Building AI-Assisted Biomarker Selection

Reducing Clinical Failure by Accelerating Biomarker Adoption in Trials with an Exhaustive and Intuitive Selection Platform.

BenchSci

Executive Summary

Problem

Databases for Biomarkers are Incomplete and Convoluted

Biomarkers dependably reduce clinical failures and advance the likelihood of approval throughout the drug development process; however, used in all phases in just 7.1% of clinical trials, biomarkers remain vastly underutilized. This phenomenon is rooted in a time-consuming and inefficient adoption process - one engendering sub-optimal biomarker selection. At the core of the pain point experienced by scientists are convoluted and incomplete databases - all of which house no more than 61% of biomarker-relevant information, failing to offer integrated insights and figures.

Solution

An Exhaustive, AI-Assisted Biomarker Selection Platform

We curate a comprehensive database of all open- and closed- access biomarker data by extracting from scientific publications and databases with NLP. BenchSci's existing proprietary ML models for textual analysis and image recognition are subsequently applied to decode and organization the information within decoded within curated resources. Through an intuitive biomarker selection platform integrated figures and insights, scientists can rapidly search for biomarker-related information for diverse range of workflows with a comprehensive portfolio of experimental filters.

Outcome

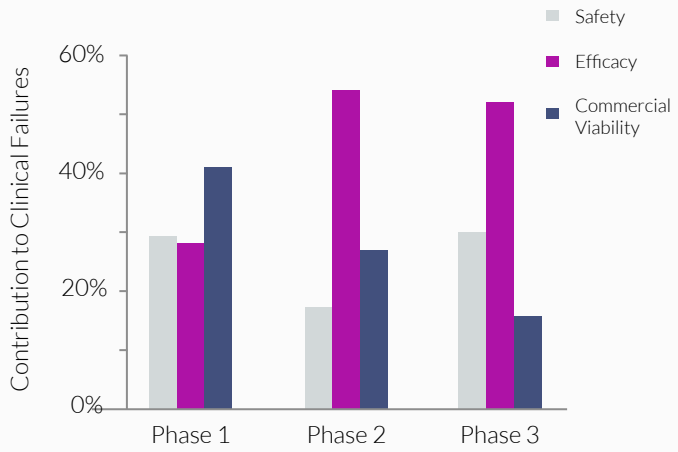
Empowering Scientists by Accelerating Biomarker Adoption

By empowering scientists to select the right biomarkers with velocity, BenchSci's biomarker selection program is well-positioned to exponentially increase the speed and quality of their life-saving research. By deliver an offering which synthesizes exhaustivity, an intuitive interface, and integrated insights, BenchSci can fill a glaring gap in an increasingly flourishing market, resolving a pain point in clinical development in the process and enabling scientists to deliver new medicines to patients 50% faster by 2025 with lower resource expenditure.

Inadequate Efficacy Derails Trials and Bloats Costs

Most Clinical Failures Derive from Efficacy Issues...

Efficacy-related complications adversely affect all stages of the drug development pipeline, driving **29%**, **54%**, and **52%** of failures in Phase I, Phase II, and Phase III trials respectively.[1]



...Which Drive Up Average Clinical Costs

Spurred largely by the lack of efficacy, the median cost of Phase I, II, and III clinical trials escalates to **$3.4**, **$8.6**, and **$21.4 million** respectively for a total of **$33.4 million** through Phase III.[2]

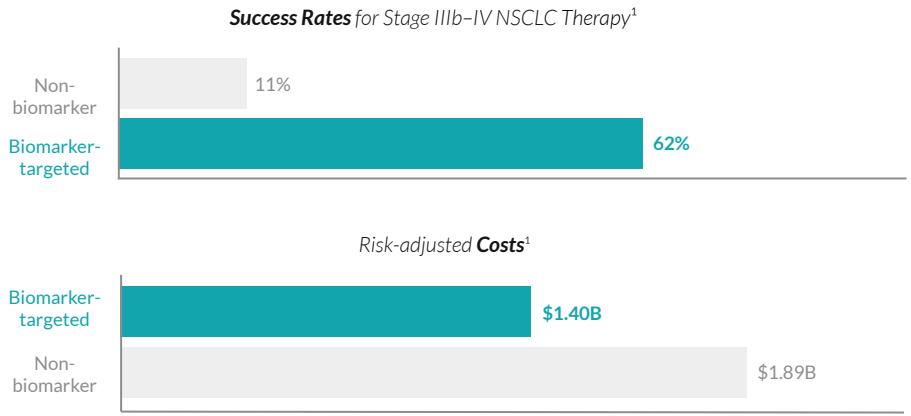
[1] Henderson, L. et al (2013) 'Reasons for Clinical Failures by Phase', Applied Clinical Trials, 22(12), pp. 12.
[2] Martin, L. et al. (2017) 'How much do clinical trials cost?', Nature Reviews Drug Discovery, 16(6), pp. 381–382.

With Biomarkers, Trials Succeed Sooner

Efficacy Biomarkers Cut Costs and Eliminate Uncertainty

In lieu of costly and prolonged resource expenditure in accruing sufficient information for clinical endpoints, a biomarker-driven approach enables scientists to predict a drug's expected clinical benefit more rapidly, allowing for shorter follow-up periods and smaller sample sizes, as illustrated in the figures to the right.

Furthermore, the usage of surrogate endpoints can drastically increase the comprehensive pass rate and drive down costs, as delineated in the case study of Stage IIIb–IV NSCLC Therapy.



True Endpoint Trial

A typical cardiovascular trial with true endpoints.[2]

Event	Endpoint	Sample Size	Length
Myocardial Infarction	Death	4000	5 yrs
Myocardial Infarction	Death	4000	5 yrs
Stroke	Stroke	25000	5 yrs

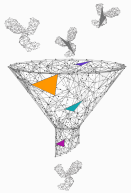
Surrogate Endpoint Trial

The trial with biomarkers as surrogate endpoints.[2]

Endpoint	Sample Size	Length
Coronary artery patency	200	90 min
Ejection fraction	30	2–4 wks
Diastolic blood pressure	200	1–2 yrs

[1] Falconi, A., Gilberto, P, & Parker, J. (2014) 'Biomarkers and Receptor Targeted Therapies Reduce Clinical Trial Risk in Non–Small-Cell Lung Cancer', Journal of Thoracic Oncology, 9(2), pp. 163–169.
[2] Wittes, J. & Lakatos, E. (1989) 'Surrogate endpoints in clinical trials: Cardiovascular diseases', Statistics in Medicine, 8, pp. 415–425.

Fail Early, Fail Fast – and Increase the Likelihood of Approval



Biological Markers Capture and Promote Clinical Benefit

Biomarkers are characteristics that are objectively measured and evaluated as an indicator of a biological process. In the setting of clinical trials, biomarkers generally aid in identifying populations for a study, monitoring therapeutic responses, and identifying side effects.

Predictive

Used to identify individuals more likely to respond – positively or adversely – to a particular drug.

Diagnostic

Determine if a patient has a condition covered by treatment or if they should be enrolled in a trial.

Prognostic

Measured at a defined baseline, they indicate an increased/lower likelihood of a future clinical event.

Monitoring

Assesses the status of a disease or for exposure evidence to a product or environmental agent.

Hyper-early Elimination of Unsuccessful Programs

Diagnostic companies face the challenge of coming up with assays that target specific sub-group-omics attributes of patients, while pharmaceutical companies need to target drugs at similar attributes (or attributes on the relevant pathways); for both parties, the economics must support smaller patient populations.[1]

Biomarkers enable these goals to be achieved, finding high levels of biological material that indicates no response or an adverse reaction. In doing so, CROs can fail drugs rapidly that don't meet a higher standard of efficacy provided by diagnostic and drug discovery techniques. With advances in the pharmaceutical and diagnostics field, earlier detection of failures will allow for a higher quality of trials to be pushed into the clinical pipeline, therefore procuring increased approval rates.

[1] Amplion (2017) 'Fail Early, Fail Fast – and Increase the Likelihood of Approval'.

