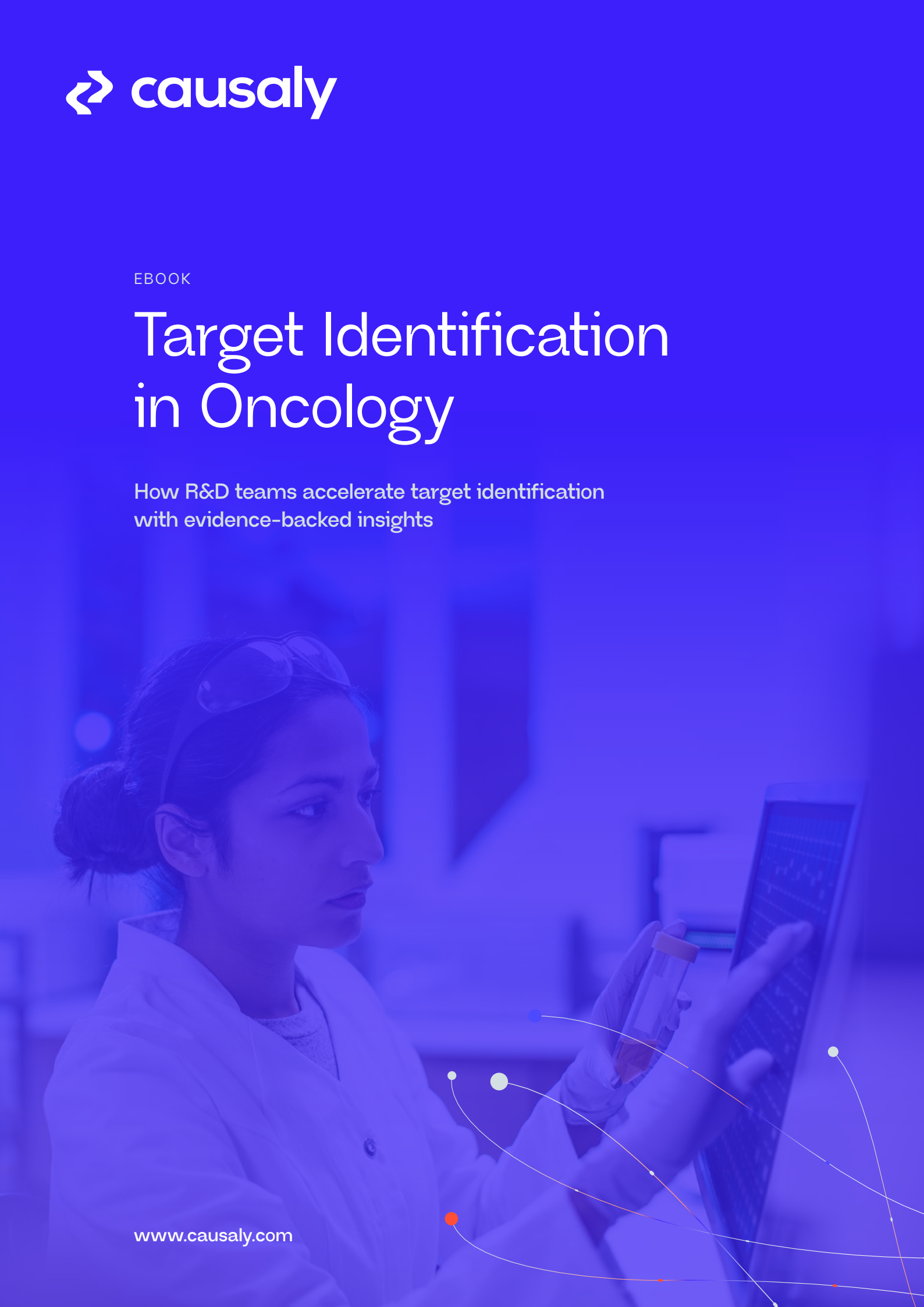
causaly

EBOOK

Target Identification in Oncology

How R&D teams accelerate target identification with evidence-backed insights



www.causaly.com

Foreword

Oncology dominates pharmaceutical R&D activity, accounting for 41% of all clinical trials in 2024,[1] driven by high unmet needs, a large and expanding pipeline across solid and hematological cancers, and continued innovation in novel modalities.

No other therapeutic area requires anything close to this level of scientific and financial commitment. The scale of that investment makes the quality of target identification decisions more consequential.

The volume of scientific evidence has outgrown the tools used to navigate it. The gap between what is published and what any single team can systematically review continues to widen. Most target decisions today are made on a fraction of the literature, without a reproducible record of how that fraction was chosen.

This eBook shows how current research and development processes are extending timelines and introducing risk, and how leading teams are using agentic research to accelerate discovery, increase accuracy, and build a defensible evidence base that prevents failure

The volume of scientific evidence has outgrown the tools used to navigate it. The gap between what is published and what any single team can systematically review continues to widen.

