Mount Sinai Data Warehouse Town Hall

May 15, 2025

Icahn School of Medicine at Mount Sinai

The MSDW Team

Patricia Kovatch
Professor & Dean for Scientific Computing

Farhan Mahmood
Director MSDW

Sharon Nirenberg MD
Senior IT Director, Physician Informaticist

Naomi So MD
Physician Informaticist

Teja Ganta MD
Physician Informaticist

Darius Boopal
Healthcare Data Engineer

Christopher Veal
MSDW Technology Specialist

Jiani Xiang
Clinical Data Specialist

Timothy Quinn PhD
Principal Data Architect

Priscilla Acquaye
MSDW Support Analyst

Rupan Hossain
Database Administrator

Shivaji Punukollu
Database Administrator

Jacob Weiser
Healthcare Data Engineer

Agenda

1. **MSDW Operations**

2. **MSDW Major Accomplishments**
   - Integration of Somatic Genomic Results into Leaf
   - Digital Pathology Slide Diagnoses

3. **Epic for Research**

4. **MSDW Roadmap May 2025 – November 2025**

MSDW Operations

MSDW Data Sets Delivered

- From October 2024 to April 2025, the MSDW team closed 87 data requests

- There has been a significant decrease in the data set delivery time from October 2024 to April 2025 due to increased operational efficiency
  - Increased operational efficiency by implementing KPIs and SLAs