Biomarkers Dependably Advance Clinical Development

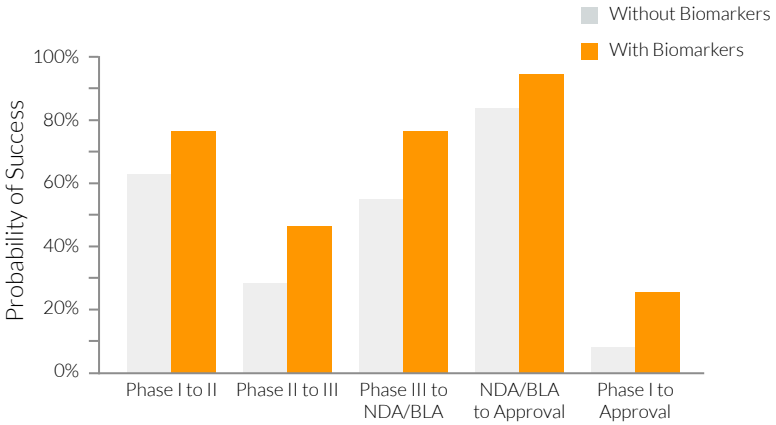
Biomarkers have Applicability at Every Stage in Clinical Trials

Phase I	Phase II	Phase III
• Identify surrogate biomarkers for toxicity and safety by protein expression analysis. • Define optimal dosage by maximizing utility functions. • Stratify and enrich the target population with response biomarkers.	• Discover prognostic biomarkers that inform on likely patient health outcome irrespective of drug treatment. • Discover predictive biomarkers to predict benefit of treatment compared to baseline patient condition.	• Monitor therapeutic side- effects or supra-therapeutic confounders and confirm the drug's response signature. • Assess the assay, its characteristics, and optical reproducibility conditions to validate analytical methods.

Beyond surrogate endpoints, biomarkers are capable of delivering advancements throughout drug discovery, encompassing everything from risk stratification in patients with previously untreated follicular lymphoma receiving anti–CD20-based therapy[1] to augmented prediction of prognosis in colorectal cancer.[2]

Biomarkers are Notably Used in Patient Selection

Biomarkers for patient enrollment (inclusion or exclusion criteria) have seen a dramatic increase in adoption since the genome was sequenced; phase transitions incorporating a selection biomarker for patient stratification promote the progress of success notably, engendering an increase in the LOA from Phase I to Approval from **8.4%** to **25.9%**.[3]



[1] Sohani, A. et al (2020) 'Biomarkers for Risk Stratification in Patients With Previously Untreated Follicular Lymphoma Receiving Anti-CD20-based Biological Therapy', American Journal of Surgical Pathology, 45(3), pp. 384–393.
[2] Bramsen, J. et al. (2017) 'Molecular-Subtype-Specific Biomarkers Improve Prediction of Prognosis in Colorectal Cancer', Cell Reports, 19(6), pp. 1268–1280.
[3] Thomas, D. et al. (2016) 'Clinical Development Success Rates 2006-2015', Bio Industry Analysis, pp. 18

Despite Their Utility, Biomarkers Remain Underutilized

Just 7.1% of Trials Use Biomarkers In All Stages[1]

Despite exhibiting numerous advantages, biomarkers remain heavily underutilized in clinical trials. Adopting a phase-by-phase approach, the literature delineates that only **7.1%** of all drug development paths use biomarkers in all phases, despite delivering a higher LOA and cost-based advancements across all clinical stages.

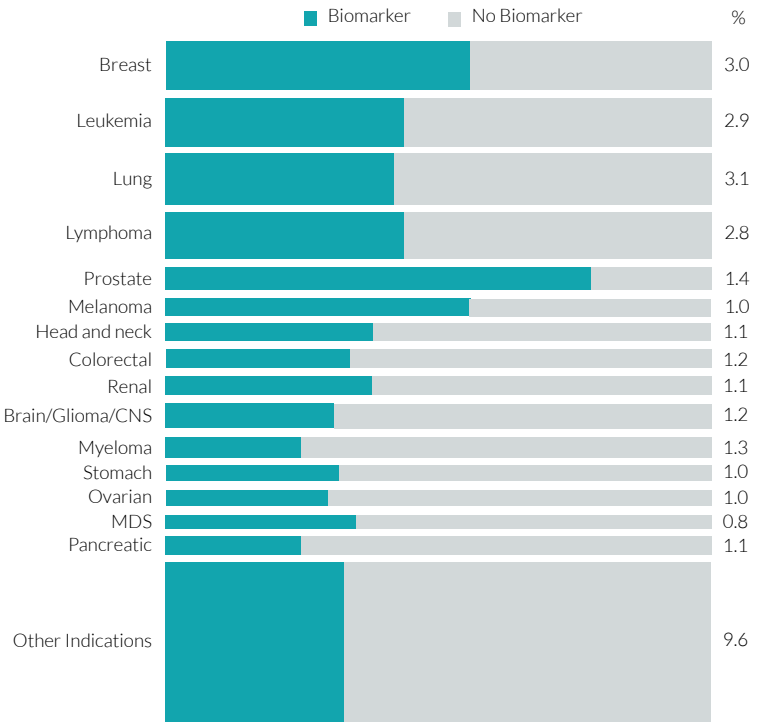
"Testing for these biomarkers is a necessary component of oncology medicine, as the results guide patients and their health care providers towards the most promising course of treatment. However, despite their importance, biomarker tests are underutilized."[2]

- Eli Lilly and Company

Only 34.7% of Oncology Trials Use Any Biomarker[3]

Biomarkers play a crucial role in cancer treatment, individualizing treatment and optimizing therapy. However, just **1,987** of **5,723** oncology trials in 2019 used a biomarker in any phase, as reflected in the figure (right), where box width and the secondary y-axis values capture the % of cumulative trials started between 2000 and 2018.

Oncology Clinical Trials, by Indication and Biomarker Use[4]



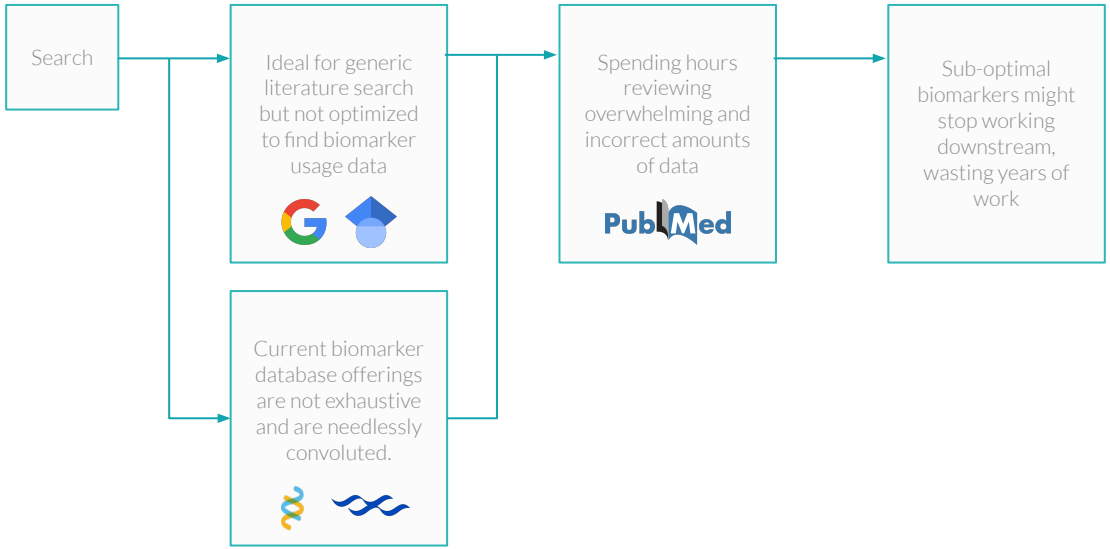
[1] Wong, C. & Siah, K. (2019) 'Estimation of clinical trial success rates and related parameters', Biostatistics, 20(2), pp. 273–286.
[2] Eli Lilly and Company (2020) 'Overcoming Barriers to Biomarker Testing in Cancer'.
[3] Patel, K. et al. (2020) 'Biomarker-driven efficiencies in clinical trials', Clarivate, pp. 14.
[4] Abrahams, E. et al (2021) 'The Evolution Of Biomarker Use In Clinical Trials For Cancer Treatment', Personalized Medicine Coalition, pp. 15.