CHAPTER 1

The Increasing Complexity of Identifying Targets in Oncology

Oncology is the largest area of pharmaceutical R&D investment, with global sales projected to exceed $400+ billion by 2029 and potentially much higher by 2035, driven by innovation in targeted therapies, immunotherapies, and advanced modalities.[2]

Yet, attrition rates remain exceptionally high, with only around 5% of drugs entering Phase I and reaching approval. With lack of efficacy accounting for roughly half of clinical development failures, particularly in late-stage trials.[3]

Cancer biology adds layers of complexity to research. Tumor heterogeneity creates subclonal populations with distinct resistance mechanisms, meaning a target may be relevant in one patient or tumor region but not in another, which complicates both biomarker validation and drug efficacy.[4] Acquired resistance arising through secondary mutations, pathway bypass, phenotypic transformation, and microenvironmental adaptation is also a leading cause of treatment failure, even when initial target selection is correct.[5]

$400B+

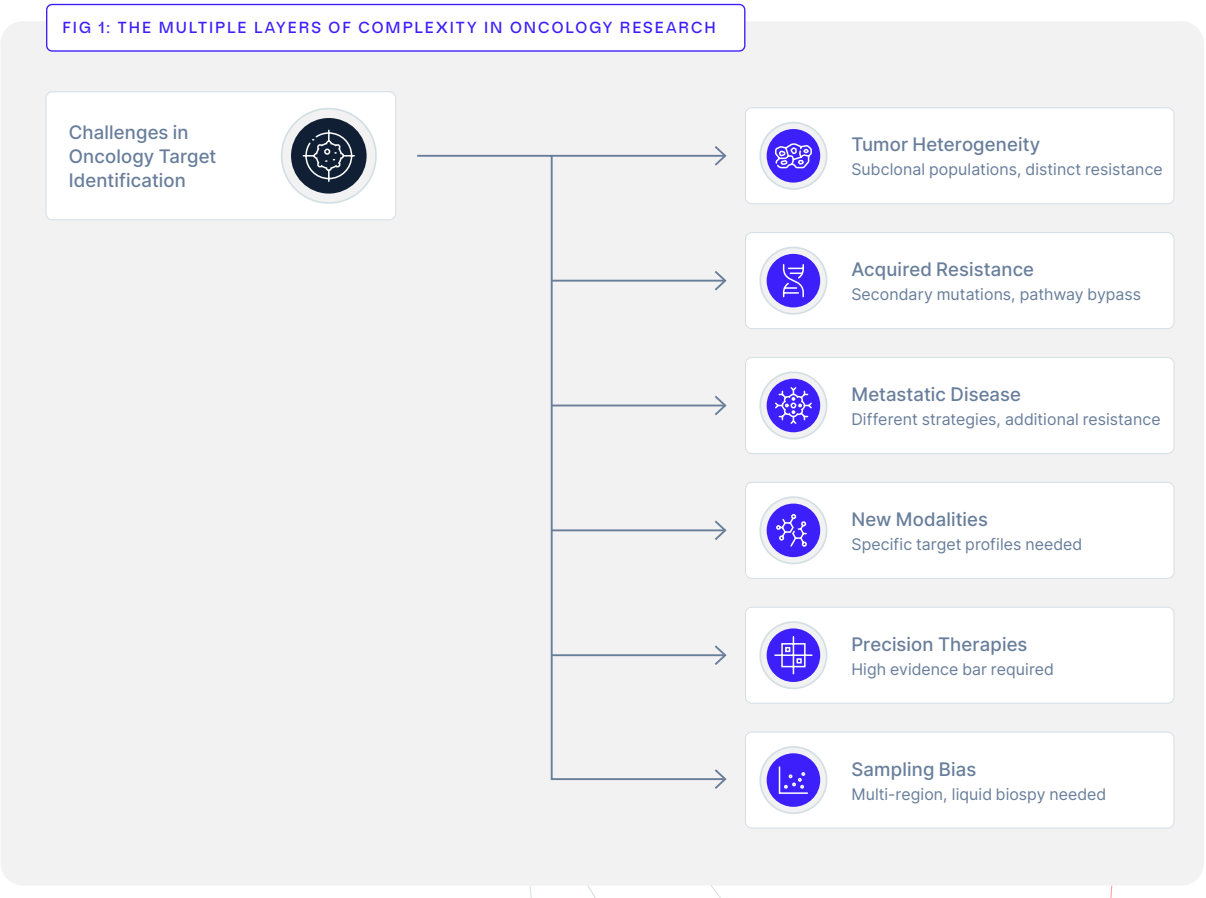
in global oncology sales expected by 2029, intensifying the need for confident target identification

Metastatic disease introduces further challenges, requiring different strategies from primary tumors and presenting additional resistance mechanisms that must be considered separately.[6] At the same time, each new modality, such as CAR-T, RNA therapeutics, and targeted immunotherapy demands its own specific target profile, with success depending on identifying targets that are both biologically relevant and physically accessible.[7]

The field is moving decisively toward molecularly targeted therapies, raising the bar for evidence at every stage. Determining which targets are causally relevant (across cancers, patient populations, and modalities) demands evidence that is not only comprehensive and traceable, but robust enough to support patient stratification and regulatory approval.