Top Department Requesting MSDW Data Sets

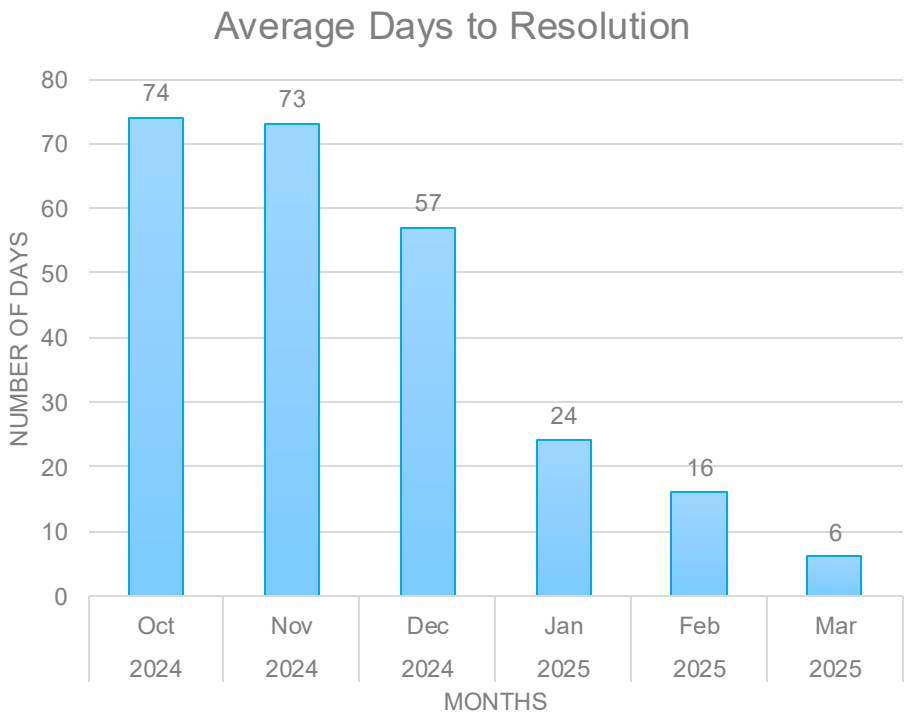
Apr 2024 – Sept 2024

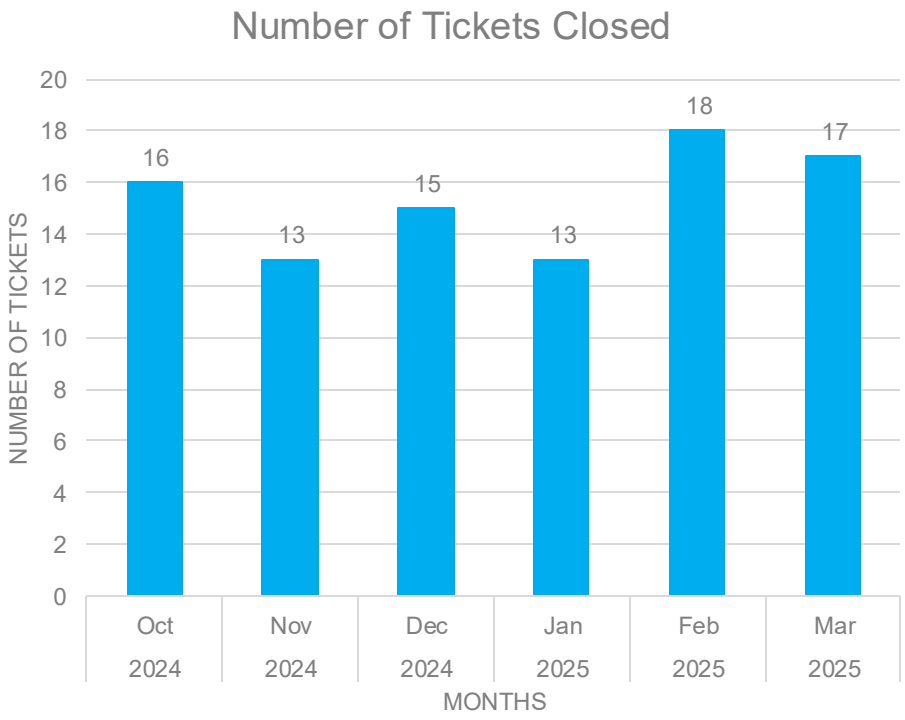
Department	Tickets (Number)
Population Health	15
Genetics and Genomics	10
Medicine	9
Medical Oncology	8
Psychiatry	7

Oct 2024 – Apr 2025

Department	Tickets (Number)
Medicine	22
Genetics and Genomics	16
Psychiatry	15
Population Health	14
Surgery	12