Root Cause: The Search for Biomarkers is Long and Inefficient

Biomarker Selection is Currently Inefficient

This misalignment between utility and usage stems from the momentous difficulty in searching for an appropriate biomarker. The conventional method of finding a primary biomarker capable of working well in an experimental context entails ferreting around the literature (unoptimized for specific searches) and databases (thoroughly inexhaustive).

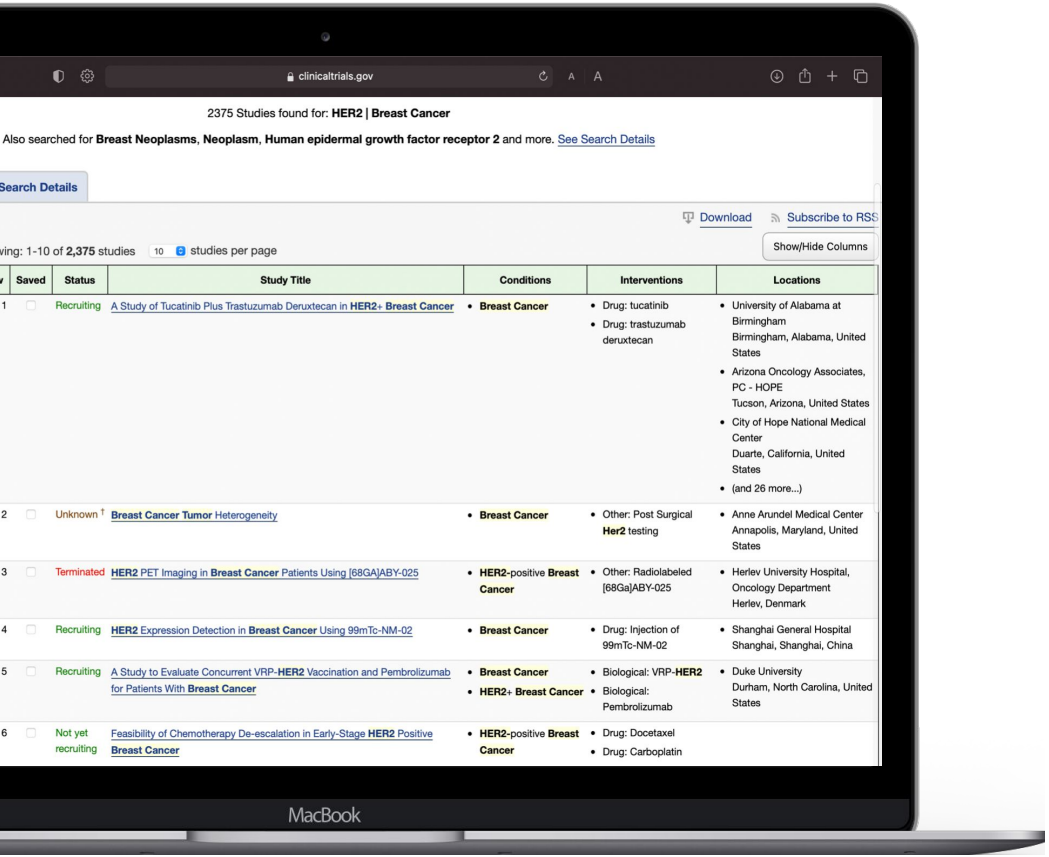
In doing so, scientists fritter away a colossal amount of time, reviewing an overwhelming amount of data, devoid of any reliable basis on which to rapidly select an ideal biomarker - a phenomenon which lends itself to an alarming corollary: if sub-optimal biomarkers are therefore selected, they could fail to perform as expected downstream, wasting years of hard work and squandering costly resources.



"Remember the last scene from Raiders of the Lost Ark where the Ark of the Covenant becomes lost in a sea of identical wooden crates inside a gigantic warehouse? The Ark of the Covenant [is like] how hugely valuable — but largely inaccessible — biomarker data is currently housed. Sure, it's there — but good luck finding it.

- Adam, Carroll, Amplion

[1] Carroll, A. (2015) 'How to Accelerate Outcomes of Biomarker Information Analysis'.



Pain Point: Biobanks are Complex and Incomplete

Tortuous Registries Contain <61% of Biomarkers[1]

Although the web's biomarker contents are considered collectively exhaustive, no single biobank contains over **61%** of biomarker-relevant information, impelling scientists into a nasty dilemma: save time by using a single database at the expense of anticipated biomarker performance, or optimize for performance by availing oneself of multiple sources, depleting an inordinate amount of time in the process. Irrespective of which choice is made, these engines are all widely regarded as difficult to navigate and intrinsically inconducive to optimal biomarker selection. Indeed, the synthesis of inaccessible search mechanics and inexhaustive contents incentivizes scientists to disregard biomarkers altogether.

"Right now you have to go into one of these databases, and it's like a labyrinth – you'd spend forever manually searching for the right biomarkers...a lot of these search engines are missing a huge amount of data anyway, so you are unlikely to find the right one."[1]

- Emma Eyer, Amplion

[1] Eyer, E. (16/07/2021, 14:17), 'An Exposition of Biomarker Databases.

Current Solutions Fail to Address this Problem

	GOBIOM	BiomarkerBase	charles river	BIONDA	Our Solution (Prospective) BenchSci
7,000+ Biomarkers	✓	✓	✓	✗	✓
8+ Data Source Types	✓	✓	✓	✗	✓
Easy-to-use Interface	✓	✓	✗	✗	✓
Varied Search Capability	✓	✗	✗	✓	✓
ML for Figure Prioritization	✗	✗	✗	✗	✓
Rapid, Integrated Insights	✗	✗	✗	✗	✓

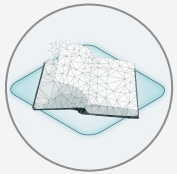
No Biomarker Offering on the Market Achieves Exhaustivity, Integration, and an Intuitive Interface

Recognizing the excruciating difficulty faced by scientists in identifying optimal biomarkers, some firms have thrown together solutions in an attempt to alleviate this pain point. However, although many of these databases are rich in quantity, they all lack functionality which would truly resolve the root cause of biomarkers' under-usage Most simply hyperlink to the original publication rather than offering integrated insights/figures, initiating yet further goose chases for an increasingly frustrated client. Furthermore, most simply cater to a single workflow, offering features "catered towards the business development side of things", rather than scientists in the thick of the drug development process.[1]

[1] Eyer, E. (16/07/2021, 14:17), 'An Exposition of Biomarker Databases.

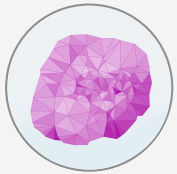
Solution Overview: An Exhaustive, AI-Assisted Biomarker Platform

1 Curate an Exhaustive Collection of Biomarker Information



First, we curate a comprehensive database of all open- and closed- access biomarker data by extracting from scientific publications, leveraging BenchSci's strategic partnerships, and confirming exhaustivity with independent validations. BenchSci's existing NLP Curation infrastructure can be repurposed to maximize breadth of coverage in trial extraction, avoid redundancy, and ensure that assimilated data is wholly relevant.

2 Decode and Organize the Data with Machine Learning



By applying BenchSci's proprietary ML models, biomarker specs and success can be decoded like a PhD biologist, and biomarkers can be linked to use cases and biomedically relevant concepts with advanced bioinformatics and ontologies. Furthermore, BenchSci's image recognition technology can be applied to unlock the value of biomarker-relevant figures, and elucidate image prioritization by linking captions and their corresponding figures.

3 Present Results and Insights in an Easy-to-use Interface

Through an intuitive biomarker selection platform, scientists can rapidly search for biomarker-related information with a comprehensive portfolio of experimental filters. By clicking on any given search result, the user is presented with all the relevant figures and basic details regarding the selected trial, alongside digestible insights and conclusions, enabling scientists to rapidly select the optimal biomarkers with confidence in a matter of seconds.

A Comprehensive Collection Mandates Multiple Sources

ClinicalTrials.gov[1]

A database of privately and publicly funded clinical studies conducted around the world.

Count	Type	Cost	Domain
38,074 Biomarkers	DX, PX, PD	Free	Public

MarkerDB[2]

A freely available electronic database consolidating information on all known clinical trials.

Count	Type	Cost	Domain
27,759 Biomarkers	PR, CH, GE, PX, DX, EX	Free	Private

EDRN Biomarkers[3]

A national network or the development, evaluation, and validation of cancer biomarkers.

