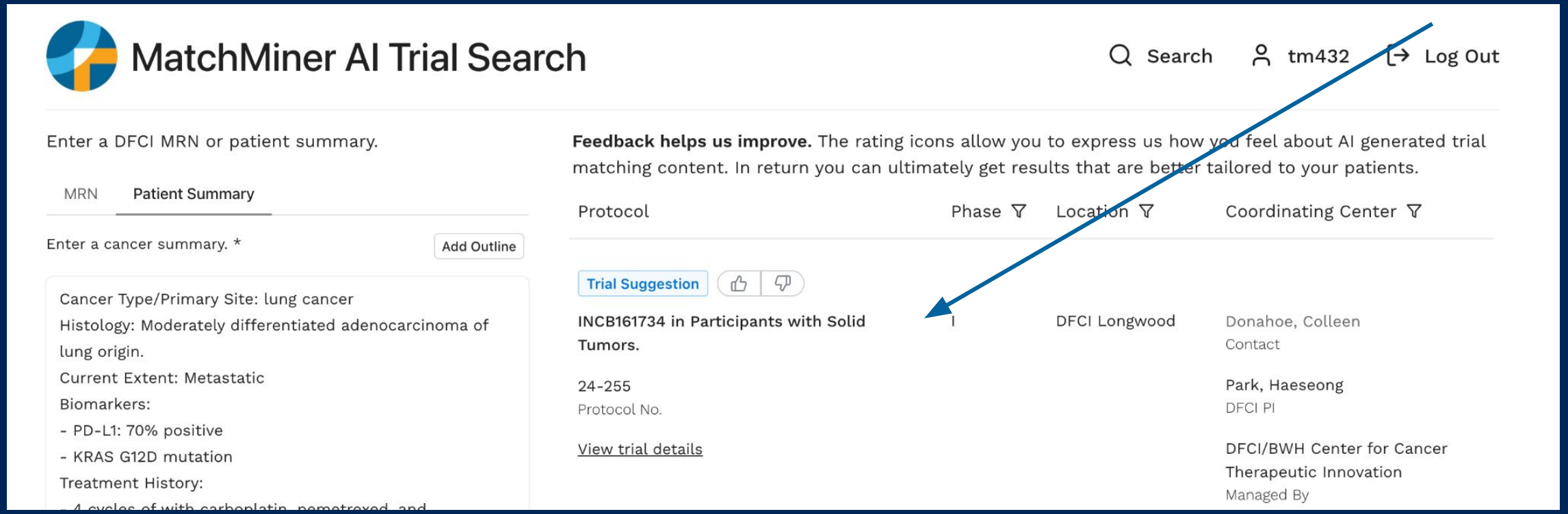
- click into the appropriate box
- add or change information
- press submit



