Executive Summary

Clinical attrition is among the most persistent challenges in drug development. Across therapeutic areas, roughly nine out of ten candidates that enter clinical trials fail to reach patients, and late-stage failure is substantially expensive in both cost and time [Hay 2014; Sun 2022]. The root cause is well understood: the gap between preclinical models and human biology is wide, and the models used to select candidates may fail to predict how those candidates will behave in humans [Ineichen 2024]. The problem is most acute in oncology, where xenograft mouse models have shown limited predictive value for clinical activity [Johnson 2001; Gould 2015].

InSilicoTrials has developed MOL2CLIN, a fully integrated in silico workflow that evaluates a new compound from its molecular structure through a simulated clinical trial outcome. Rather than replacing experimental work, MOL2CLIN provides a rigorous, mechanistic, human-data-calibrated decision layer at the earliest stage of drug development. The result is a structured, quantitative answer to the most consequential question in R&D: is this compound worth developing for patients?

Applied to oncology, MOL2CLIN produces a structured set of outputs (predicted PK, target engagement, tumor response, safety profile, virtual trial readouts) connected through typed mechanistic relationships that form a knowledge graph linking molecular structure to predicted clinical outcome. This structure allows sponsors to ingest MOL2CLIN results directly into their internal knowledge graphs and consume them through agent frameworks, contributing structured, provenance-tagged evidence at the candidate selection stage rather than another disconnected vendor output.

This whitepaper describes the MOL2CLIN workflow applied in oncology, its scientific foundations, the human-in-the-loop decision architecture that governs each stage, the structure of its outputs, and a concrete illustrative example in EGFR-mutant non-small cell lung cancer (EGFRm NSCLC).

The Challenge: High Attrition, High Cost

Oncology drug development is the most expensive failure-prone enterprise in modern medicine. A single Phase 3 trial can cost hundreds of millions of dollars and consume years of scientific effort [DiMasi 2016]. When those trials fail, which they do the overwhelming majority of the time, the cost is borne not only by the sponsoring company but by the broader ecosystem of patients awaiting better treatments.

The dominant cause of late-stage failure is not a lack of scientific ingenuity. It is a systematic mismatch between the biological models used to select candidates and the human biology those candidates must ultimately address [Gould 2015; Mak 2014]. Preclinical in vivo models capture some aspects of tumor biology but fail to reproduce the complexity of human pharmacokinetics [Gould 2015], tumor microenvironment [Day 2015], and interpatient variability.

The strategic consequence is that the pharmaceutical industry spends enormous resources carrying forward compounds that were unlikely to work in humans from a very early stage. A decision support layer anchored in human clinical data, applied at the candidate selection stage, would change that calculus fundamentally.

A second, related challenge is becoming visible across the industry: despite substantial AI investment, attrition and cycle-time outcomes have not moved in proportion to the inputs [Jayatunga 2022]. Much of that gap reflects fragmentation. Point AI tools optimized for narrow tasks generate outputs that never reach the systems where development decisions are actually made, because those outputs arrive as standalone reports rather than as structured evidence that can be merged into a sponsor's internal knowledge base. Closing the outcomes gap requires AI tools whose outputs are structured, identifier-bearing, and reusable across the candidate-to-trial workflow: exactly the form that knowledge graphs and agent frameworks consume.

MOL2CLIN for Oncology: A Four-Stage Simulation Workflow

The MOL2CLIN workflow takes a newly designed compound, represented initially only by its molecular structure in SMILES format, and generates a quantitative prediction of how it is likely to behave in real cancer patients. The workflow is structured as a sequential four-stage pipeline with integrated human review checkpoints at each transition. Scientists retain full interpretive control throughout, while the platform surfaces the quantitative evidence needed to make rigorous, defensible decisions.

Each simulation stage is reviewed by a scientist before the next stage is run. The platform produces the evidence; the team makes the call. This structure converts a black-box prediction into an auditable, iterative decision flow, aligning with the joint FDA/EMA *Guiding Principles of Good AI Practice in Drug Development* (January 2026), which call for human-centric design and transparent documentation of the data underlying model predictions.