Count	Type	Cost	Domain
17,915 Cancer Biomarkers	DX, PD, EF, EX	Free	Public

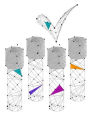
Key

Diagnostic: **DX**; Prognostic: **PX**;
Predictive: **PD**; Protein: **PR**;
Chemical: **CH**; Genetic: **GE**;
Exposure: **EX**; Efficacy: **EF**

An exhaustive compilation of open- and closed- access biomarker data is at the heart of a curation strategy. The bubbles (left) present information on 3 examples of prominent sources on biomarkers, while the following information summarizes the main 3 data sources:

Real-world experimental data from hundreds of thousands of scientific publications, including closed-access papers.

Leveraging BenchSci's existing strategic partnerships with leading scientific publishers (eg. Springer Nature, Wiley).



Independent validations from organizations to confirm exhaustivity & experimental validity (eg. Euromene, BmDR).

[1] ClinicalTrials.gov (Accessed: 21/07/2021, 18:40)
[2] MarkerDB.ca (Accessed: 21/07/2021, 18:47)
[3] EDRN.nci.nih.gov (Accessed: 21/07/2021, 18:53)

NLP Algorithms Mine Literature for Biomarker-based Trials

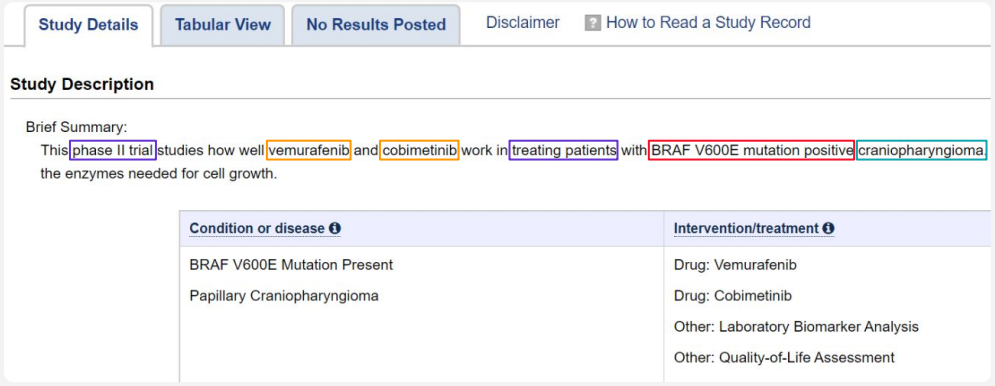
Curation NPL Maximizes Breadth of Coverage in Trial Extraction

Systematic reviews and meta-analyses of biomarker data are considered labor-intensive and time-consuming. By automating the extraction of quantitative data from primary studies with NLP algorithms, the curation of biomarker-based trials can be greatly accelerated.[1] Moreover, since BenchSci's NLP infrastructure is already optimized for the extraction of reagent-relevant data, refinement of existing algorithms to a biomarker context further reduces costs and time.

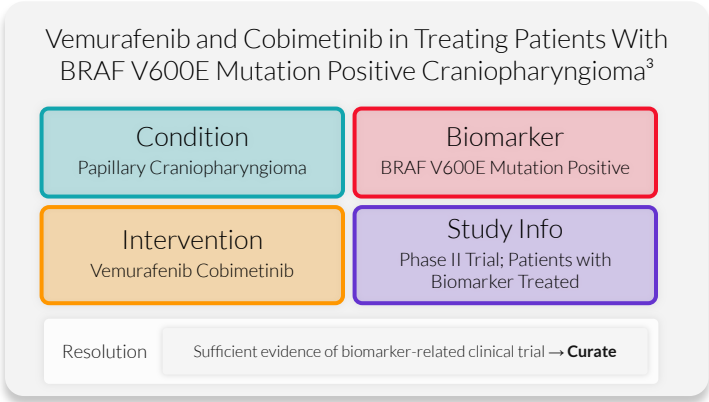
"It is more costly to get biomarker information from public sources than you think. The volume of information requires technology assistance to be efficient. AI approaches like Natural Language Processing (NLP) are the best bet for efficient gathering of clinical information."[2]

- BiomarkerBase

An illustration of an NLP algorithm's identification of a biomarker-based trial



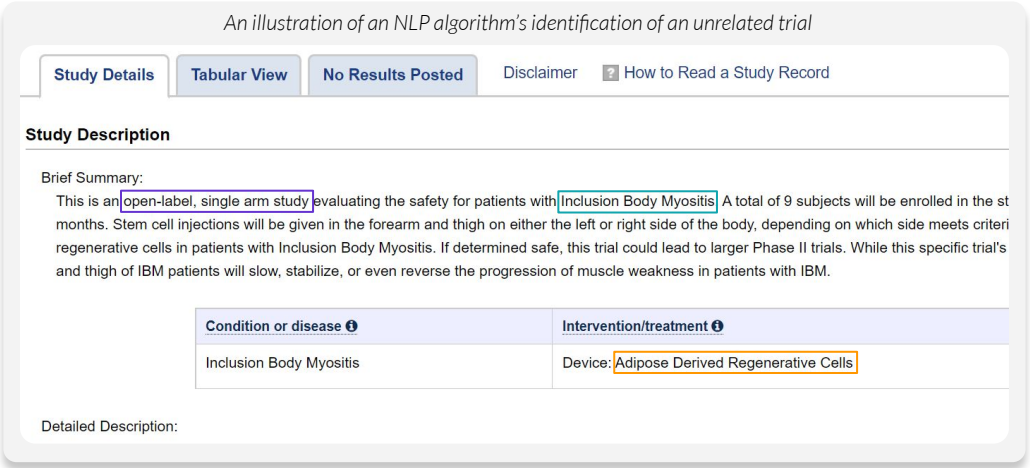
Information Segmentation and Concept Extraction for Resolution Identification



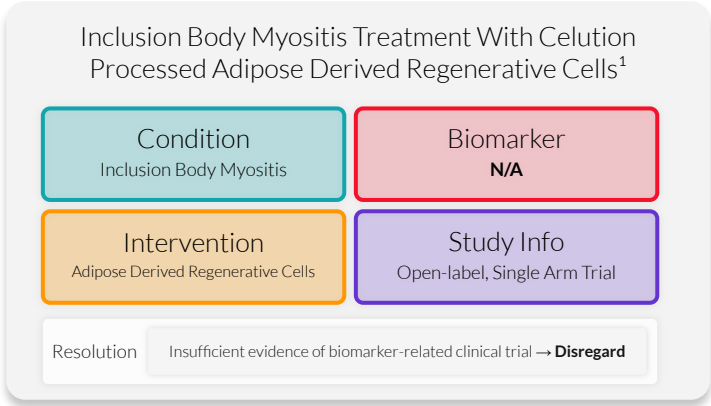
[1] Pradhan, R. et al. (2019) 'Automatic extraction of quantitative data from ClinicalTrials.gov to conduct meta-analyses', Journal of Clinical Epidemiology, 105.
[2] Amplion (2016) 'Benefits of an NLP Approach to Biomarker Identification in Clinical Trials'.
[3] Alliance for Clinical Trials in Oncology (2017) 'Vemurafenib and Cobimetinib in Treating Patients With BRAF V600E Mutation Positive Craniopharyngioma'.