Sampling bias further complicates target identification & validation, making multi-region and liquid biopsy approaches increasingly important alongside conventional tissue analysis

FIG 1: THE MULTIPLE LAYERS OF COMPLEXITY IN ONCOLOGY RESEARCH



CHAPTER 2

Traditional Target Identification Process

There are various approaches to target identification, depending on the organization, therapeutic area, and available data.

One common method begins with disease understanding: mapping dysregulated pathways, and known genetic drivers, signaling relationships before any candidate is evaluated, grounding target discovery in disease biology and mechanistic context.[8] From there, data mining then draws on genomics, transcriptomics, proteomics, and public databases. Integrating multi-omics (identifying targets that are genetically altered, transcriptionally upregulated, and functionally active in-patient tumor samples) provides convergent evidence across independent multi-data types that is superior to single-measure methods.[9]

Target prioritization applies structured criteria such as strength and direction of the target-disease evidence, genetic validation from GWAS and CRISPR knockout models, druggability, safety signals, expression in human tumor tissue, and clinical actionability. These criteria are widely used in both academic and industry settings to ensure that only the most promising and tractable targets advance. In parallel, network biology approaches including protein-protein interaction mapping, pathway analysis, and synthetic lethality screens help surface candidates at key biological junctions that might otherwise be missed through direct search methods.[10]

The highest-priority candidates are then validated experimentally in cancer cell lines and patient-derived organoids which are then taken into animal models for in vivo confirmation. Orthogonal approaches (genetic knockdown, chemical probes, rescue experiments) are run independently to reduce false positives, with rigorous controls and blinding throughout to ensure results hold up when reproduced elsewhere.[11]

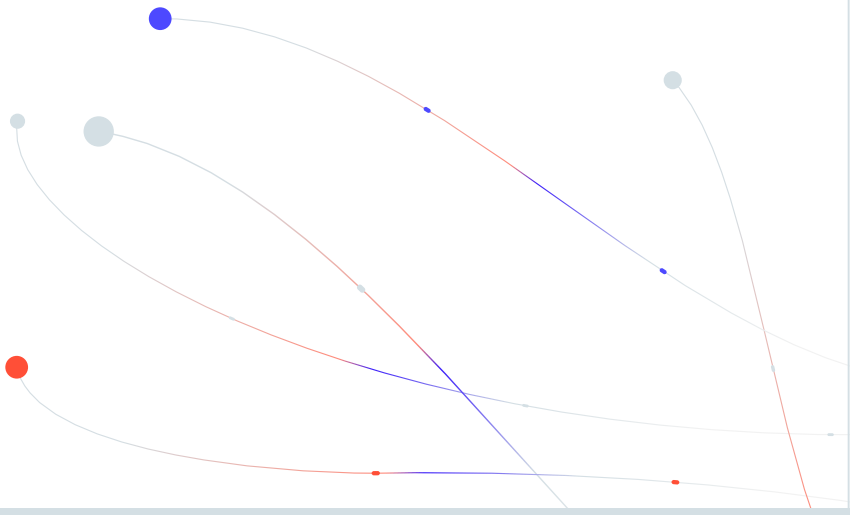
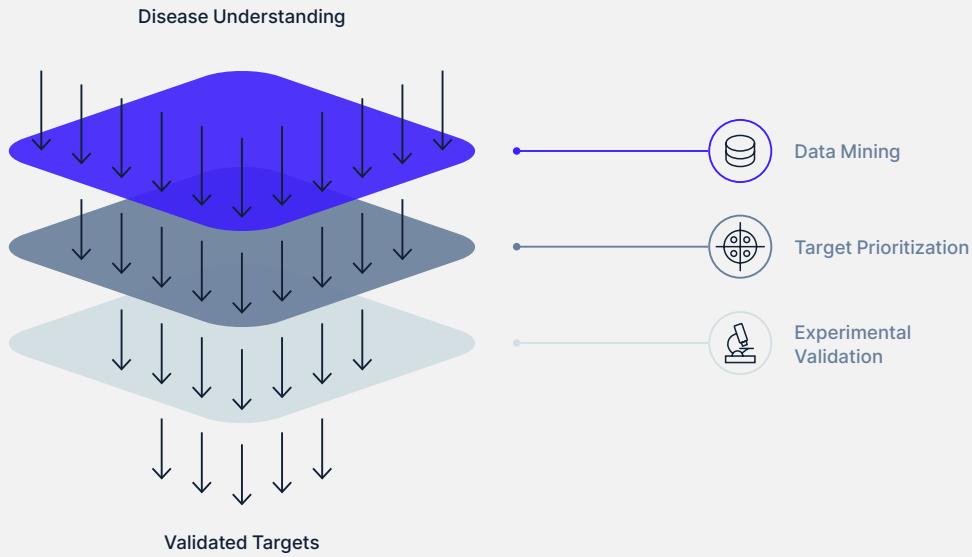