Each stage emits typed, identifier-bearing outputs: a compound node carrying its SMILES string and physicochemical descriptors; a target node carrying the gene and protein identifiers, the activating and resistance mutations, and the predicted binding affinity against each variant; exposure and occupancy nodes carrying the predicted PK and engagement parameters with their uncertainty ranges; and trial outcome nodes carrying the simulated endpoints. These outputs are not isolated data points; they are connected by explicit mechanistic dependencies:

- Exposure *drives* occupancy,
- Occupancy *drives* tumor inhibition
- Predicted endpoints *benchmark against* the standard of care.

Together with their stable identifiers and typed relationships, MOL2CLIN's outputs form a knowledge graph that links every prediction back to the molecular structure it derives from. This structure is what allows the outputs to be ingested into sponsor knowledge graphs and consumed by agent frameworks without an intermediate normalization step. Figure 1 provides an overview of the workflow while the full schema is illustrated on the IST-01 example in the next section.

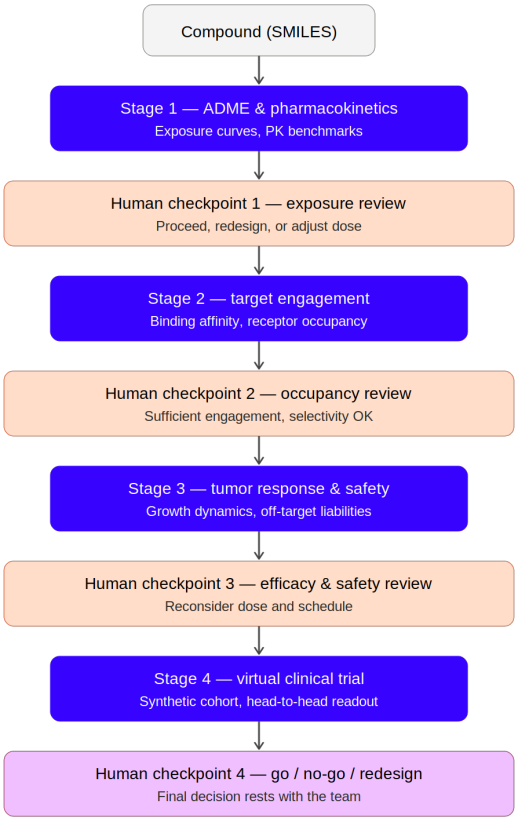

Figure 1: The MOL2CLIN workflow. Four simulation stages (deep blue) are separated by four human review checkpoints (peach). The final checkpoint (pink) is the binding go / no-go / redesign decision

Stage 1: ADME and Pharmacokinetic Prediction

Starting from molecular structure alone, the platform's ADME/PK Predictor generates key pharmacokinetic parameters governing how the drug is expected to move through the human body: oral bioavailability, systemic clearance, volume of distribution, elimination half-life, and unbound plasma fraction. Predictions are generated directly from molecular structure (SMILES) without reliance on experimentally derived physicochemical descriptors on in vitro data. The underlying AI framework is trained on a large, rigorously curated dataset comprising more than 37,000 validated experimental measurements and has undergone extensive internal and external validation across multiple ADME/PK endpoints.

Rather than producing single-point estimates, the predictor uses a conformal prediction framework to generate uncertainty-aware parameter ranges. Prediction intervals reflect the degree to which a query compound differs from compounds represented in the training data, providing scientists with a transparent view of both predicted behavior and associated uncertainty.

The predicted parameters are then fed into a mathematical pharmacokinetic model that simulates drug concentration over time across a range of doses and dosing schedules. The resulting exposure curves are compared against the InSilicoTrials library of pre-fitted PK

models for hundreds of approved oncology agents and against published clinical PK data retrieved by the platform for any approved drug in the same class.

PK is increasingly recognized across the industry as a critical enabler of reverse translational research, the question many sponsors have framed as "can ML models predict human PK?" MOL2CLIN's ADME/PK stage is built to answer that question directly, combining structure-based ML prediction with mechanistic simulation to estimate human drug exposure, and exposing the resulting parameters in a form that can be consumed by sponsor-side knowledge graphs and agents for downstream scoring.

Human-in-the-Loop Checkpoint 1

The scientist reviews simulated exposure curves and PK benchmark comparisons. If the exposure profile looks unrealistic, or if the compound cannot achieve the concentrations needed at a safe dose, the team can stop here, flag the compound for structural redesign, or adjust the dose range before any further modelling is run.