Optimized Textual Analysis Disregards Unrelated Trials

An illustration of an NLP algorithm's identification of an unrelated trial



Information Segmentation and Concept Extraction for Resolution Identification



WR: A Third Dimension to a MECE-based Extraction Approach

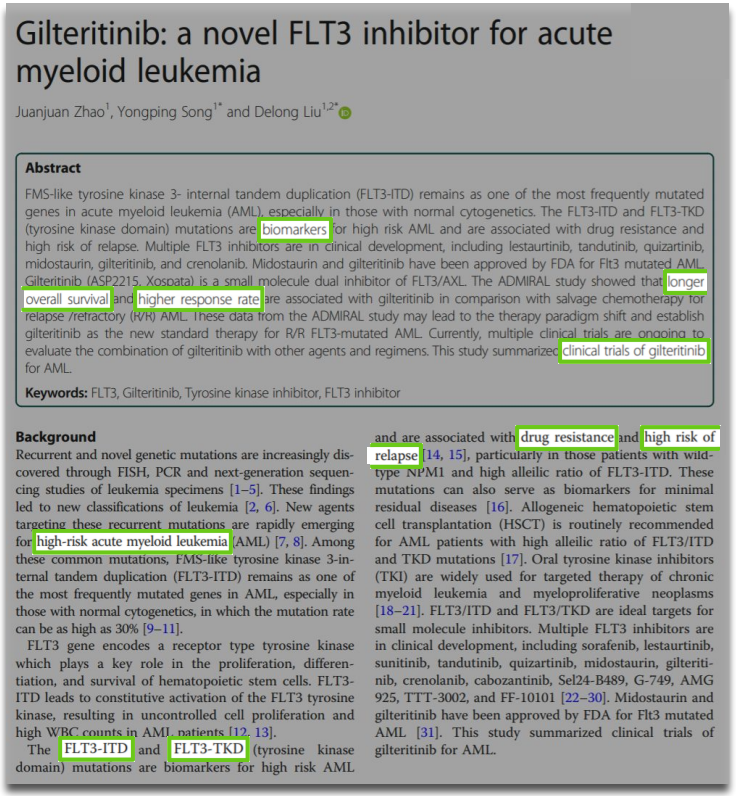
Insofar as it is essential for NLP algorithms to identify relevant trials and literature and avoid duplication, it is equally critical that irrelevant papers and sources are disregarded. By doing so, the amalgam of curated data will not just be Mutually Exclusive and Collectively Exhaustive (MECE), but also Wholly Relevant (WR), ensuring scientists ultimately encounter no unrelated information. To this end, NLP methods must distinguish between trials which might merely reference a biomarker, and those fundamentally founded on a biomarker-based approach.

"The initial step…[in developing] an NLP algorithm to automate the manual process…was to exclude irrelevant information by segmenting the information into different sections; [this is achieved through] concept extraction - a knowledge-driven annotation and indexing process to identify phrases referring to concepts of interests [or disinterest] in an unstructured text.

- Fu, S. et al.

[1] University of Kansas Medical Center (2021) 'Inclusion Body Myositis Treatment With Celution Processed Adipose Derived Regenerative Cells'.
[2] Fu, S. et al. (2020) 'Natural Language Processing for the Evaluation of Methodological Standards and Best Practices of EHR-based Clinical Research',

Repurposing ML to Decode Biomarker Specs and Success



Gilteritinib: a novel FLT3 inhibitor for acute myeloid leukemia

Juanjuan Zhao[1], Yongping Song[1*] and Delong Liu[1,2*]

Abstract

FMS-like tyrosine kinase 3- internal tandem duplication (FLT3-ITD) remains as one of the most frequently mutated genes in acute myeloid leukemia (AML), especially in those with normal cytogenetics. The FLT3-ITD and FLT3-TKD (tyrosine kinase domain) mutations are biomarkers for high risk AML and are associated with drug resistance and high risk of relapse. Multiple FLT3 inhibitors are in clinical development, including lestaurtinib, tandutinib, quizartinib, midostaurin, gilteritinib, and crenolanib. Midostaurin and gilteritinib have been approved by FDA for Flt3 mutated AML. Gilteritinib (ASP2215, Xospata) is a small molecule dual inhibitor of FLT3/AXL. The ADMIRAL study showed that longer overall survival and higher response rate are associated with gilteritinib in comparison with salvage chemotherapy for relapse /refractory (R/R) AML. These data from the ADMIRAL study may lead to the therapy paradigm shift and establish gilteritinib as the new standard therapy for R/R FLT3-mutated AML. Currently, multiple clinical trials are ongoing to evaluate the combination of gilteritinib with other agents and regimens. This study summarized clinical trials of gilteritinib for AML.

Keywords: FLT3, Gilteritinib, Tyrosine kinase inhibitor, FLT3 inhibitor

Background

Recurrent and novel genetic mutations are increasingly discovered through FISH, PCR and next-generation sequencing studies of leukemia specimens [1–5]. These findings led to new classifications of leukemia [2, 6]. New agents targeting these recurrent mutations are rapidly emerging for high-risk acute myeloid leukemia (AML) [7, 8]. Among these common mutations, FMS-like tyrosine kinase 3-internal tandem duplication (FLT3-ITD) remains as one of the most frequently mutated genes in AML, especially in those with normal cytogenetics, in which the mutation rate can be as high as 30% [9–11].

FLT3 gene encodes a receptor type tyrosine kinase which plays a key role in the proliferation, differentiation, and survival of hematopoietic stem cells. FLT3-ITD leads to constitutive activation of the FLT3 tyrosine kinase, resulting in uncontrolled cell proliferation and high WBC counts in AML patients [12, 13].

The FLT3-ITD and FLT3-TKD (tyrosine kinase domain) mutations are biomarkers for high risk AML and are associated with drug resistance and high risk of relapse [14, 15], particularly in those patients with wild-type NPM1 and high alleilic ratio of FLT3-ITD. These mutations can also serve as biomarkers for minimal residual diseases [16]. Allogeneic hematopoietic stem cell transplantation (HSCT) is routinely recommended for AML patients with high alleilic ratio of FLT3/ITD and TKD mutations [17]. Oral tyrosine kinase inhibitors (TKI) are widely used for targeted therapy of chronic myeloid leukemia and myeloproliferative neoplasms [18–21]. FLT3/ITD and FLT3/TKD are ideal targets for small molecule inhibitors. Multiple FLT3 inhibitors are in clinical development, including sorafenib, lestaurtinib, sunitinib, tandutinib, quizartinib, midostaurin, gilteritinib, crenolanib, cabozantinib, Sel24-B489, G-749, AMG 925, TTT-3002, and FF-10101 [22–30]. Midostaurin and gilteritinib have been approved by FDA for Flt3 mutated AML [31]. This study summarized clinical trials of gilteritinib for AML.

Reading Experiments like a PhD Biologist with ML

Having curated biomarker-relevant papers, machine learning can be applied to decode and organize the data therein. In particular, an algorithm is trained to identify specific data points, specs, and criteria within the literature, ranging from biomarker-relevant specs (assay, therapeutic area, matrix) to surrounding details regarding the eligibility criteria and the study itself:

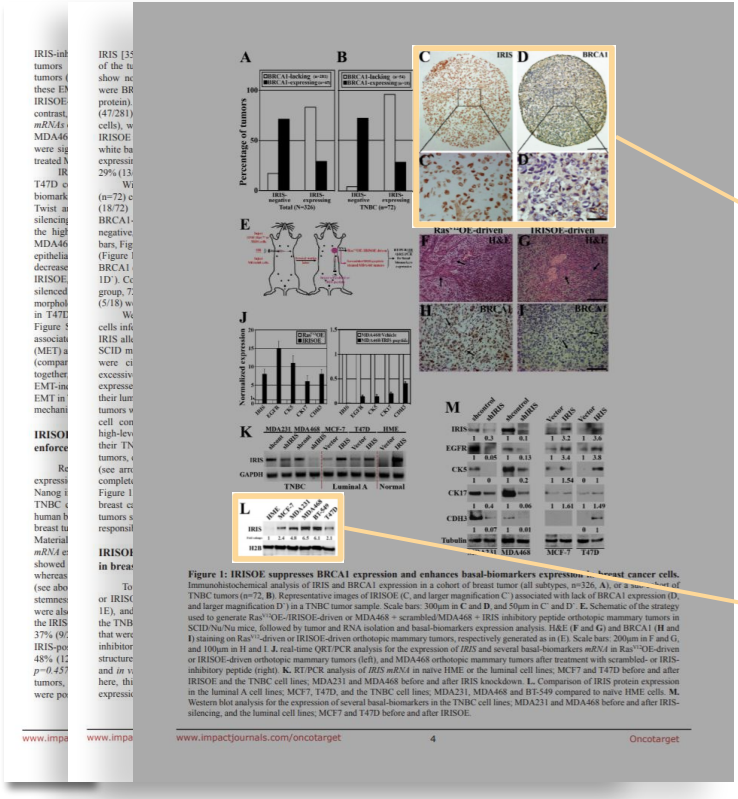
- Biomarkers
- Higher Response Rate
- High-risk Acute Myeloid Leukemia
- Longer Overall Survival
- Clinical Trials of Gilteritinib
- High Risk of Relapse
- FLT3-ITD
- FLT3-TKD
- Drug Resistance
- Small Molecule Inhibitors

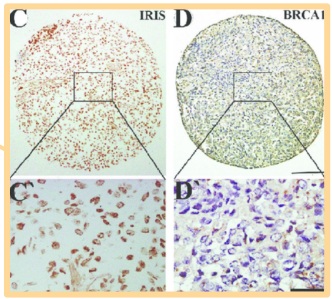
Connect Biomarkers to Use Cases with Bioinformatics

To overcome challenges with entity homonymy and synonymy, BenchSci's existing advanced bioinformatics and ontologies can be repurposed to connect biomarkers to specific use cases and biomedically relevant concepts, offering an additional layer of insight to the user at minimum additional costs.