FIG 2: TRADITIONAL TARGET IDENTIFICATION PROCESS



Where there are evidence gaps

The challenge at every stage is systematic coverage. The majority of compounds designed against cancer drug targets do not progress to become approved drugs, mainly due to a lack of efficacy and/or unmanageable toxicity. Robust target evaluation is therefore required before progressing through the drug discovery process to reduce the high attrition rate. There is a wealth of publicly available databases that can be mined to generate data as part of a target evaluation. It can, however, be challenging to learn what databases are available, how and when they should be used, and the associated limitations.[12]

In areas such as non-small cell lung cancer (NSCLC), the target landscape is both extensive and rapidly evolving. Alongside well-established driver mutations like EGFR, ALK, KRAS, and ROS1, ongoing research continues to identify new molecular targets and therapeutic strategies. However, biomedical searching is inherently vulnerable to selection bias, and incomplete retrieval can distort which targets appear most promising. This underscores the need for structured, comprehensive search approaches to ensure that critical evidence is not overlooked

CHAPTER 3

Evolution of Target Identification: Evidence Quality as the Foundation of Target Identification

Before using any discovery process, it helps to be clear about what the evidence needs to demonstrate. A well-supported target in oncology needs to be backed by multiple independent lines:

Genetic evidence

GWAS or genetic studies directly linking the target gene to cancer progression or maintenance.

Mechanistic evidence

Experimental data showing how the target contributes to disease pathology, including pathway and signaling relationships.

Clinical evidence

Data from human studies establishing the target's relevance in patient populations.

Safety and selectivity

Assessment of tumor specificity, expression in normal tissues, and toxicity risk when the target is modulated.

Directional evidence

The direction of the relationship matters. This encompasses both;

- (i) the effect direction: whether modulation of the target worsens or improves disease and
- (ii) the causal direction: whether the target drives disease processes or reflects a downstream consequence.

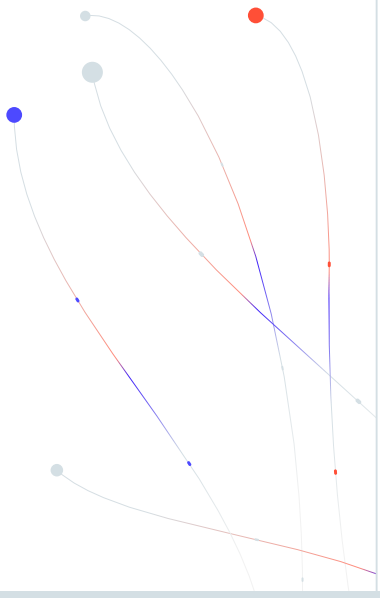
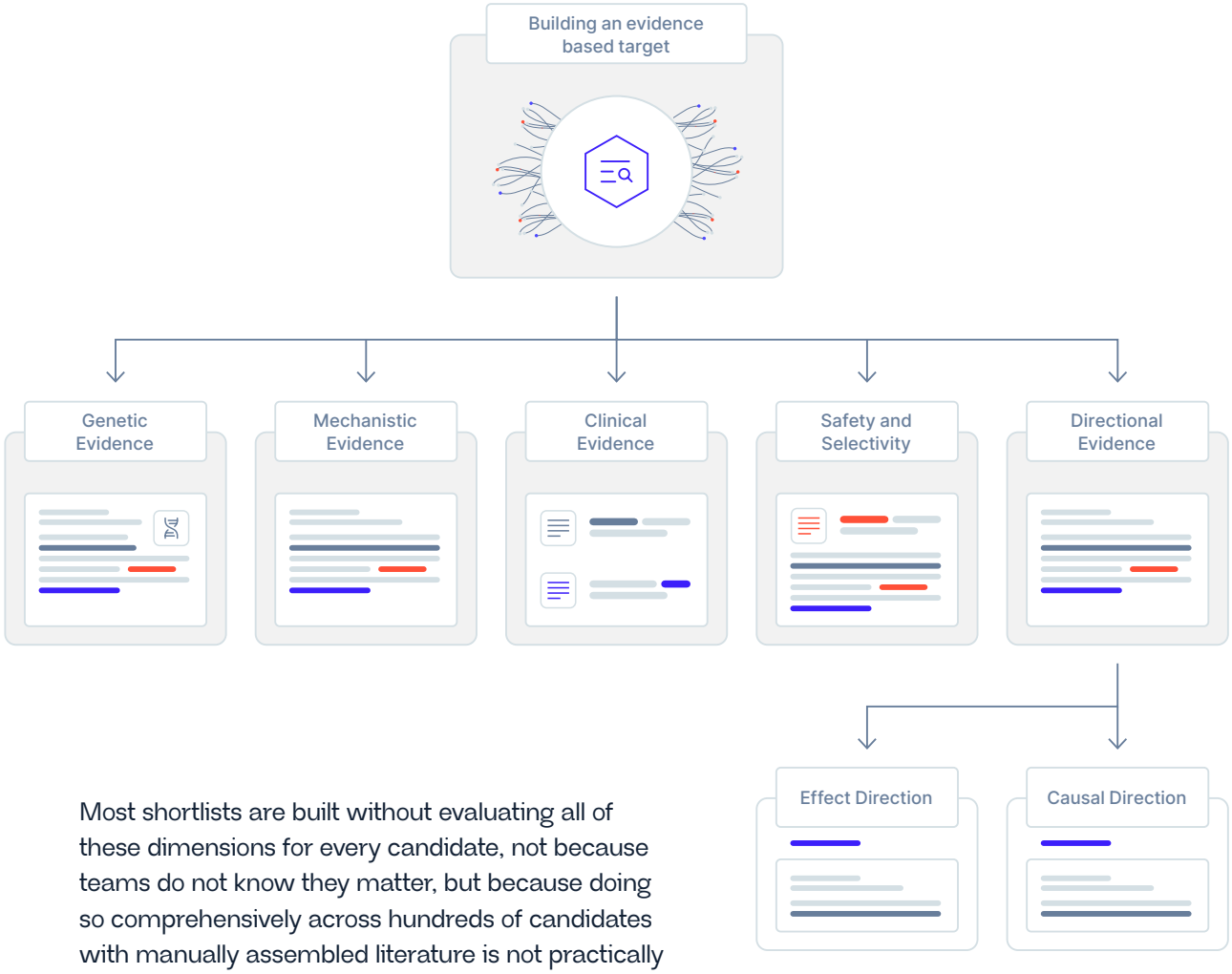