Using MatchMiner-AI

2 Trial Suggestions and Details

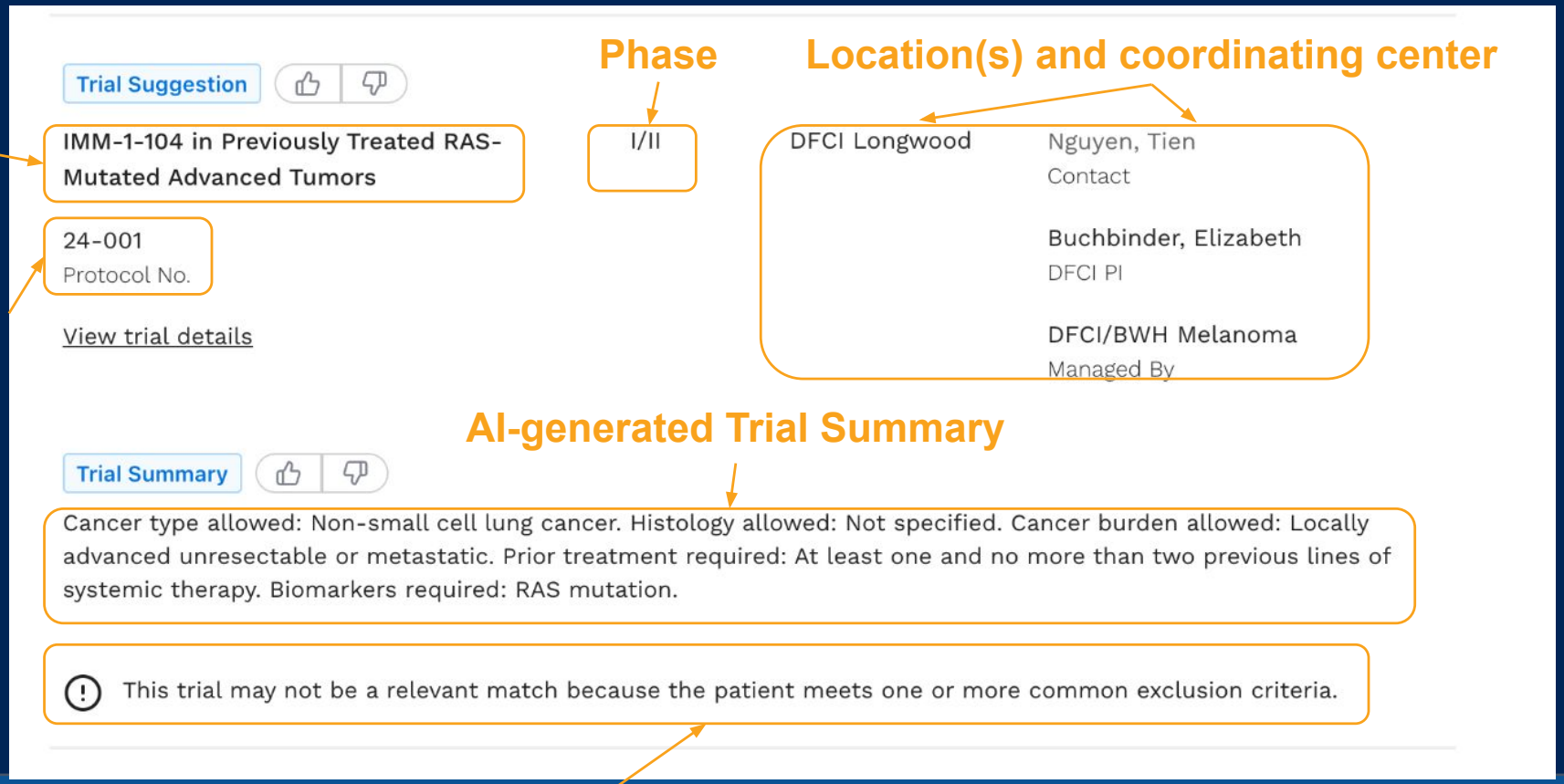
Trial suggestions



Using MatchMiner-AI

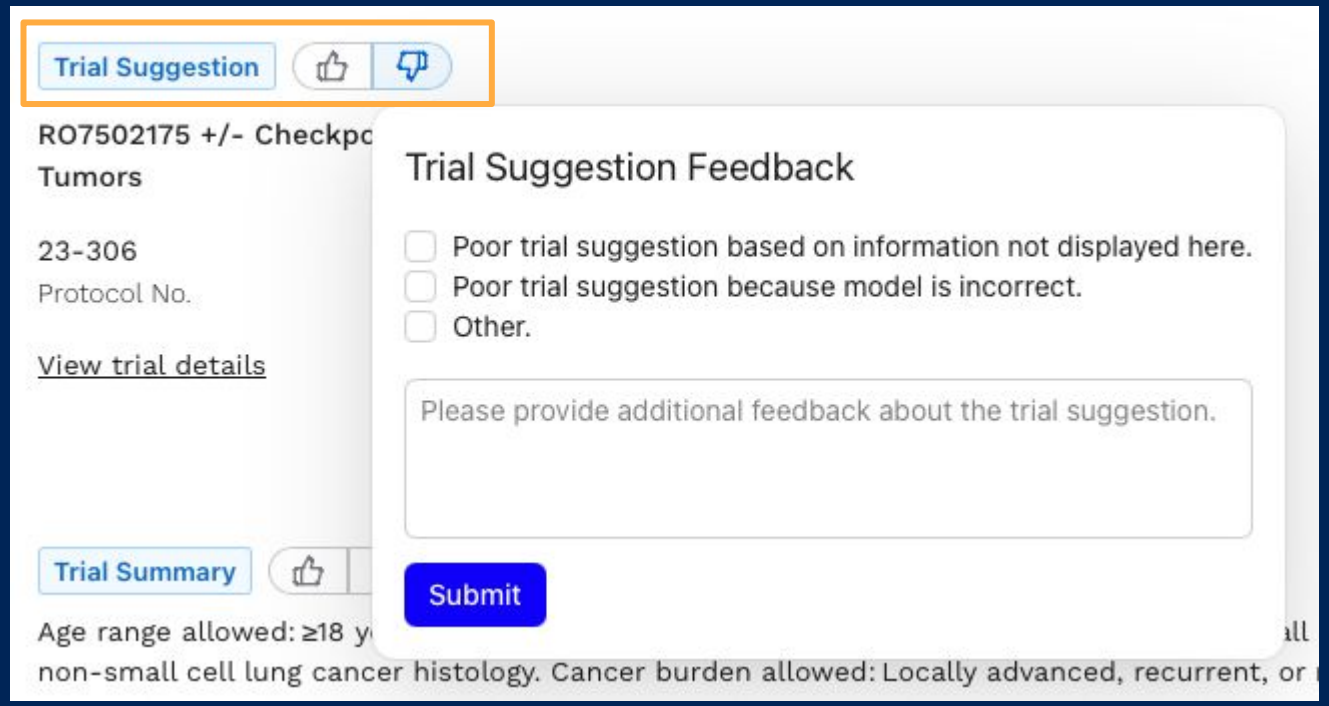
2

Trial Suggestions and Details



Using MatchMiner-AI

2 Trial Suggestions and Details

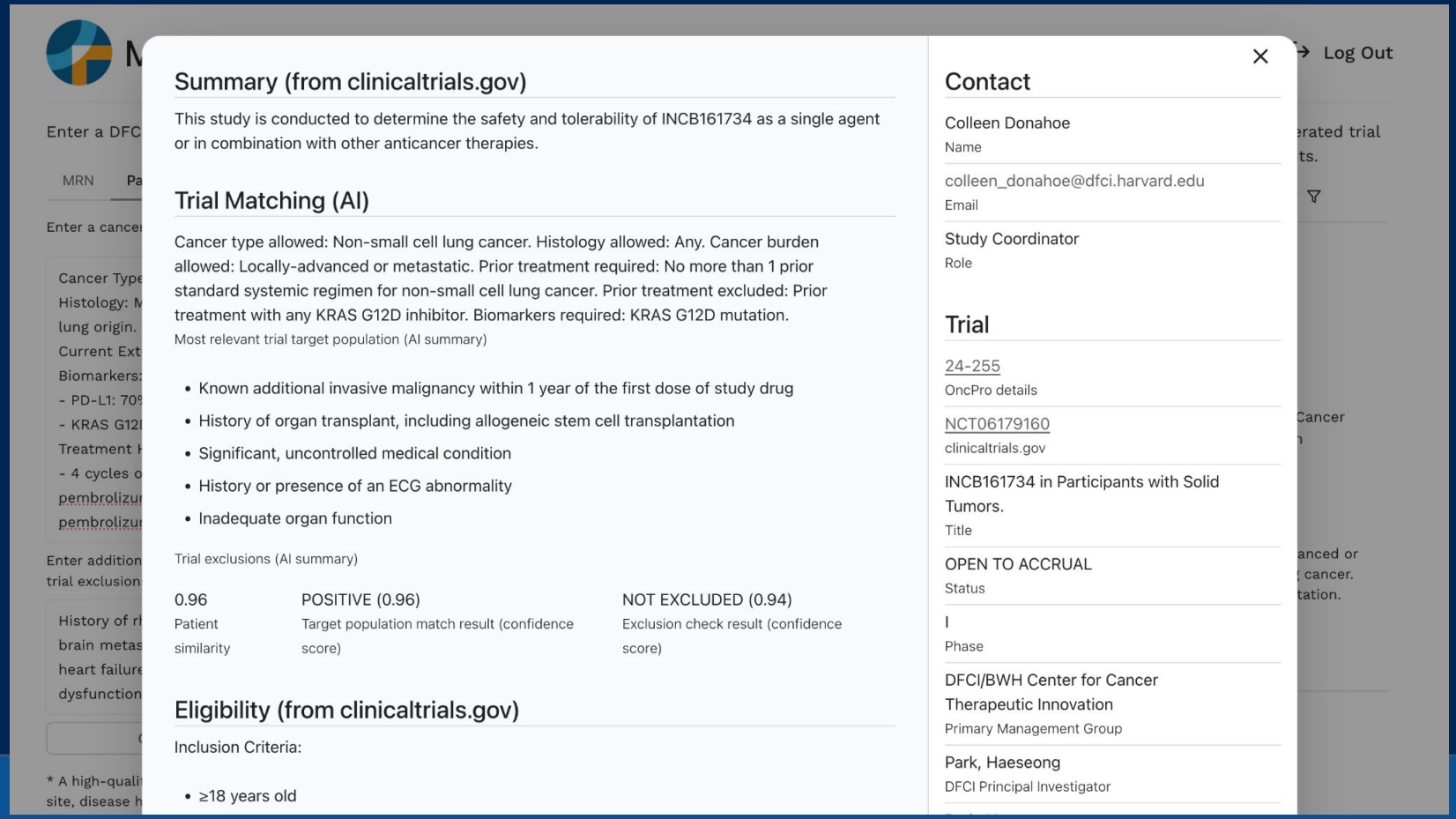


Using MatchMiner-AI

2

Trial Suggestions and Details

View Details



Using MatchMiner-AI

3 Filtering Results

Feedback helps us improve. The rating icons allow you to express us how you feel about AI generated trial matching content. In return you can ultimately get results that are better tailored to your patients.

Protocol	Phase	Location	Coordinating Center
Trial Suggestion	I I/II II III IV		
CERTIS1 D8410C00001 IVRS 22-447 Protocol No. View trial details		DFCI Longwood	Partridge, Kathryn Contact Wen, Patrick DFCI PI DFCI/BWH Neuro-Oncology Managed By

Trial Summary

Age range allowed: NA. Sex allowed: both. Cancer type allowed: breast cancer. Histology allowed: invasive carcinoma, HER2-negative. Cancer burden allowed: recurrent locally advanced or metastatic disease, refractory to standard therapy or no standard therapy exists. Prior treatment required: NA. Prior treatment excluded: more than one prior line of PARP inhibitor–based regimen. Biomarkers required: loss-of-function mutation in BRCA1 or BRCA2 or PALB2 or RAD51C or RAD55D (germline or tumor). Biomarkers excluded: NA.