Custom Data Set Average Days to Resolution



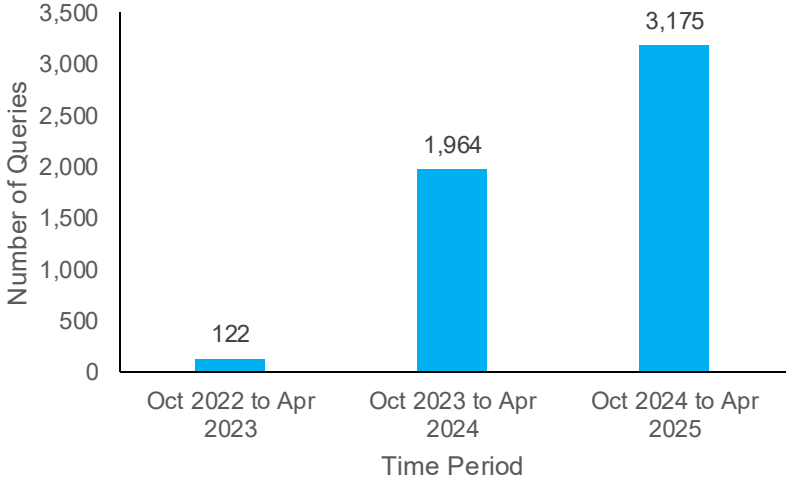


Leaf Utilization: Patient Cohorts & Custom Search Facets

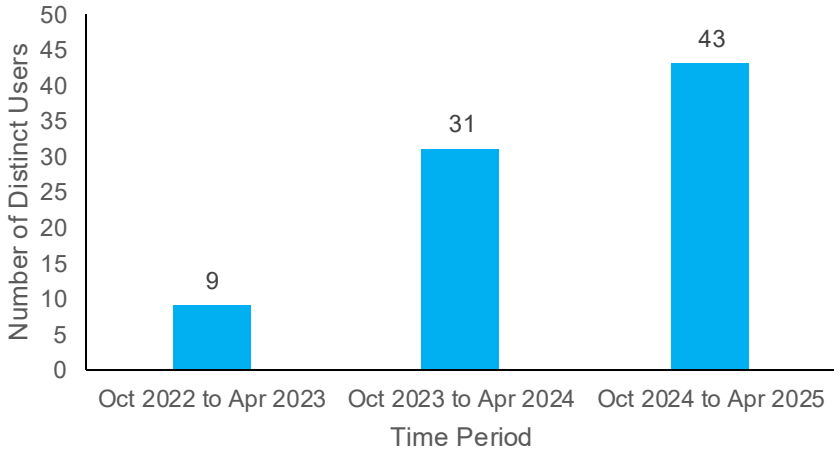
	Queries Executed (N)	
Patient Cohort / Feature	April 2024 – Oct 2024	Nov 2024 - Apr 2025
Cancer Staging from Cancer Registry	196	153
BioMe BioBank	102	102
Imaging Modality/ Body Part from IRW 2.0	NA	80
Digitized Pathology Slides Cohort	5	29
BioMe BioBank Global Diversity Array (Sema4)	6	16
BioMe BioBank Whole Exome Sequencing (Regeneron)	0	5
Imaging Research Warehouse 1.0	0	4
Cancer Patient Cohort	2	4
Cancer Institute Biorepository	14	3
Imaging Research Warehouse 2.0	9	1
BioMe BioBank Global Screening Array (Regeneron)	0	0
Total Leaf Queries	**797**	**1006**

TriNetX Utilization Continues to Increase

Number of Queries Run between October 2024 & April 2025 Compared to Previous Years

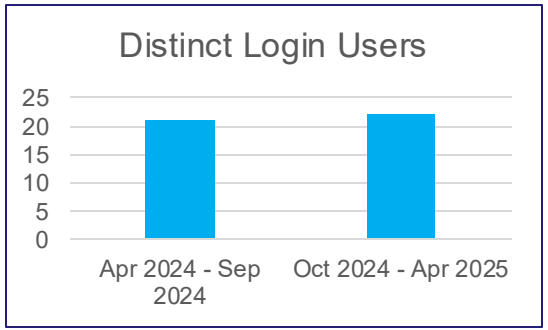


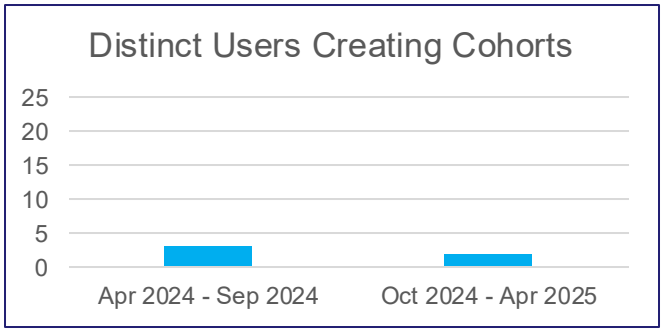
Number of Unique Users between October 2024 & April 2025 Compared to Previous Years

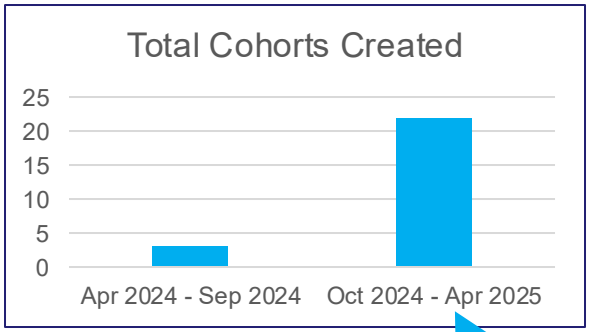


There has been a substantial increase in both users and queries run since the previous year's report

ATLAS Utilization: April 2024 – April 2025







A single user created 21cohorts in October 2024. This user has not logged in again since October 2024