FIG 3: A EVIDENCE-BASED TARGET IDENTIFICATION PROCESS IN ONCOLOGY



Most shortlists are built without evaluating all of these dimensions for every candidate, not because teams do not know they matter, but because doing so comprehensively across hundreds of candidates with manually assembled literature is not practically achievable. Something gets dropped or missed.

Usually, the hardest dimensions to uncover are directionality, safety signals buried in tangential studies, and contradictory findings that never make it onto a typical PubMed results page

CHAPTER 4

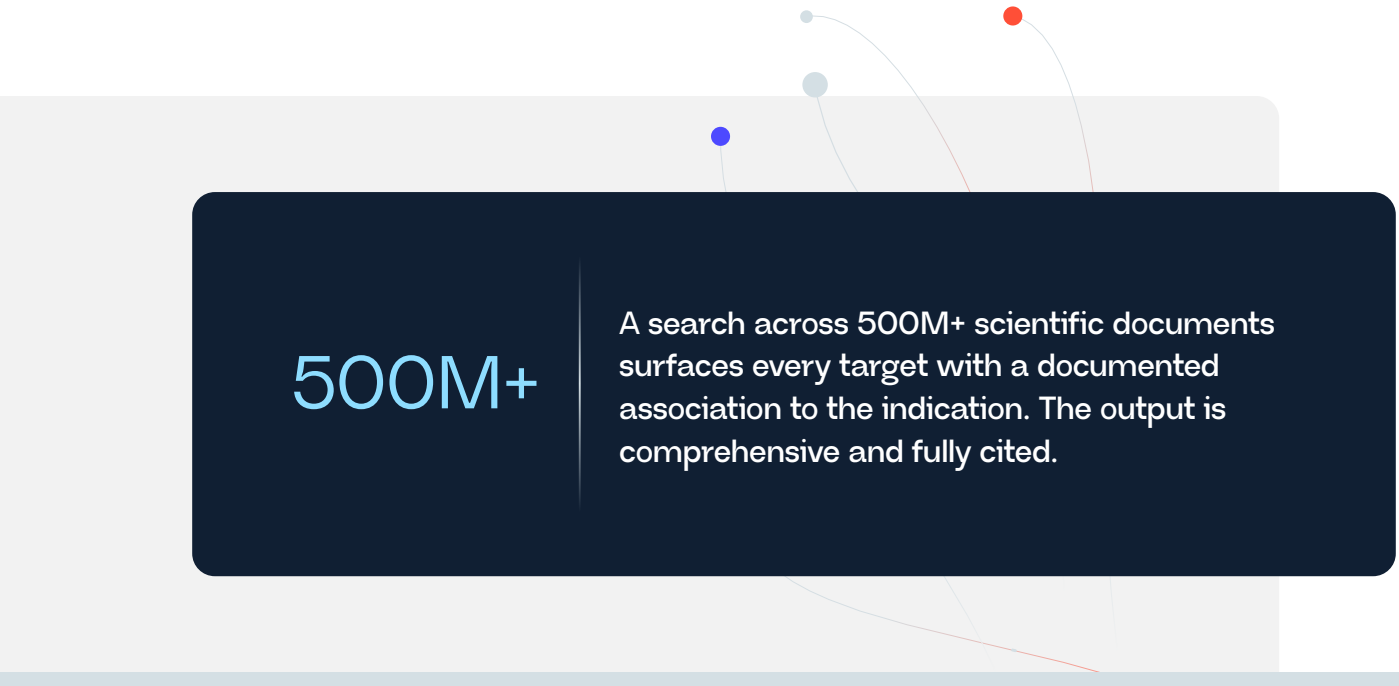
Reducing Errors with Complete Evidence to Improve Target Identification Outcomes

Fewer errors, faster decisions: how Causaly streamlines target identification.