Stage 2: Target Engagement and Receptor Occupancy

Once the systemic exposure profile is established, the platform's binding affinity prediction stage estimates key target engagement parameters directly from molecular structure. Using the SMILES representation of the compound, the predictor infers target-specific binding affinity (KD), describing how tightly the drug binds to its intended protein target, together with cellular potency measures such as IC50, reflecting the concentration required to inhibit cancer cell proliferation by 50% in a relevant cellular context. Together, these parameters capture both the molecular interaction and its expected functional cellular consequence.

Predictions are generated through a structure-based bioactivity profiling framework that performs multi-fingerprint similarity searching against a high-quality reference library of more than 632,000 experimentally characterized bioactive compounds spanning over 6,000 protein targets in the ChEMBL database. By leveraging experimentally observed structure–activity relationships across a broad pharmacological space, the platform enables target affinity estimation directly from molecular structure without requiring experimentally measured binding data for the query compound.

These binding parameters are integrated into a receptor occupancy model that uses the exposure profile from Stage 1 to calculate the fraction of target protein occupied by the drug at each point in the dosing cycle. The platform pays particular attention to occupancy at the trough, the lowest point between doses, since sustained engagement between administrations is typically required for durable clinical activity.

Human-in-the-Loop Checkpoint 2

The scientist reviews receptor occupancy curves. Are occupancy levels sufficient across the dosing interval? Does the selectivity profile raise tolerability concerns? Does the compound

engage unintended targets at clinically relevant concentrations? At this point the team decides whether to proceed or to revisit the compound design.

Stage 3: Tumor Response and Safety Prediction

The tumor response model is where in silico oncology prediction achieves its most distinctive value. Rather than assuming a proportional relationship between target occupancy and tumor shrinkage, the platform simulates how a tumor grows and contracts over time under drug treatment, accounting for the balance between pharmacological effect and the tumor's intrinsic growth dynamics.

For targeted therapies such as kinase inhibitors, the model captures cytostatic and cytotoxic effects on the proliferating cell population, the delay between drug exposure and measurable tumor shrinkage, and the emergence of acquired resistance over the treatment course. The tumor biology parameters, including baseline growth rate and cellular proliferation kinetics, are derived from large public databases of clinical tumor data rather than from compound-specific experiments, ensuring that the biological context is grounded in real patient observations.

In parallel, the platform evaluates the drug's predicted safety profile at the intended clinical dose. For cytotoxic agents, dedicated models predict the impact on bone marrow progenitor cells and the risk of neutropenia. For targeted therapies, the selectivity profile from Stage 2 is mapped against known off-target safety liabilities.

Human-in-the-Loop Checkpoint 3

Before running the full trial simulation, the scientist reviews the predicted efficacy signal and the safety profile together. Is the predicted tumor response clinically meaningful? Are there off-target liabilities, predicted toxicities at the required dose, or structural alerts warranting further investigation? This is the right moment to reconsider dose and schedule before investing in the final simulation.

Stage 4: Virtual Clinical Trial Simulation

The final stage applies the predicted drug effects to a virtual patient population, a synthetic cohort that reflects the real diversity of cancer patients in terms of clinically relevant baseline characteristics such as age, body weight, disease stage, and baseline tumor burden. Each virtual patient runs through the tumor response model with their individual characteristics, generating simulated tumor measurements at each treatment assessment point.

The platform applies the same response criteria used in real oncology trials: a tumor shrinkage of at least 30% from baseline constitutes a response, while growth of more than 20% from the nadir constitutes progression. Where a synthetic patient population exists for

the relevant indication, it can serve as a comparator arm, enabling a head-to-head prediction that directly informs trial design.

Human-in-the-Loop Checkpoint 4

After the trial simulation, the clinical development team reviews the full readout. This is the decision point that matters most: go, no-go, or redesign. The platform provides a quantitative, reproducible basis for that decision, but the final judgement remains entirely with the scientists and clinicians who understand the full strategic context.

Three Exit Paths at Every Checkpoint

Each checkpoint is not a gate with two outcomes but a router with three. Approve advances the candidate to the next simulation stage. Adjust and rerun loops back into the current stage with revised inputs (a different dose, a different schedule, a different assumption). Request redesign loops all the way back to medicinal chemistry, where a new compound enters Stage 1. This structure is what makes the workflow iterative: early evidence of a problem feeds directly into compound or protocol improvement rather than ending the program.

