



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

1

2

3

4

MatchMiner-AI

Using 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

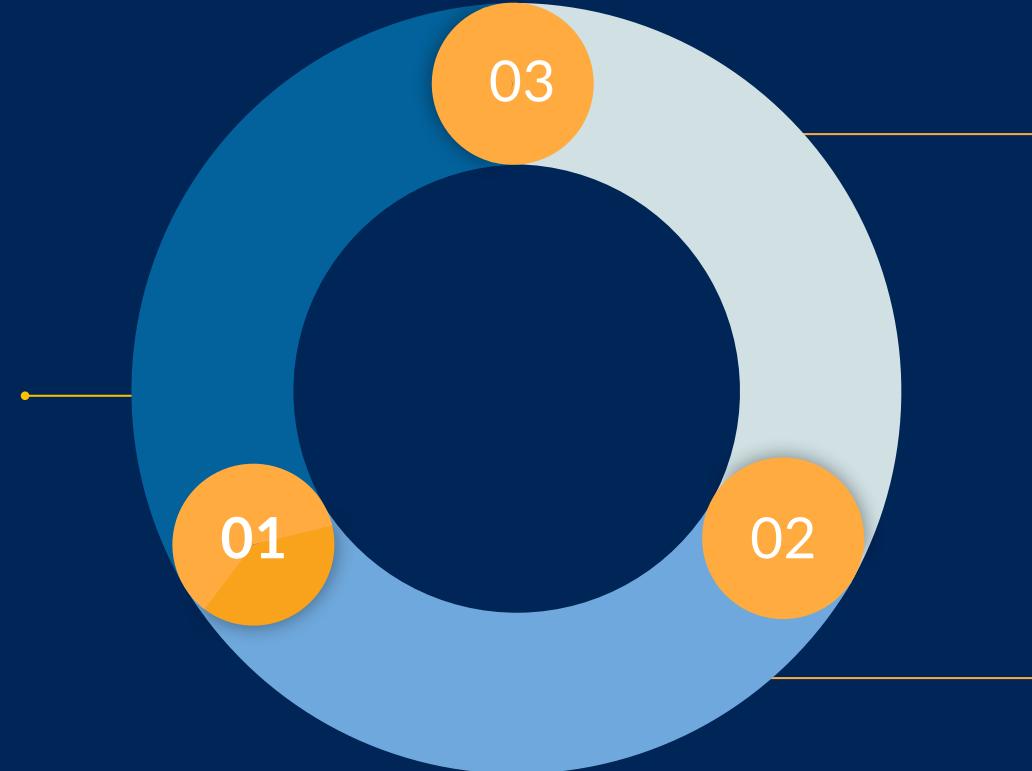
Not all trials are genomically driven.

~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

1

We focus primarily on core clinical criteria.

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



How it works

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.

Historical 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

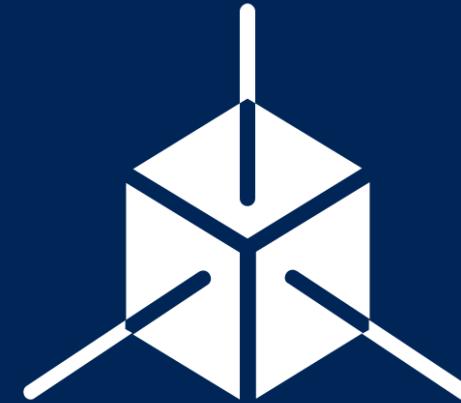
A three-step trial pre-screening tool.

a

Trial Space Model

Input:

Patient Summary
Clinical Trial Summary



How it works

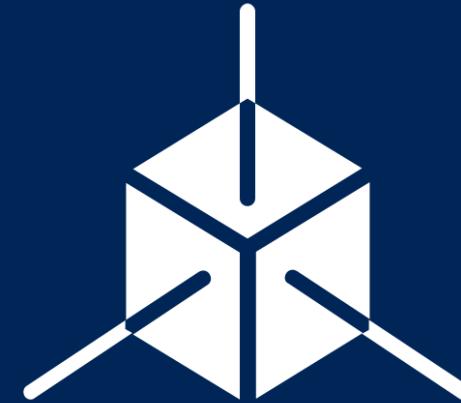
5

A three-step trial pre-screening tool.