A target identification project in Causaly begins with disease understanding rather than a candidate list. Scientists use Agentic Research to map the molecular and cellular processes associated with the indication (the dysregulated pathways, the known drivers, the signaling relationships), building a structured view of the disease landscape from the knowledge graph with over 70M+ directional relationships before any filtering begins.

From there, a search across 500M+ scientific documents surfaces every target with a documented association to the indication. The output is comprehensive and fully cited. The goal at this stage is coverage, making sure nothing relevant is missed, and no candidate is excluded before the evidence is seen. Because Causaly's extraction is systematic, the list reflects the full scope of what is available in the literature, not what a keyword search would return.

Prioritization applies filters across multiple dimensions simultaneously: target class, strength of genetic evidence, validation in human tissue or animal models, safety signals, novelty. Multi-hop analysis surfaces targets implicated through indirect mechanisms that a direct search would never reach. Thousands of possibilities are narrowed down to tens of high-priority candidates, with every filtering decision traceable.



500M+

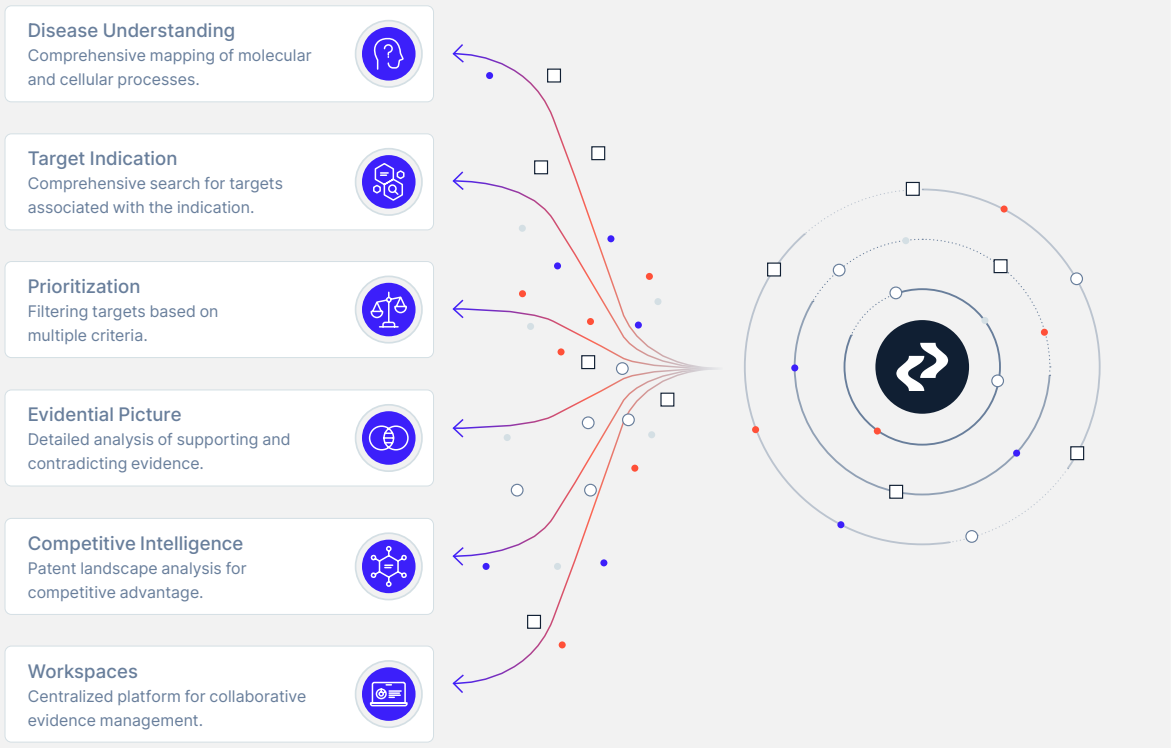
A search across 500M+ scientific documents surfaces every target with a documented association to the indication. The output is comprehensive and fully cited.

For shortlisted candidates, Bio Graph provides the full evidential picture: what supports the association, what contradicts it, how the evidence has evolved, and, for competitive intelligence, Pipeline Graph shows what the patent landscape looks like. Delta Analysis compares and contrasts the evidence base across multiple candidates or indications side by side. The output becomes a documented, evidence-backed rationale with every claim cited, every limitation visible, and one-click access to the original source directly within the platform.

Everything is captured in Workspaces. As programs progress and the field moves, Causaly's alert functionality keeps the evidence database current with new papers, trials, and datasets giving all cross functional teams access to the compounding intelligence.

This process ensures the team's understanding evolves together, rather than drifting apart through separate searches run at different times ●

FIG 4: EVIDENCE-BASED TARGET IDENTIFICATION WITH CAUSALY



CHAPTER 5

How R&D Teams Are Making Better, Evidence-Backed Target Decisions with Causaly

The three cases below come from oncology R&D teams at different stages and sizes. What they share is a common starting point: a target identification process producing results teams could not fully stand behind, the need to change that before committing further resources, and a search for a more transparent, evidence-driven way to prioritize targets, one that Causaly helped provide.