- Utilization of ATLAS remains low with few users who are logging in creating and saving cohorts

Outreach to MSDW Users: April 2024 – April 2025

Event	Participants
Epic for Research Training	200
TriNetX Training Session	30
Leaf and ATLAS Training Session	36
Digital pathology training session	62
Leveraging Electronic Health Record (EHR) Data for Data Analysis	66
Digital Concierge	410
Total	623

MSDW Major Accomplishments

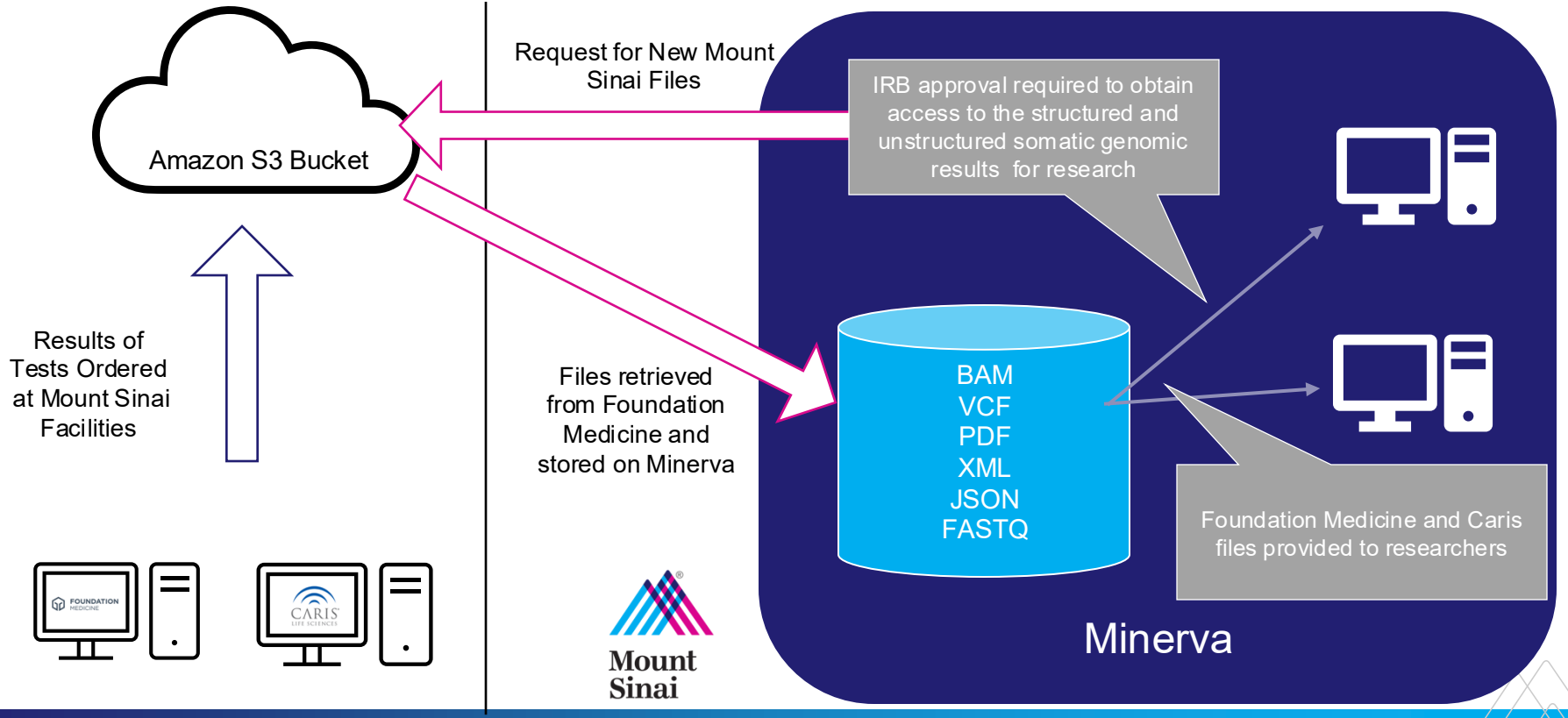
Obtaining Somatic Genomic Results from External Vendors