[1] Zhao, J., Song, Y. & Liu, D. (2019) 'Gilteritinib: a novel FLT3 inhibitor for acute myeloid leukemia', Biomarker Research, 7(1).

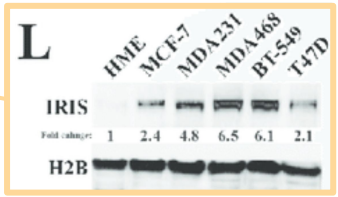
Unlocking the Value of Figures with Image Recognition





Applying Existing Tech for Biomarkers

To further expedite the drug development process for scientists, BenchSci's existing proprietary image recognition technology can be repurposed to extract images in biomarker-relevant literature. Since this setting not so dissimilar to reagent and antibody figures – fundamentally, the 2 scenarios are equivalent – BenchSci's optimized models can be rapidly/effectively reused.



Achieving ML-based Figure Prioritization

By linking captions with their corresponding figures, BenchSci's technology can identify the most valuable image to display. This ML-guided approach can be continually leveraged to elucidate figure presentation priorities.

[1] Sinha, A. et al. (2016) 'BRCA1-IRIS overexpression promotes and maintains the tumor initiating phenotype: Implications for triple negative breast cancer early lesions', Oncotarget, 8(6).

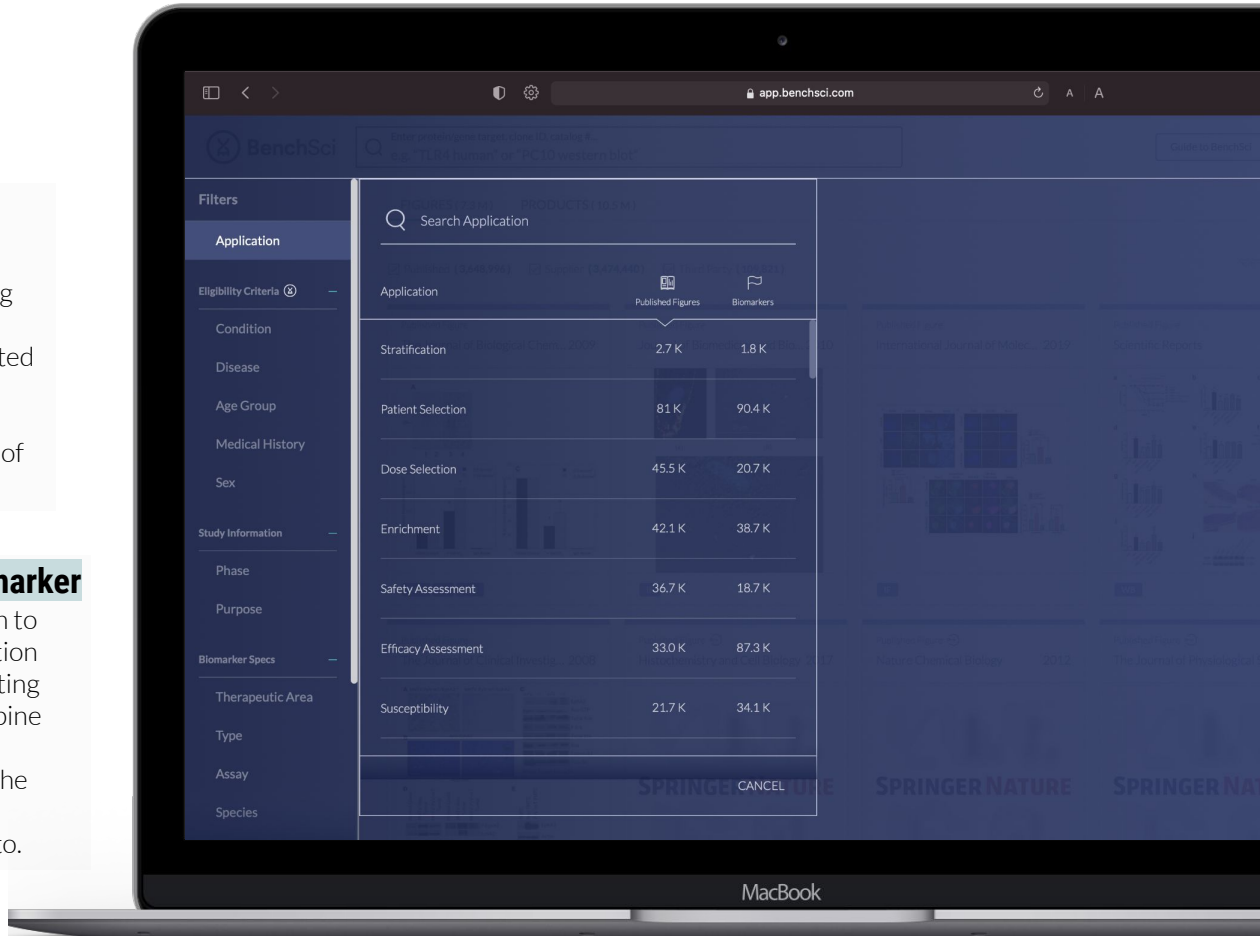
A Platform Catering to Diverse Workflows

A Comprehensive Portfolio of Filters

Foundational to any biomarker (or indeed reagent) selection interface is an intuitive, easy-to-use filtering mechanism. By devising a novel list of experimental variables (which correspond to the categories collected by the ML for Decoding), the platform's searching capabilities can be optimized for biomarker-based inquiries, while preserving the signature personality of BenchSci's existing marquee offerings.