a

Trial Space Model

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



How it works

5

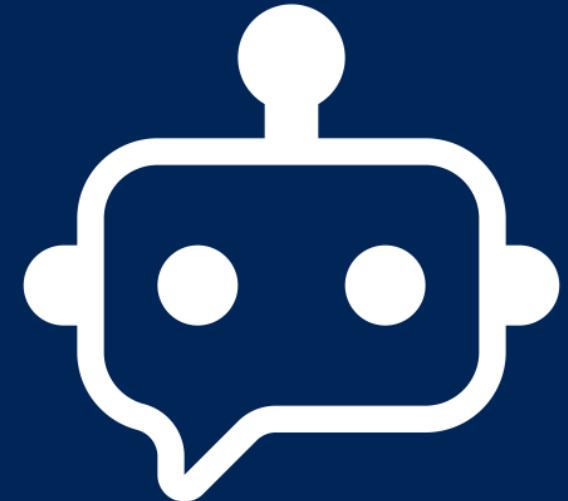
A three-step trial pre-screening tool.

b

Trial Checker Model

Input:

Patient Summary
Clinical Trial Summary



How it works

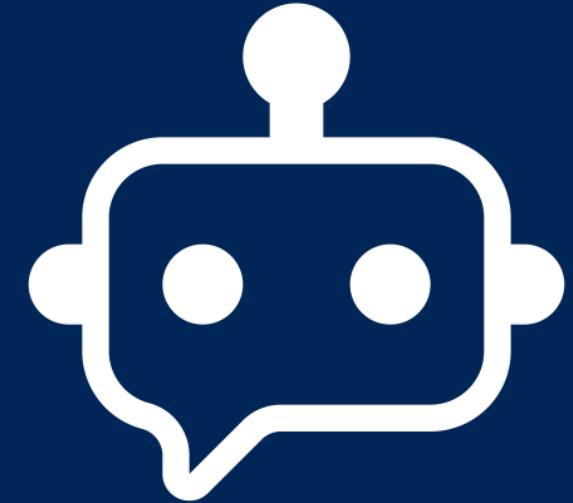
5

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

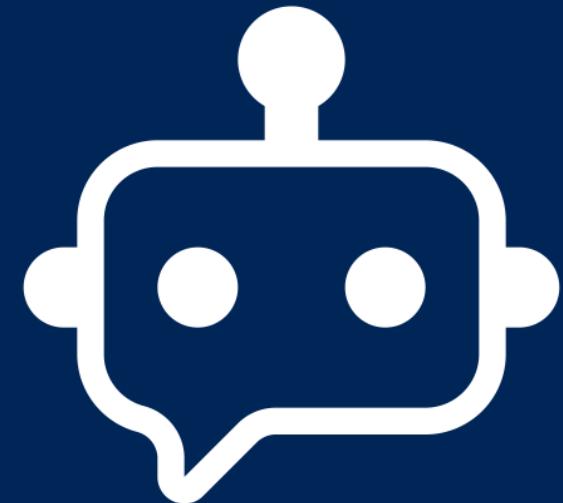
A three-step trial pre-screening tool.

c

“Boilerplate” Exclusions Model

Input:

Patient “Boilerplate” Summary
Clinical Trial “Boilerplate”
Summary



How it works

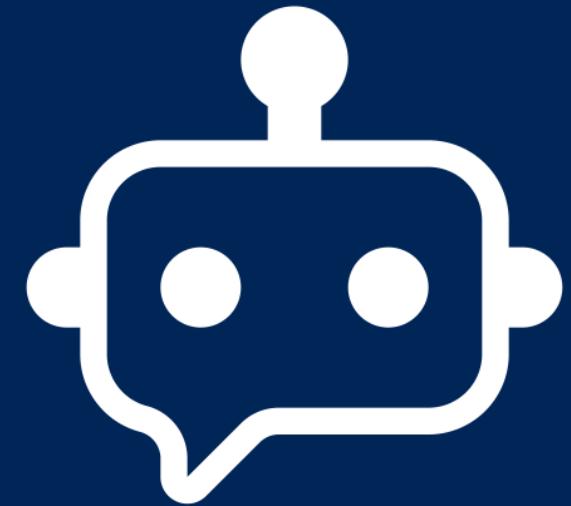
5

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

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](https://github.com/matchminerai/matchminerai). Examples of inference code are at [this https URL](https://github.com/matchminerai/matchminerai_inference). 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](https://github.com/matchminerai/matchminerai_demos).