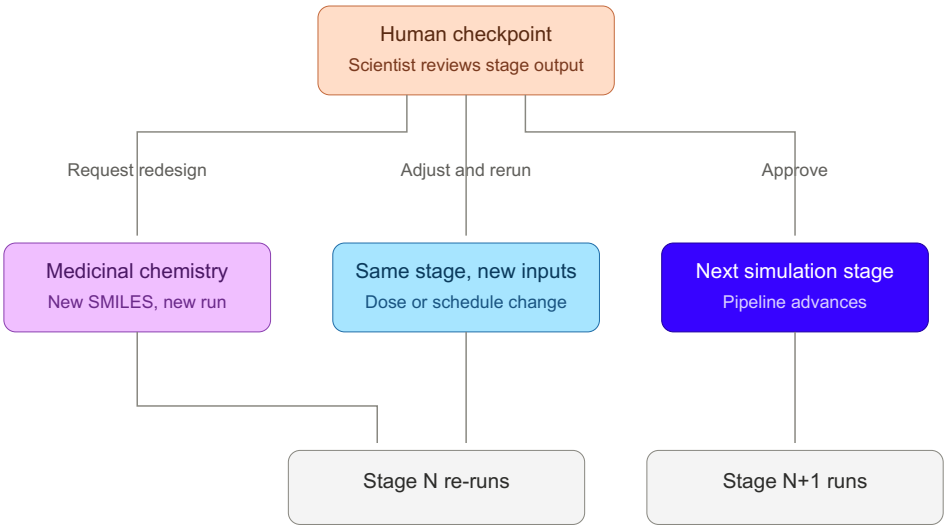

Figure 2: Decision flow at each checkpoint. Every gate has three exit paths, not two, supporting iteration on both the compound and the protocol.

Illustrative Example: IST-01 in First-Line EGFRm NSCLC

These predictions are illustrative. IST-01 is a hypothetical compound and the comparator values are taken from published FLAURA data. The point is not the specific numbers but what they represent: a structured, mechanistic, head-to-head evaluation of a new candidate against the current standard of care, completed before a single patient is enrolled.

To make the workflow concrete, consider a hypothetical fourth-generation EGFR tyrosine kinase inhibitor, designated IST-01, evaluated in first-line EGFR-mutant non-small cell lung cancer. The defining feature of a fourth-generation agent is retained activity against the C797S resistance mutation, which is the primary on-target mechanism of failure for the third-generation class. The current standard-of-care monotherapy benchmark for this indication is osimertinib, approved on the basis of the FLAURA trial [Soria 2017], in which osimertinib delivered a median progression-free survival of 18.9 months versus 10.2 months for first-generation EGFR-TKI comparators. Any new first-line EGFR-TKI monotherapy must demonstrate meaningful improvement over the FLAURA benchmark to be commercially viable, a high bar that the MOL2CLIN workflow can evaluate before significant clinical investment is made.

Starting from the molecular structure of IST-01 as a SMILES string, the platform generates the following outputs through the full four-stage workflow.

ADME and PK

Predicted oral bioavailability of 64% and half-life of 22 hours, supporting a once-daily dosing profile consistent with the approved EGFR-TKI class.

Target Engagement

IC50 of 2.4 nM against activating EGFR mutations with retained activity against the C797S resistance mutation that limits all currently approved third-generation EGFR-TKIs. Modeled receptor occupancy at the predicted clinical dose remains above 85% at trough, above the threshold expected for sustained activity in this class, with no off-target safety liabilities flagged at the predicted clinical dose.

Virtual Trial Design

600-patient virtual Phase 3 study (to match the FLAURA trial sample size) with 30-month follow-up using a FLAURA-aligned synthetic population of first-line EGFRm NSCLC patients, with osimertinib monotherapy as the comparator arm.