- **Project objectives**
  - To link the phenotypic and somatic genomic data on Minerva facilitating the use of somatic genomic data for analytics, clinical research and clinical operations
  - To make raw and structured somatic genomic results from external vendors available to the Mount Sinai research and clinical community

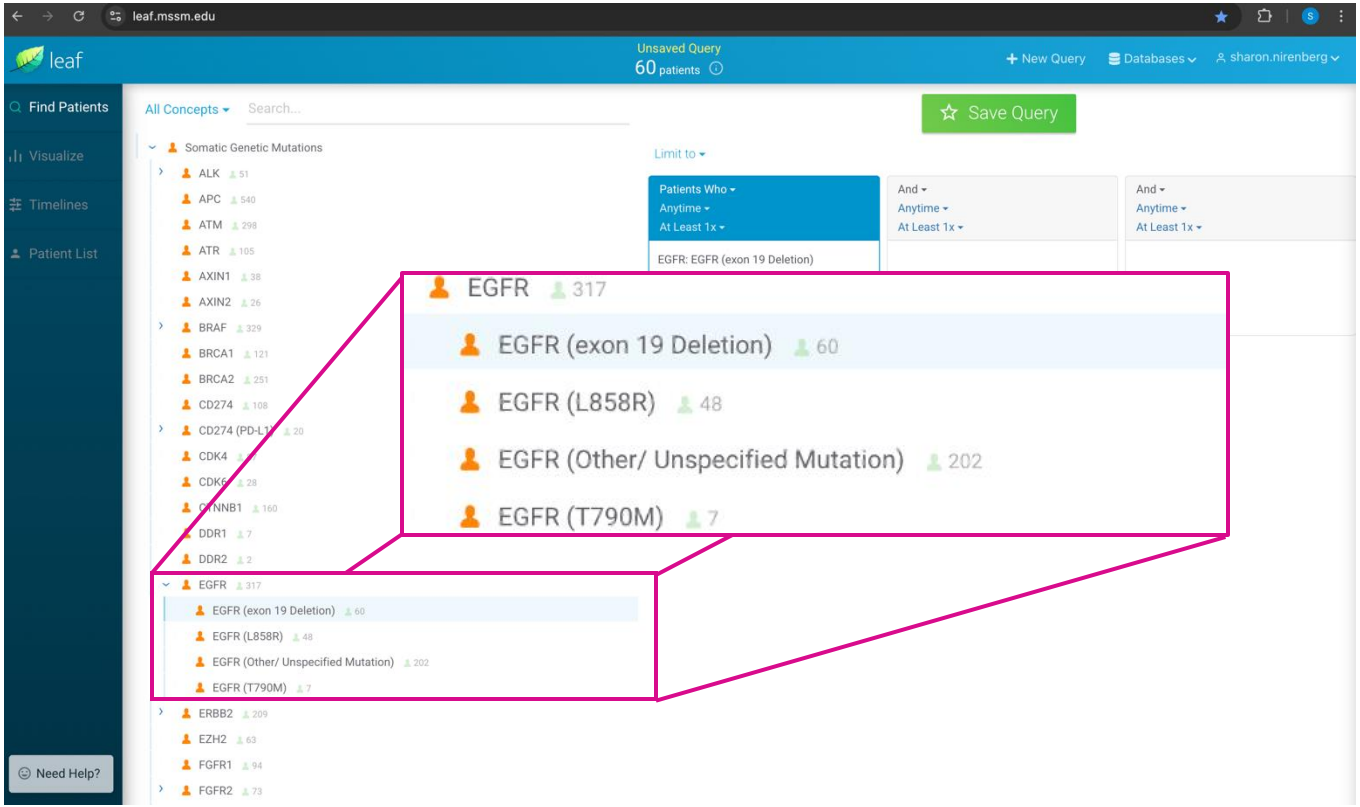
- **Collaboration with Mount Sinai Innovation Partners (MSIP)**
  - MSIP ensuring contracts address Mount Sinai's best interest for use of somatic genomic data