Using MatchMiner-AI

3 Filtering Results

Feedback helps us improve. The rating icons allow you to express us how you feel about AI generated trial matching content. In return you can ultimately get results that are better tailored to your patients.

Protocol	Phase	Location	Coordinating Center
Trial Suggestion	☐ I ☐ I/II ☐ II ☐ III ☐ IV		
CERTIS1 D8410C00001 IVRS		DFCI Longwood	Partridge, Kathryn Contact
22-447 Protocol No.			Wen, Patrick DFCI PI
View trial details			DFCI/BWH Neuro-Oncology Managed By

Select funnel icon

Then, choose which phase(s)

Trial Summary

Age range allowed: NA. Sex allowed: both. Cancer type allowed: breast cancer. Histology allowed: invasive carcinoma, HER2-negative. Cancer burden allowed: recurrent locally advanced or metastatic disease, refractory to standard therapy or no standard therapy exists. Prior treatment required: NA. Prior treatment excluded: more than one prior line of PARP inhibitor–based regimen. Biomarkers required: loss-of-function mutation in BRCA1 or BRCA2 or PALB2 or RAD51C or RAD55D (germline or tumor). Biomarkers excluded: NA.

Using MatchMiner-AI

Clinician Feedback

04

Many new commercial trial matching platforms coming online. Why not just adopt one of those?

Benefits of Open Source Software

Many new commercial trial matching platforms coming online. Why not just adopt one of those?

- By going open source, we can fully and objectively evaluate its impact.

Benefits of Open Source Software

Many new commercial trial matching platforms coming online. Why not just adopt one of those?

- By going open source, we can openly collaborate with other cancer centers.

Benefits of Open Source Software

Many new commercial trial matching platforms coming online. Why not just adopt one of those?

- By going open source, we can evolve our models, based on clinical feedback.

Benefits of Open Source Software

We have separate AI models that can detect progressive cancer based on imaging reports and progress notes

Research question: Will proactively notifying oncologists about trial options when their patients have disease progression help increase trial accrual?

Proactive Notifications

You may also receive these email notifications.

You may opt out, if you like.

Proactive Notifications

Email us at:
matchminer@dfci.harvard.edu

Please provide feedback on patient summaries and trial matches.

Please send us your ideas and feedback