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



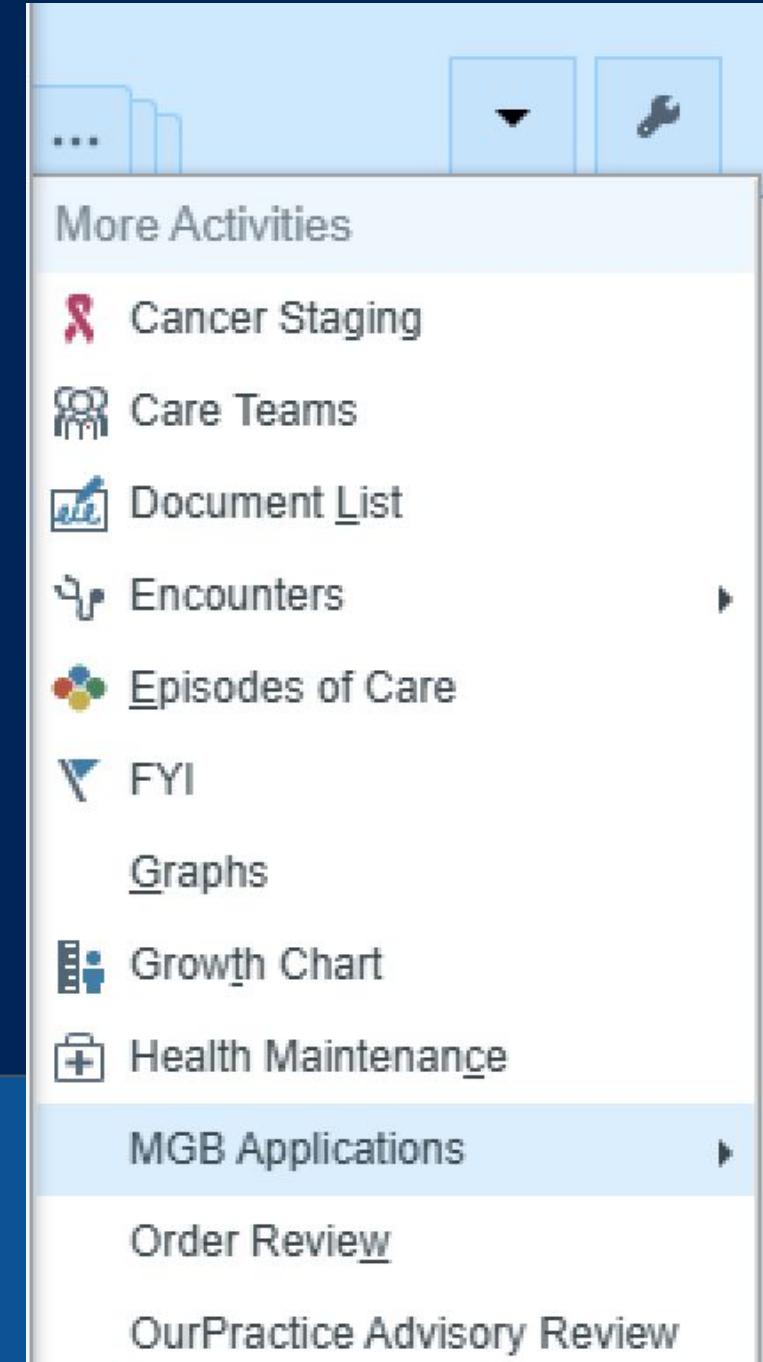
Using MatchMiner-AI

1

Access via EPIC.

→ MGB Applications

Using MatchMiner-AI



1

Access via EPIC.