Predicted vs Benchmark Endpoints

Endpoint	IST-01 (Simulated)	Osimertinib (FLAURA)
Overall response rate	85%	80% (FLAURA, investigator)
Median progression-free survival	22.1 months	18.9 months

A Knowledge Graph View of the IST-01 Run

The four-stage workflow produces a structured set of outputs at each stage: predicted PK parameters, target binding affinities, receptor occupancy curves, tumor growth dynamics, safety predictions, and virtual trial endpoints. These outputs are connected by explicit mechanistic dependencies, which together form a knowledge graph linking the molecular structure of the compound to the predicted clinical outcome through every intermediate step. Figure 3 shows that graph for IST-01.

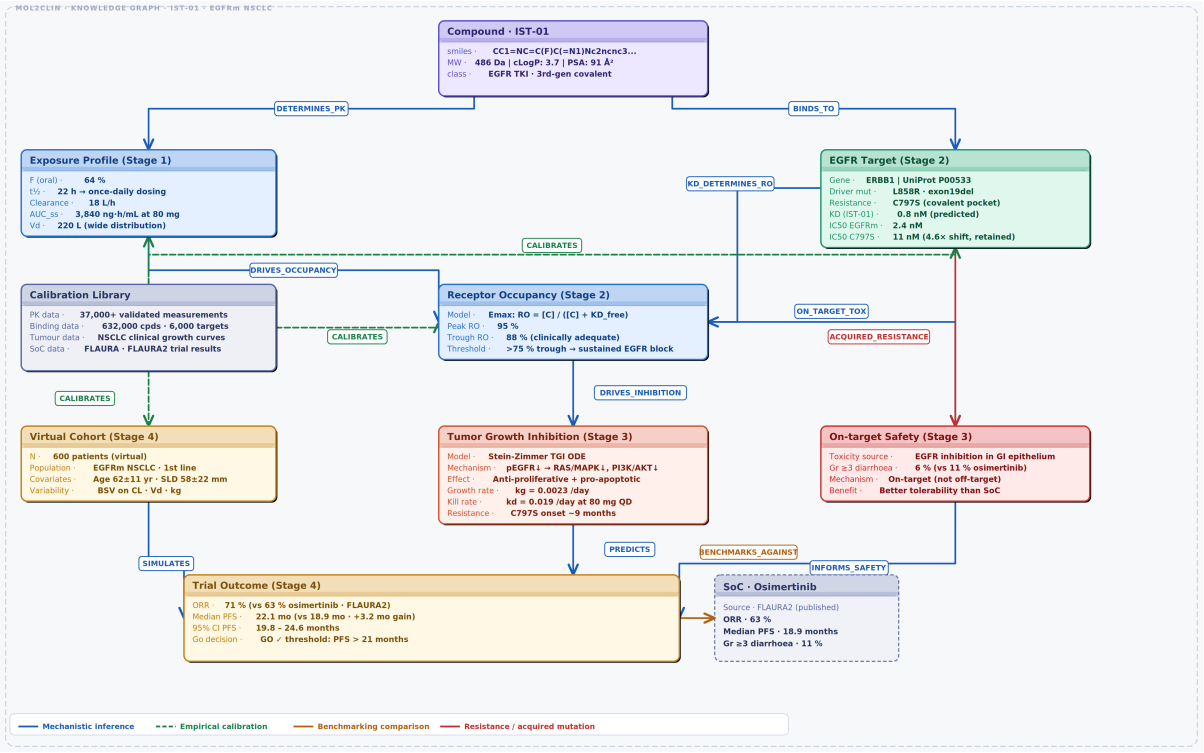

Figure 3. Knowledge Graph representing the use case of the MOL2CLIN workflow.

Nodes. Each node represents one entity in the prediction chain and carries its attributes as properties:

- The *Compound* node (IST-01) carries its SMILES, molecular weight, cLogP, PSA, and class annotation.

- The *Target* node carries the gene symbol (EGFR), UniProt accession (P00533), driver mutation (L858R / exon19del), resistance mutation (C797S), and predicted KD and IC50 against each variant.
- *Exposure Profile*, *Receptor Occupancy*, *Tumor Growth Inhibition*, and *On-target Safety* nodes carry the predicted parameters from Stages 1–3, including the model specifications used and the uncertainty ranges around each prediction.
- *Virtual Cohort* and *Trial Outcome* nodes carry the Stage 4 cohort composition and the simulated endpoints (ORR, median PFS, 95% CI, go/no-go threshold).
- A *Calibration Library* for the data behind each prediction.