- **Genomic results returned to Minerva**
  - Results include both structured and raw genomic data
  - File formats received include BAM, FASTQ, VCF, PDF, XML, JSON and CSV
    - File types available vary by vendor

Data Flow from Somatic Genomic Testing Vendors to Mount Sinai



Search Select Genetic Mutations in Leaf



Pathogenic Gene Mutations Exposed in Leaf

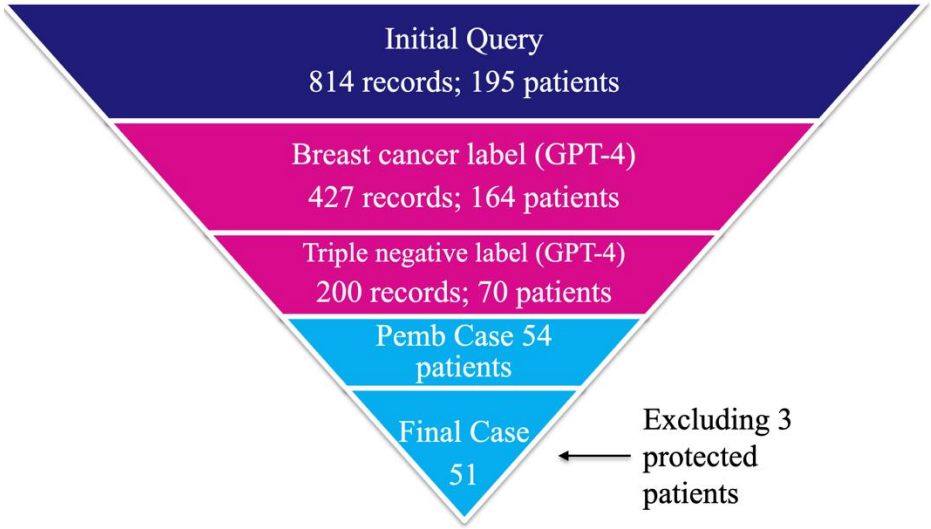
GENE	MUTATION DETAILS	GENE	MUTATION DETAILS	GENE	MUTATION DETAILS
ALK	*ELM4-ALK fusion*	**DDR1**		**MSI-H**	
APC		**DDR2**		**MTOR**	
ATM		**EML4**		**MYC**	*Amplification*
ATR		**EGFR**	*L858R, T790M, Exon 19 Deletion*	**NTRK1**	*NTRK1 fusion*
AXN1		**ERBB2**	*Amplification*	**NTKR2**	*NTRK2 fusion*
AXN2		**EZH2**		**NTRK3**	*NTRK3 fusion*
TCF7L2		**FGFR1**		**PIK3CA**	*E542K, E545K, H1047R*
RNF43		**FGFR2**	*FGFR2 fusion*	**PTEN**	
BRAF	*V600E, V600K*	**FGFR3**	*S249C*	**RET**	
BRCA1		**FLT1**		**RNF43**	
BRCA2		**FLT4**		**ROS1**	*ROS1 rearrangement*
CD274 (PD-L1)	*Amplification*	**IDH1**	*R132H*	**TCF7L2**	
CDK4		**IDH2**	*R140Q*	**TP53**	
CDK6		**KDR**		**VEGFA**	
CTLA4		**KRAS**	*G12C, G12D, G12V, Wild type*		
CTNNB1		**MET**			

Using AI to Identify Patients for Cohort Inclusion

Background: Using Large Language Models (LLM) to enhance cohort identification. Conventional methods (keywords, ICD codes, structured queries) struggle with nuanced clinical criteria.

Study goal: Identify **Stage 1,2,3 TNBC** patients who (1) received **neoadjuvant pembrolizumab** (2) underwent **definitive surgery**. **Exclude:** Stage 4 at diagnosis.

Solution: Used **GPT-4 via Azure API** (**HIPAA-compliant)** to process unstructured pathology reports and notes securely.