MGB Applications



MatchMiner

Desens Worksheet Sidebar

DFCI Archer

DFCI Pathways

Digital Care

Emerson Chart

Emerson NE Hand Viewer

Family Doctors Viewer

Findhelp

Harmony (ECNS)

Harmony (MEE)

HMA Viewer

Infectious Disease

Rx InsightRX

MatchMiner

Medbridge

Solution Health Chart

SSH Chart

Using MatchMiner-AI

1

Access via EPIC.

Dual Access Users (MatchMiner Genomics + MatchMiner-AI)

EPIC launch detected. Choose an application.

Choose MatchMiner Experience

You have access to both MatchMiner Genomics and MatchMiner-AI.

About these tools

MatchMiner Genomics identifies potentially relevant genomically driven clinical trials for patients with available genomic sequencing by matching molecular and limited clinical features to trial eligibility criteria. MatchMiner Genomics can also be used to view genomic sequencing results.

[GO TO MATCHMINER GENOMICS](#)

MatchMiner-AI uses artificial intelligence to suggest potentially relevant clinical trials for *any* Dana-Farber patient, including those without genomic testing. Trial suggestions are generated from EHR data and predictive models to broaden trial discovery.

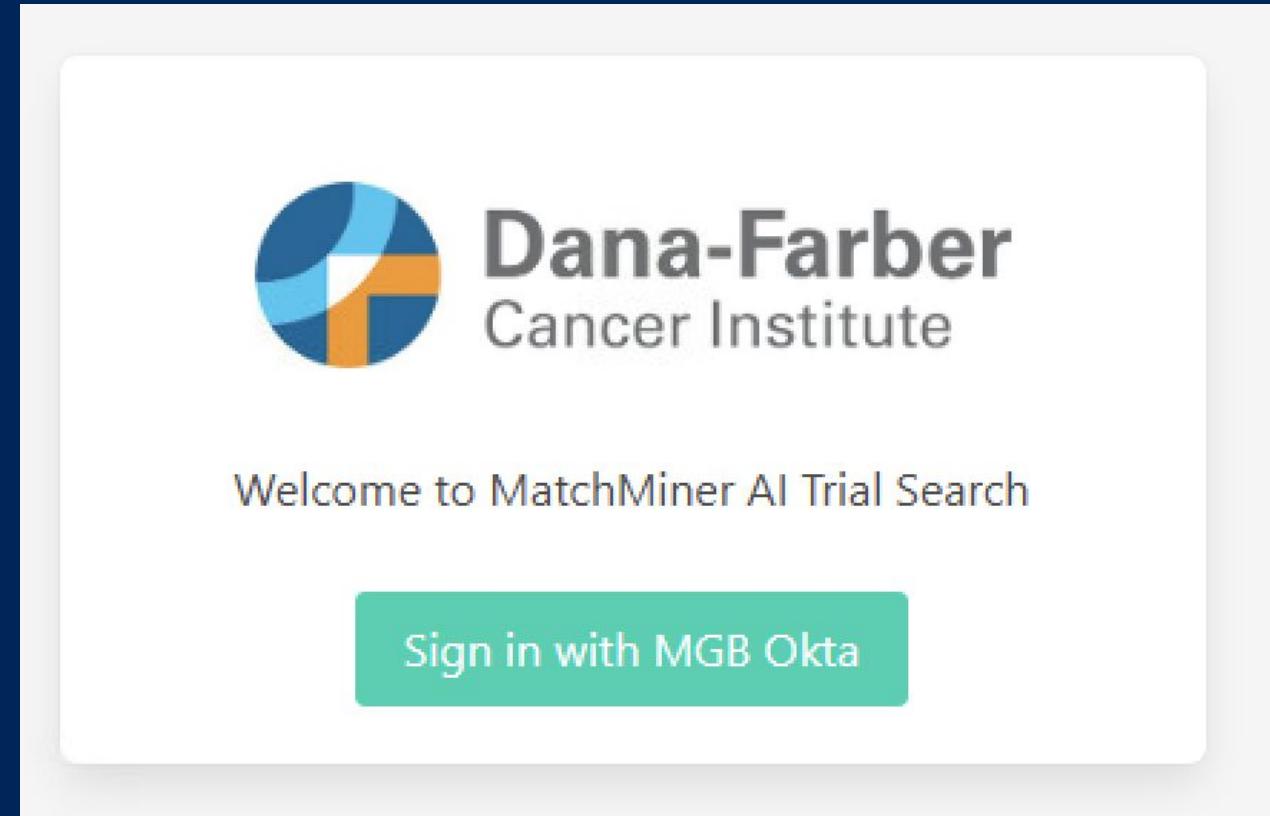
[GO TO MATCHMINER-AI](#)