Filter by Application, Eligibility, Study, Biomarker

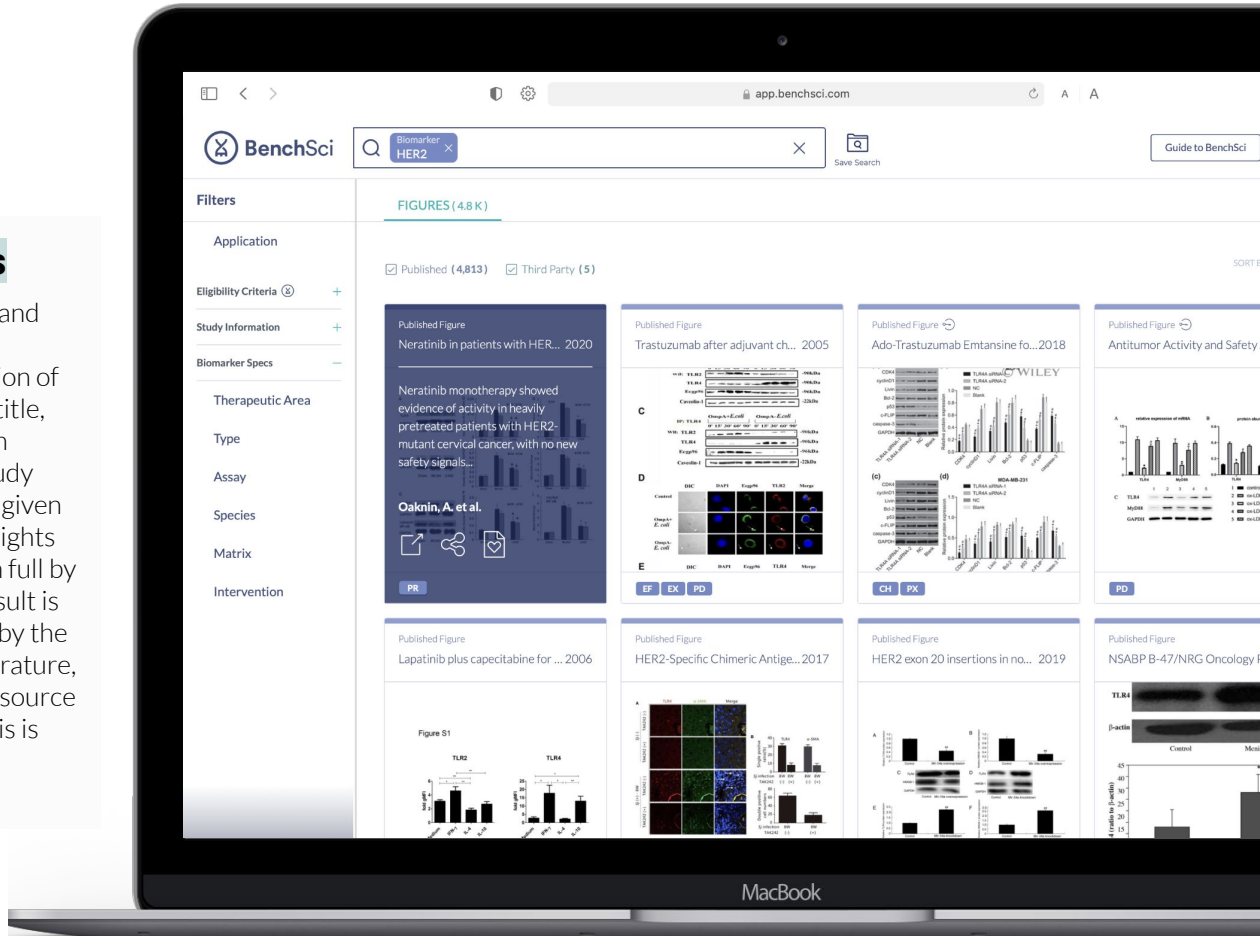
The scientist has the opportunity to limit their search to a specific biomarker use case, from patient stratification to risk/susceptibility. As is essential to Benchci's existing platform, the user should have the capability to combine filters/labels across categories, which include information relevant to the trial's eligibility criteria, the broader details surrounding the study, or indeed the biomarker itself, ensuring all workflows are catered to.



Searching for HER2 Biomarker Insights

Select the Right Biomarkers In Minutes

Having handpicked an assortment of specific labels and filters, the interface presents all matching trials and papers from it's dynamic database, offering a selection of information for each result. First, the user sees the title, and publication date at the top of each result, shown above a miniaturized teaser of the most relevant study figure (see slide 16). If the scientist hovers over any given result, they are shown a teaser of the embedded insights and conclusions from the trial, which they can see in full by clicking on the result. At the very bottom of each result is the 2-letter biomarker type (see slide 12), followed by the paper's authors and options to open the original literature, share the figure, and save the biomarker for each resource related to the biomarker in question (in this case, this is the human epidermal growth factor receptor HER2 biomarker for breast cancer and stomach cancer).



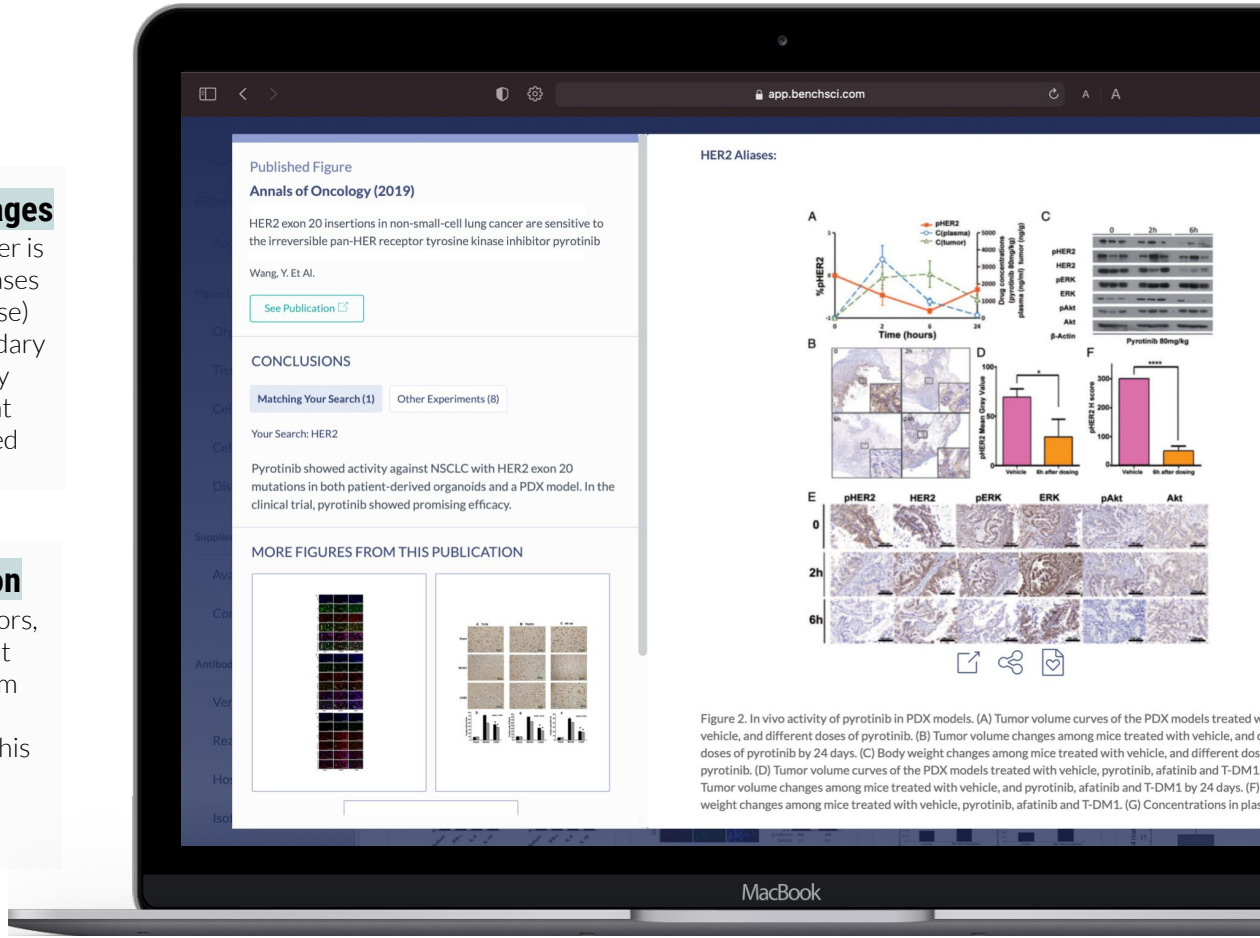
Assimilating Insights to Deliver Value

A Captioned, Priority-based Bundle of Images

Upon clicking upon a particular search result, the user is presented with the primary figures pertaining to aliases of the chosen biomarker (HER2, in this particular case) along with the corresponding caption, before secondary figures from the publication – each too appropriately titled and captioned – are also embedded; this format closely mirrors the configuration of the image-related aspects of BenchSci's reagent selection platform.

Integrated Insights Expedite HER2 Adoption

Supplementing information regarding the title, authors, and academic journal is a bite-sized snippet of insight on the trial's outcomes or findings – often pulled from the 'Conclusions' or 'Results' components of the original publication. For this particular clinical trial, this section summarizes pyrotinib's promising efficacy against NSCLC with the HER2 exon 20 mutation, signalling the drug's potential in seconds.



Finding the Biomarkers Scientists Need, Faster and Cheaper

An Intuitive Solution Addressing a Bona Fide Pain Point

Built upon powerful technology, the solution eliminates inefficiencies and errors in the entire biomarker selection process, procuring a greater likelihood of approval and enabling the mitigation of costly experimental failure.

Accelerate Projects

With the AI-assisted platform, projects are accelerated dramatically by selecting biomarkers within just 3 minutes rather than the status quo average of around 3 weeks, enabling scientists to advance assets with fewer, shorter more successful clinical experiments.