Edges. Relationships between nodes are typed and carry mechanistic meaning, distinguished in Figure 3 by color:

- *Mechanistic inference* encodes the causal chain from molecular structure to predicted trial outcome.
- *Empirical calibration* links each predictive stage to the human data it was calibrated against, making the provenance of every prediction explicit.
- *Benchmarking comparison* links predicted endpoints to the standard-of-care reference.
- *Resistance / acquired mutation* links target variants to the predicted clinical consequences of resistance or on-target toxicity.

Why this matters. Because every node carries a stable identifier and every edge carries an explicit type, the IST-01 graph can be ingested into a sponsor's existing knowledge graph and merged with whatever the sponsor already knows about the compound, the EGFR target, or prior EGFR-TKI programs. An agent traversing the graph can answer questions of the form *"why does the platform predict 22.1-month PFS for IST-01?"* by walking from Trial Outcome → Tumor Growth Inhibition → Receptor Occupancy → Exposure Profile → Calibration Library, surfacing the model, parameters, and human data behind each step. Because every prediction node also carries its uncertainty range, downstream confidence-weighted scoring frameworks can weigh evidence by its provenance rather than treating all predictions as equally reliable.

Strategic Implications for Pharma and Biotech

The practical consequence of deploying MOL2CLIN in an oncology pipeline is that the first question asked about a new candidate changes. Instead of asking how quickly a compound can be synthesized for xenograft studies, the question becomes whether the predicted target engagement, tumor response, and safety profileare strong enough to justify any further investment on behalf of patients. That is a fundamentally different and more efficient starting point.

For large pharmaceutical companies, MOL2CLIN adds a rigorous in silico triage layer to complement and prioritize experimental programs. Portfolio teams can apply the workflow systematically across compound series, generating quantitative comparisons that support go/no-go decisions with a level of mechanistic detail that internal scoring methods rarely achieve. MOL2CLIN's outputs are structured along two axes that make them knowledge-graph-compatible: edges named after the mechanistic relationship they encode, and nodes that act as aggregating entities, each one bringing together everything the platform knows about a given input, intermediate prediction, or reference. This structure fits naturally into a sponsor's internal knowledge base, with each run contributing to the graph the sponsor is building. The result can be consumed directly by confidence-weighted scoring frameworks, in which each piece of evidence is weighted by its provenance and uncertainty rather than treated as equally reliable.

For emerging biotech companies with limited resources, the value is even more direct. The cost of a single clinical failure can be existential. MOL2CLIN provides a way to pressure-test a lead compound against the clinical standard of care before committing to IND-enabling studies, identifying the compounds that deserve full investment on the path to patients and those that need further optimization.

Within the broader landscape of AI for clinical development, MOL2CLIN occupies a distinct position. Most AI-driven simulation platforms operate at the trial-execution layer: they optimize protocol design, adaptive enrollment, and statistical power for compounds that have already been selected for clinical development, typically by training large statistical or digital-twin models on real-world and EHR data. That work is valuable, but it begins after the most consequential decision, which compound to advance, has already been made.

MOL2CLIN operates upstream of that decision. The workflow starts from molecular structure, before a candidate has been nominated, and chains mechanistic models of human PK, target engagement, tumor dynamics, safety, and trial response to produce a quantitative go/no-go artefact prior to IND-enabling investment. Three properties distinguish the approach. First, it is mechanistic rather than purely data-driven, meaning predictions are biologically interpretable and explain why a compound is likely to succeed or fail. Second, predicted exposure, target engagement, efficacy and safety profiles can be contextualized against published clinical data and existing pharmacological models for approved oncology agents, allowing compounds to be benchmarked against standards of care. Third the platform produces a decision-grade evaluation artefact, a virtual head-to-head assessment, that can support progression decisions using predefined quantitative criteria rather than qualitative interpretation alone.

Taken together, these properties address the recurring failure modes that have prevented AI pilots in pharma R&D from translating into pipeline outcomes. MOL2CLIN targets the layer of the value stack where the bottleneck actually sits (candidate selection) rather than the easier layers below it. It ships not as a model but as a workflow, with mechanistic models wrapped in tools, benchmarks, and human-in-the-loop checkpoints. This white paper focuses on oncology application, but MOL2CLIN can be customized to many other therapeutic areas. And because its outputs are structured for direct ingestion into sponsor knowledge graphs, evidence from each run accumulates in the sponsor's knowledge base rather than ending up as a standalone report.