MATCHMINER - DANA-FARBER CANCER INSTITUTE (2016 - 2026)

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

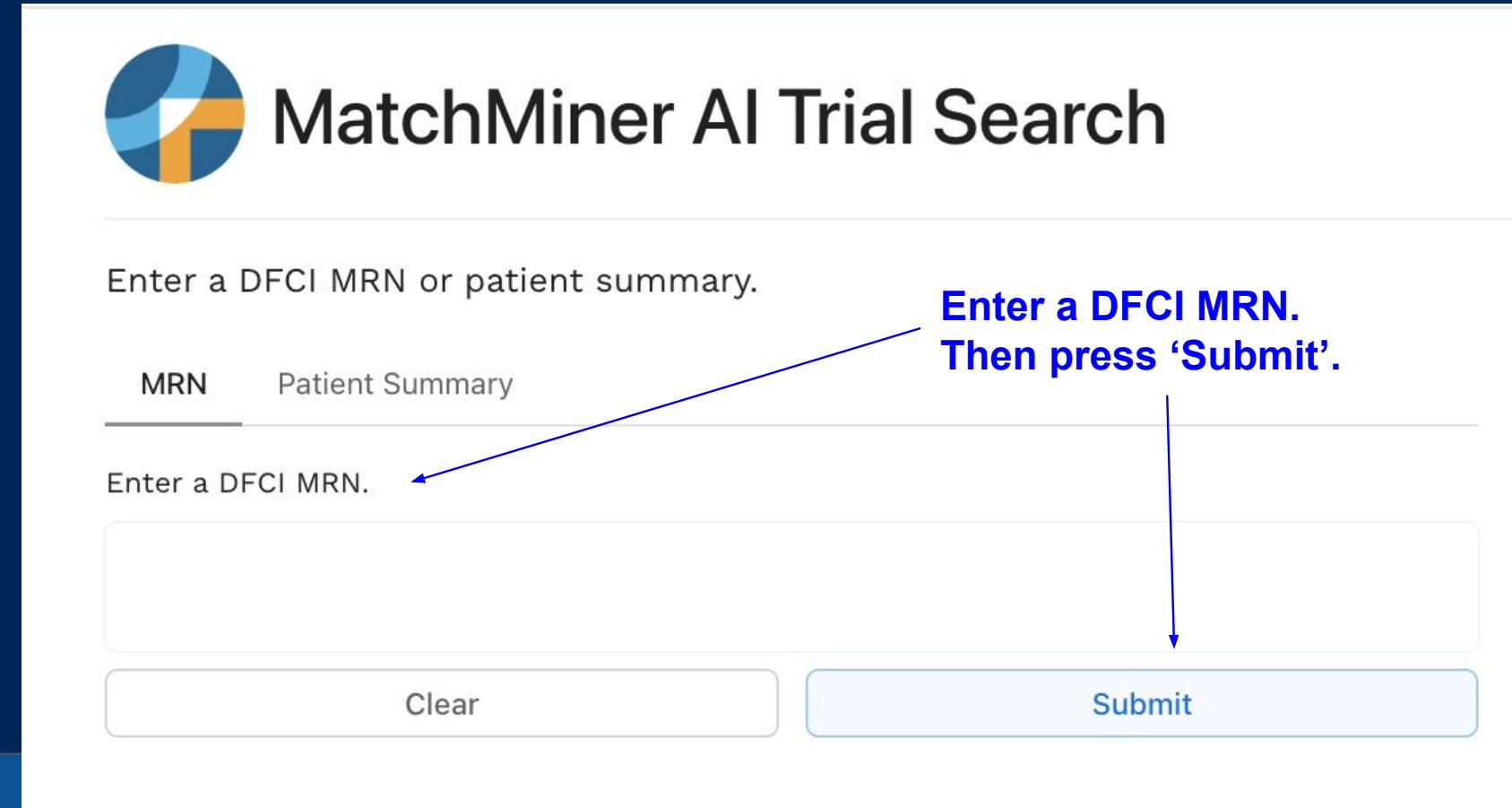
Using MatchMiner-AI

1

Patient Summary

Website Access

EPIC will take you directly to the patient summary



The image shows the MatchMiner AI Trial Search interface. At the top, there is a logo consisting of overlapping blue and orange circles. To the right of the logo, the text "MatchMiner AI Trial Search" is displayed in a large, bold, black font. Below the logo and text, there is a search input field with the placeholder "Enter a DFCI MRN or patient summary." To the left of the input field, there are two tabs: "MRN" and "Patient Summary". A blue arrow points from the text "Enter a DFCI MRN." to the "MRN" tab. Another blue arrow points from the text "Enter a DFCI MRN. Then press 'Submit'." to the "Submit" button. At the bottom of the interface, there are two buttons: "Clear" on the left and "Submit" on the right.

Enter a DFCI MRN or patient summary.

MRN Patient Summary

Enter a DFCI MRN.

Clear Submit

Enter a DFCI MRN.
Then press 'Submit'.

Using MatchMiner-AI

1

Patient Summary

MRN Patient Summary

Enter a cancer summary. *

Add Outline

Cancer Type/Primary Site: lung cancer

Histology: Moderately differentiated adenocarcinoma of lung origin.

Current Extent: Metastatic

Biomarkers:

- PD-L1: 70% positive
- KRAS G12D mutation

Treatment History:

- 4 cycles of carboplatin, pemetrexed, and pembrolizumab, then transitioned to pemetrexed and pembrolizumab maintenance therapy

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



MRN Patient Summary

Enter a cancer summary. *

Add Outline

Cancer Type/Primary Site: lung cancer

Histology: Moderately differentiated adenocarcinoma of lung origin.

Current Extent: Metastatic

Biomarkers:

- PD-L1: 70% positive
- KRAS G12D mutation

Treatment History:

- 4 cycles of carboplatin, pemetrexed, and pembrolizumab, then transitioned to pemetrexed and pembrolizumab maintenance therapy

Enter additional relevant conditions that are commonly clinical trial exclusions. **

History of rhabdomyolysis. No evidence of uncontrolled brain metastases, lack of measurable disease, congestive heart failure, pneumonitis, renal dysfunction, liver dysfunction, HIV, or hepatitis infection.

Using MatchMiner-AI

1

Patient Summary

- ## Modify a summary
- click into the appropriate box
 - add or change information
 - press submit

Enter a DFCI MRN or patient summary.

MRN Patient Summary

Enter a cancer summary. *

Add Template

Age:
Sex:
Cancer Type/Primary Site:
Histology:
Current Extent:
Biomarkers:
Treatment History:

Patient Summary Template

Enter additional relevant conditions that are commonly clinical trial exclusions. **

Common trial exclusion conditions

Clear Submit

* A high-quality cancer summary includes cancer type/primary site, disease histology, current extent of disease, relevant biomarkers, and treatment history.

** Include any conditions that commonly preclude clinical trial enrollment, such as poor performance status, pneumonitis, and active brain metastases.

Using MatchMiner-AI

2

Trial Suggestions and Details

Trial suggestions

MatchMiner AI Trial Search

Enter a DFCI MRN or patient summary.

MRN Patient Summary

Enter a cancer summary. *

Add Outline

Cancer Type/Primary Site: lung cancer
Histology: Moderately differentiated adenocarcinoma of lung origin.
Current Extent: Metastatic
Biomarkers:

- PD-L1: 70% positive
- KRAS G12D mutation

Treatment History:
4 cycles of carboplatin, paclitaxel, and

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
INCB161734 in Participants with Solid Tumors. 24-255 Protocol No.	Phase ▾	Location ▾	Coordinating Center ▾
View trial details	I	DFCI Longwood	Donahoe, Colleen Contact
		Park, Haeseong DFCI PI	DFCI/BWH Center for Cancer Therapeutic Innovation Managed By