Eliminate Costs

By selecting the right biomarkers quicker, scientists can reduce the hard cost of consumables, total pay given to patients (with a lower sample size), and a host of other supplies, to the tune of **$4 million** in initial cost savings value and a further **$52k** in monthly cost savings for every **50k biomarkers**, as the chance that fail to work downstream reduces..[1]

Empower Scientists

By restoring weeks of research time and minimizing costly resource expenditure, the platform empowers organizational purpose and enables scientists to predict drug efficacy more quickly than conventional clinical endpoints, identify risk factors and individuals at risk, and a host of other benefits.[2]

Impact Business

By leveraging a turnkey biomarker solution powered by advanced proprietary machine learning models, organizations can invest time and resources more meaningfully throughout the drug development process, further compounding the pool of expected benefits associated with rapid biomarker selection.

[1] Amplion (2016) 'Benefits of an NLP Approach to Biomarker Identification in Clinical Trials'.
[2] Selleck, M., Senthil, M. & Wall, N. (2017) 'Making Meaningful Clinical Use of Biomarkers', Biomark Insights, 12:1–7

Structurally Simpatico to Ensure Feasibility

Mirroring the Existing Approach Guarantees an Effortless Transition

By constructing a solution structurally equivalent to BenchSci's current marquee offerings, no insurmountable feasibility challenges – or indeed particular novel scenarios – will be encountered. The bulk of the work will entail repurposing the same curation and decoding strategies to assimilate and organize biomarker data, and making minor, almost negligibly significant adjustments to the interface to construct a biomarker-specific engine. In a nutshell, the only significant difference between BenchSci's current antibody/reagent interface and our proposed biomarker selection platform is that biomarker-relevant information is curated, decoded and presented rather than antibody/reagent data.

AI-Assisted ~~Antibody~~ Biomarker Selection

Using AI to decode open- and closed-access data on **biomarkers** and present published figures with actionable insights, allowing researchers to reduce time, money, and uncertainty in clinical trials.

Curate	Decode	Present
Curate the world's largest collection of **biomarker** data from life science experiments and reagent catalogs with NLP.	Decode and organize the **biomarker** information with proprietary machine learning algorithms for text and images.	Provide rapid **biomarker** insights with comprehensive filtering, assimilated figures, and integrated conclusions.

Barriers & Assumptions

The Proprietary Information

In an attempt to realize a competitive edge, some unique/proprietary biomarker-related data is restricted from external sources, posing a potential challenge to the accumulation of an exhaustive database; however, BenchSci's partnerships and their own straddled proprietary info may serve as a valuable foothold on which to negotiate with potentially reluctant sources.

A Competitive Landscape

With a handful of existing competitors in the biomarker selection setting, existing customers and clients – already comfortable with the products they currently use – may be unwilling to switch to BenchSci's offering. However, should BenchSci deliver a more intuitive and comprehensive solution at competitive pricing, this bucket of potential customers will recognize that BenchSci's platform represents a significantly more complete, valuable service.

A Layperson's Reluctance

With just 7.1% of trials using biomarkers at all stages, prospective clients unaccustomed to a biomarker-driven approach to clinical trials may be reluctant to make the transition. However, by gaining traction and validation who currently engage and struggle with biomarker selection, this group of hard liners will recognize the sheer value of integrating biomarkers into their drug development pipeline.

A 15-month Go-to-market Strategy for Nov '22 Launch

A Waterfall Roadmap for the First Iteration and Beyond

By repurposing BenchSci's core machine learning and interface infrastructure, the offering can be on the market within 15 months.

Aug '21
Conduct further market analysis to identify obstacles, crystalize the unique value proposition, and validate viability.

Sep '21
Hire ML engineers, R&D scientists, & biomarker specialists. Start gauging scope of curation required for exhaustivity.

Oct '21
Assemble a complete list of these sources, begin curation, and retrain NLP models to identify biomarker-relevant sources.

Dec '21
Fine-tune NLP infrastructure to optimize curation, contact partners to collect all possible data on biomarkers.

Feb '22
Post-completion of curation, retrain ML models for text and images to decode and organize the data in accumulated papers.

Apr '22
Validate experimental labels/criteria with experts, assign them using repurposed ML algorithms to all curated literature.

Jun '22
Begin developing the interface, building off the existing platform and renaming filters, labels, & searching for a biomarker setting.

Aug '22
Integrate all decoded biomarker data with the updated platform to deliver an intuitive, turnkey application for efficient searching.

Sep '22
Beta-test the offering with a handful of trusted clients, implementing feedback to optimize the user experience.

Oct '22
Finalize pricing and integration of new platform with existing offerings. Recruit early adoptees to build traction.

Nov '22
Launch the product commercially, perhaps coupled with Series C funding, taking the next big leap to expedite R&D.

2023
Iterate the offering as novel partnerships are formed, innovative recommendations are implemented, and new data is gathered.

A 3-Stage Scaling Strategy to Maximize ROI

Stage II
Having been established as the leading biomarker offering, BenchSci would be well positioned to expand to the **34.7%** of CROs which use biomarkers in some capacity, playing up the platform's reduction of time, costs, & uncertainty.

Stage III
With a hefty minority of CROs as clients, BenchSci's biomarker selection leaps into the radar of the **65.3%** of organizations around the world which had never previously used biomarkers. Supplementing the platform's critical acclaim and with effective messaging on the benefits procured by a biomarker-driven approach, BenchSci is now well-positioned to expand and dominate a flourishing market.

Stage I
By first targeting the **7.1%** of CROs which leverage biomarkers at all clinical stages, BenchSci can obtain a strong foothold in the market. In doing so, BenchSci compounds value for existing clients and secures new partnerships, bringing in sufficient revenue to iterate on the offering & further outreach.

"Pfizer is interested in establishing alliances to develop and/or access biomarker studies."[1]
- Paul E. Young, Pfizer

[1] Pfizer (2021) 'Pfizer's Interest in Developing Alliances'.

Bringing New Medicine to Patients 50% Faster by 2025

A Market Blossoming towards $92.1 Billion by 2025[1]

Expected to rise at a CAGR of **16.8%** through 2025, the global biomarker technologies market is growing towards a **$92.1 billion** cap by 2025.[1] These favorably-blowing winds represent a colossal opportunity for BenchSci to put the world's biomarker knowledge at scientists' fingertips in the fraction of the time and work towards the firm's larger ambition of bringing new drugs to the market 50% faster by 2025. Indeed, by returning time to scientists & increases research velocity, BenchSci's biomarker selection platform in turn grows the global biomarker market , enlarging the share of rewards to be reaped.

Completing a Job Half-Done by Competitors' Offerings

No competitors on the market currently offer a complete biomarker package consisting of a comprehensive collection, intuitive interface, and integrated insights and figures. Should BenchSci deliver an offering at the intersection of all 3 attributes, the company stands to rapidly surpass competitors' offerings, which themselves represent a job half-done and therefore are incapable of delivering as much value as all 3 core attributes in unison. The synthesis of this large gap in the market and this very same market's projected flourishing in the next 4 years bodes positively for this considerable endeavor.

Realizing an Exciting Future for BenchSci

By empowering scientists to select the right biomarkers with velocity, BenchSci's biomarker selection program is well-positioned to exponentially increase the speed and quality of their life-saving research.

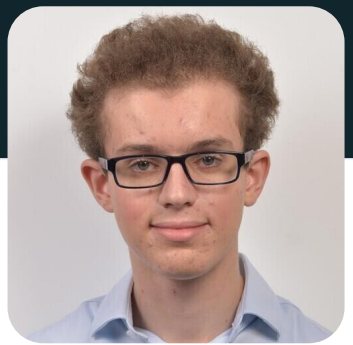
In transitioning BenchSci's proprietary ML curation and decoding algorithms from a preclinical research context to a clinical trial setting, the significant delays and costs straddled by the later stages of drug development can be mitigated, paving the way for scientists to bring new, impactful medicines to the market.

[1] BCC Research (2021) 'Explosion of Publications, Clinical Trials Driving Growth of Biomarkers Technology'.

Thank You, BenchSci

Anna Heck

Alex Koubaa



Mir Ali Zain

Liesl Anggijono

We are so excited to have this opportunity to help increase the success of clinical trials and help people get the medication they need faster! We have had an such an eye-opening and amazing experience working with the awesome BenchSci team. Special thanks to Jelena, our mentor for guiding us through this journey. Thank you for the opportunity to help you tackle a real life problem that will impact billions in the future.