Renal Cell Carcinoma: five targets a manual process would have missed

A global oncology-focused pharmaceutical company needed new targets for solid and liquid tumors, starting with RCC. They had defined tumor antigen criteria, including antigen expression in humans, and needed a shortlist that could withstand internal scrutiny and not just a list of candidates, but a documented case for each one.

Using Causaly, the team evaluated the full landscape of potential tumor antigens, assessed mechanistic safety risks early to de-risk selection, and used Delta Analysis to identify targets shared across multiple indications for lifecycle planning. The output was ten high-potential candidates, five of which were newly identified targets backed by traceable evidence the team could present at review.

> "Causaly is key for us to save time because we spend a lot of resources on personnel-related costs. If we can save time identifying new targets and starting new programs, we save money and extend the life of the company."
>
> DIRECTOR OF RESEARCH AND DEVELOPMENT,
> GLOBAL ONCOLOGY-FOCUSED PHARMACEUTICAL COMPANY

Non-Small Cell Lung Cancer: a year's work in five hours

A pharma company working on NSCLC — one of oncology's most heavily researched disease areas — needed confidence that their shortlist reflected the strongest available evidence, not just the most visible targets or whatever surfaced first. Were they converging on the same crowded targets as everyone else?

The team used Causaly across the full NSCLC literature, filtered to preclinical candidates, and ranked by strength of disease evidence rather than publication frequency. Seven thousand candidate targets were reduced to one thousand in five hours. The equivalent manual process would have taken approximately one year. The shortlist included targets found only on page 200+ of the equivalent PubMed search.

7,000

candidate targets reduced to 1,000 in 5 hours

Including high-quality candidates that would never have appeared in a manual review

Innate Pharma: from fragmented sources to a shared evidence base

Innate Pharma needed to identify and prioritize tumor antigens, evaluate their biological relevance and druggability, and track the competitive landscape for antigens with proven clinical activity. Their existing process produced results that were difficult to replicate consistently especially when their scientists' compared notes, because they were often working from different evidence.

Causaly became part of the team's daily workflow. Using Workspaces, they built a shared, living record of their target evaluation work with consistent visibility on expression patterns across tumor types and stages, comparative target-indication relationships in a single view, and a citable evidence base every team member could build on.

The shared evidence base revealed therapeutic opportunities the team had not originally scoped. Indications beyond the initial focus became viable candidates, not because the science had changed, but because the team had more transparency

> "Causaly is my new friend. I use it every day. It helped us realise we could target many more tumor types than we initially thought. It has become essential for evaluating target expression and biology efficiently."

FLORENT CARRETTE,
SENIOR R&D PROJECT MANAGER, INNATE PHARMA

CHAPTER 6

Conclusion

Target identification in oncology integrates cutting-edge science, rigorous validation, and strategic decision-making. Modern workflows now leverage multi-omics, CRISPR screens, network biology, and AI-assisted prioritization with knowledge graphs and AI platforms central to integrating heterogeneous data types and surfacing actionable insights.

What causaly delivers goes beyond speed. It is the quality, traceability, and auditability of the evidence base behind every target decision that enable the integration of heterogeneous data, supporting evidence-based decision-making, and providing a holistic, auditable view of the evidence trail. Practical utility depends on implementation and integration with expert workflows, and ongoing attention to data quality and standardization remains important.