Using MatchMiner-AI

2

Trial Suggestions and Details

Trial Title	Phase	Location(s) and coordinating center
HCC Identifier	I/II	DFCI Longwood Nguyen, Tien Contact Buchbinder, Elizabeth DFCI PI DFCI/BWH Melanoma Managed By
IMM-1-104 in Previously Treated RAS-Mutated Advanced Tumors		
24-001 Protocol No.		
View trial details		

AI-generated Trial Summary

Trial Summary

Cancer type allowed: Non-small cell lung cancer. Histology allowed: Not specified. Cancer burden allowed: Locally advanced unresectable or metastatic. Prior treatment required: At least one and no more than two previous lines of systemic therapy. Biomarkers required: RAS mutation.

(!) This trial may not be a relevant match because the patient meets one or more common exclusion criteria.

AI-assessment of patient exclusion criteria

Using MatchMiner-AI

2

Trial Suggestions and Details

Trial Suggestion  

RO7502175 +/- Checkpc
Tumors

23-306
Protocol No.

[View trial details](#)

Trial Summary  

Age range allowed: ≥18 years
non-small cell lung cancer histology. Cancer burden allowed: Locally advanced, recurrent, or metastatic

Trial Suggestion Feedback

Poor trial suggestion based on information not displayed here.
 Poor trial suggestion because model is incorrect.
 Other.

Please provide additional feedback about the trial suggestion.

Using MatchMiner-AI

Trial Suggestions and Details

View Details

The screenshot displays the MatchMiner-AI platform interface. On the left, there's a sidebar with fields for 'Enter a DFCI ID', 'MRN', 'Patient Name', and 'Enter a cancer type'. Below these are sections for 'Cancer Type', 'Histology', 'Current Extent', 'Biomarkers' (listing PD-L1: 70%, KRAS G12D: 100%), 'Treatment History' (mentioning 4 cycles of pembrolizumab), and 'Enter additional trial exclusion criteria' (with a history of brain metastases). At the bottom left, a note says '* A high-quality site, disease history'.

Summary (from clinicaltrials.gov)

This study is conducted to determine the safety and tolerability of INCB161734 as a single agent or in combination with other anticancer therapies.

Trial Matching (AI)

Cancer type allowed: Non-small cell lung cancer. Histology allowed: Any. Cancer burden allowed: Locally-advanced or metastatic. Prior treatment required: No more than 1 prior standard systemic regimen for non-small cell lung cancer. Prior treatment excluded: Prior treatment with any KRAS G12D inhibitor. Biomarkers required: KRAS G12D mutation. Most relevant trial target population (AI summary)

- Known additional invasive malignancy within 1 year of the first dose of study drug
- History of organ transplant, including allogeneic stem cell transplantation
- Significant, uncontrolled medical condition
- History or presence of an ECG abnormality
- Inadequate organ function

Trial exclusions (AI summary)

0.96	POSITIVE (0.96)	NOT EXCLUDED (0.94)
Patient similarity	Target population match result (confidence score)	Exclusion check result (confidence score)

Eligibility (from clinicaltrials.gov)

Inclusion Criteria:

- ≥18 years old

Contact

Colleen Donahoe
Name
colleen_donahoe@dfci.harvard.edu
Email
Study Coordinator
Role

Trial

24-255
OncPro details
NCT06179160
clinicaltrials.gov
INCB161734 in Participants with Solid Tumors.
Title
OPEN TO ACCRUAL
Status
I
Phase
DFCI/BWH Center for Cancer Therapeutic Innovation Primary Management Group
Park, Haeseong
DFCI Principal Investigator

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  	<input type="checkbox"/> I <input type="checkbox"/> I/II <input type="checkbox"/> II <input type="checkbox"/> III <input type="checkbox"/> IV	DFCI Longwood	Partridge, Kathryn Contact
22-447 Protocol No.			Wen, Patrick DFCI PI
View trial details			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  	Select funnel icon <input type="checkbox"/> I <input type="checkbox"/> I/II <input type="checkbox"/> II <input type="checkbox"/> III <input type="checkbox"/> IV	DFCI Longwood	Partridge, Kathryn Contact
22-447 Protocol No.			Wen, Patrick DFCI PI
View trial details			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

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