Why In Silico Prediction Is the Right Starting Point for Oncology

Oncology is not like other therapeutic areas. The heterogeneity of tumors, the complexity of resistance mechanisms, the narrow therapeutic windows, and the high bar set by existing standards of care all conspire to make it the most demanding environment for drug development. It is also the area where computational prediction has the most to offer, precisely because the human clinical data needed to calibrate predictive models is available in abundance.

MOL2CLIN does not replace experimental biology. It provides a structured mechanism for asking better questions of experimental data, for understanding which parameters matter most for patients, and for identifying the compounds most likely to succeed before the investment of human subjects. The patients waiting for better cancer treatments deserve a development process that earns its way to them: experiments aimed at the questions that matter most, candidates de-risked before they ever reach a human subject, and drugs that arrive faster and at lower cost because the trial and error happened in silico, not in a clinic. That is the future MOL2CLIN is built to enable, and it begins with the decisions we make today.

References

Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. Nat Biotechnol. 2014 Jan;32(1):40-51. doi: 10.1038/nbt.2786. PMID: 24406927.

Sun D, Gao W, Hu H, Zhou S. Why 90% of clinical drug development fails and how to improve it? Acta Pharm Sin B. 2022 Jul;12(7):3049-3062. doi: 10.1016/j.apsb.2022.02.002. Epub 2022 Feb 11. PMID: 35865092; PMCID: PMC9293739.

Ineichen BV, Furrer E, Grüninger SL, Zürrer WE, Macleod MR. Analysis of animal-to-human translation shows that only 5% of animal-tested therapeutic interventions obtain regulatory approval for human applications. PLoS Biol. 2024 Jun 13;22(6):e3002667. doi: 10.1371/journal.pbio.3002667. PMID: 38870090; PMCID: PMC11175415.

Johnson JI, Decker S, Zaharevitz D, Rubinstein LV, Venditti JM, Schepartz S, Kalyandrug S, Christian M, Arbuck S, Hollingshead M, Sausville EA. Relationships between drug activity in NCI preclinical in vitro and in vivo models and early clinical trials. Br J Cancer. 2001 May 18;84(10):1424-31. doi: 10.1054/bjoc.2001.1796. PMID: 11355958; PMCID: PMC2363645.

Gould SE, Junttila MR, de Sauvage FJ. Translational value of mouse models in oncology drug development. Nat Med. 2015 May;21(5):431-9. doi: 10.1038/nm.3853. PMID: 25951530.

DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: New estimates of R&D costs. J Health Econ. 2016 May;47:20-33. doi: 10.1016/j.jhealeco.2016.01.012. Epub 2016 Feb 12. PMID: 26928437.

Mak IW, Evaniew N, Ghert M. Lost in translation: animal models and clinical trials in cancer treatment. Am J Transl Res. 2014 Jan 15;6(2):114-8. PMID: 24489990; PMCID: PMC3902221.

Day CP, Merlino G, Van Dyke T. Preclinical mouse cancer models: a maze of opportunities and challenges. Cell. 2015 Sep 24;163(1):39-53. doi: 10.1016/j.cell.2015.08.068. PMID: 26406370; PMCID: PMC4583714.

Jayatunga MKP, Xie W, Ruder L, Schulze U, Meier C. AI in small-molecule drug discovery: a coming wave? Nat Rev Drug Discov. 2022 Mar;21(3):175-176. doi: 10.1038/d41573-022-00025-1. PMID: 35132242.

Soria JC, Ohe Y, Vansteenkiste J, Reungwetwattana T, Chewaskulyong B, Lee KH, Dechaphunkul A, Imamura F, Nogami N, Kurata T, Okamoto I, Zhou C, Cho BC, Cheng Y, Cho EK, Voon PJ, Planchard D, Su WC, Gray JE, Lee SM, Hodge R, Marotti M, Rukazenkov Y, Ramalingam SS; FLAURA Investigators. Osimertinib in Untreated EGFR-Mutated Advanced Non-Small-Cell Lung Cancer. N Engl J Med. 2018 Jan 11;378(2):113-125. doi: 10.1056/NEJMoa1713137. Epub 2017 Nov 18. PMID: 29151359.