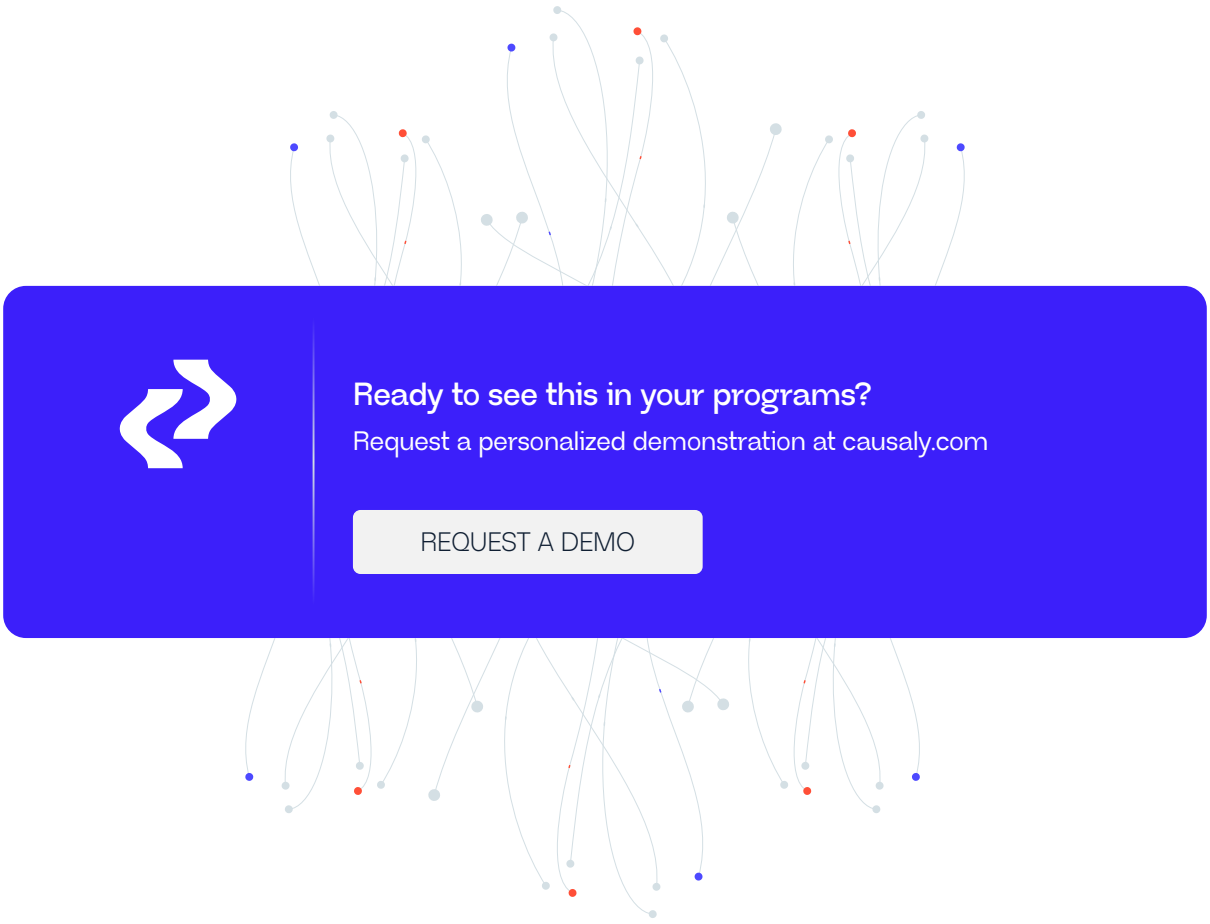
The teams described in this eBook are not working harder than their peers. They are working from a more complete, evidence-based, and auditable process. They surface both supporting and contradictory evidence, document the full evidence trail, and present traceable, cited shortlists to review committees, a major advance over traditional keyword search, which is subject to selection bias and often misses critical or contradictory findings.

Teams report up to 90%+ time savings and 5x–10x productivity improvements, with time reallocated to higher-value scientific work. ProQR, for example, met their 2024 target identification goal by Q3 and achieved 5x productivity compared to PubMed, with 50% of targets pitched informed by insights from Causaly. That is where the investment pays off, not in hours saved reading abstracts, but in the quality of what those hours now produce

In a field where target selection determines whether a program succeeds or fails years down the line, getting the evidence right at the start is the most important investment a discovery team can make.

About Causaly

Causaly is a platform for life sciences R&D that reads scientific evidence across 500M+ documents, including publications, clinical trials, patents, preprints, and GWAS data, and extracts directional, cause-and-effect relationships between biological entities with 90%+ precision. It is used by R&D teams at the top 20 global pharmaceutical companies for target identification, biomarker discovery, indication expansion, and competitive intelligence, amongst other use cases



Sources and References

1	https://www.iqvia.com/insights/the-iqvia-institute/reports-and-publications/reports/global-oncology-trends-2025
2	blog.bccresearch.com; pharmiweb.com; Henderson Raymond H. et al., 2023.
3	pubmed.ncbi.nlm.nih.gov; nature.com; Failure of Investigational Drugs in Late-Stage Clinical Development and Publication of Trial Results
4	https://pmc.ncbi.nlm.nih.gov/articles/PMC9066118/; https://pmc.ncbi.nlm.nih.gov/articles/PMC8422119/
5	https://pubmed.ncbi.nlm.nih.gov/40461793/
6	https://pubmed.ncbi.nlm.nih.gov/35840112/
7	https://pmc.ncbi.nlm.nih.gov/articles/PMC11928443/; https://pmc.ncbi.nlm.nih.gov/articles/PMC10712310/
8	https://pmc.ncbi.nlm.nih.gov/articles/PMC5469697/
9	https://pmc.ncbi.nlm.nih.gov/articles/PMC12809910/
10	https://www.biorxiv.org/content/10.1101/2025.01.21.634201v2
11	https://pmc.ncbi.nlm.nih.gov/articles/PMC11775273/; https://pubmed.ncbi.nlm.nih.gov/30912023/; https://www.seattlechildrens.org/research/partnerships/partnership-opportunities/cancer-target-validation/; https://pubmed.ncbi.nlm.nih.gov/33191402/; https://www.nature.com/articles/nrc.2017.32
12	https://www.sciencedirect.com/science/article/abs/pii/S0304419X25002094; https://www.tandfonline.com/doi/full/10.1517/17460441.2013.768984; https://www.nature.com/articles/nrclinonc.2011.34
13	https://pmc.ncbi.nlm.nih.gov/articles/PMC7961376/