Epic for Research

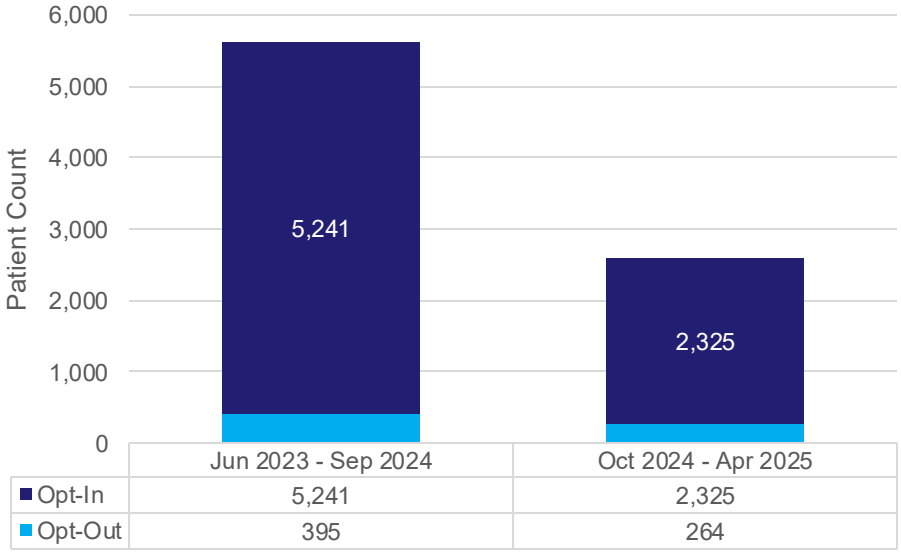
Use of Epic for Clinical Trial Recruitment

MyChart Recruitment and Clinical Trial Alerts are two Epic features for expanding clinical trial recruitment

MyChart Recruitment	Clinical Trial Alerts
• Individuals are identified as potentially eligible for a clinical trial via data stored in the electronic health record • Patient is sent a MyChart message to alert them of potential eligibility in a study • The patient can express or decline interest in the study via MyChart	• Individuals are identified as potentially eligible for a clinical trial via data stored in the electronic health record • The provider receives an alert that a specific patient may be eligible for a study • Via Epic, the provider can alert the study team if the patient is interested in participating in a study

	MyChart Recruitment	Clinical Trial Alerts
Live	13	8
In Progress	5	1

Metrics for MyChart Research Opt-Out



	Jun 2023 - Sep 2024	Oct 2024 - Apr 2025
Opt-In	5,241	2,325
Opt-Out	395	264

- In total 8,225 patients have responded to the Epic Research Consent, with 2,589 new responses since October 2024
- 659 (8.0%) of respondents have opted-out of being contacted via MyChart for research studies suggested by information in their electronic health record

MSDW Roadmap May 2025 – November 2025

MSDW Projects in Progress

	Project	Target Date	New Capabilities for Researchers
1	**Announce the ability to Identify patients with select somatic genetic mutations in Leaf**	2025-Q2	Enable researchers to obtain counts of patients with expanded list of somatic genetic mutations via Leaf
2	**Explore and prototype new approaches extract data Pathology Notes using AI**	2025-Q3	Facilitate medical research by providing quick access to relevant pathology data

MSDW Resources & Useful Links

MSDW Resources and Useful Links


Open a ticket: **https://labs.icahn.mssm.edu/msdw/open-a-ticket/**


Training and Tutorials: **https://labs.icahn.mssm.edu/msdw/training/**


About Leaf: **https://labs.icahn.mssm.edu/msdw/about-leaf/**


About ATLAS: **https://labs.icahn.mssm.edu/msdw/about-atlas/**

About TriNetX: **https://live.trinetx.com/**


Contact Us: **https://labs.icahn.mssm.edu/msdw/contact-us/**


Frequently Asked Questions: **https://labs.icahn.mssm.edu/msdw/faqs/**

Icahn School of Medicine at Mount Sinai

Thank You