Pharmacophore

ISSN-2229-5402

Journal home page: http://www.pharmacophorejournal.com

Connecting Molecular Design with Patient-Level Treatment Decisions: An Integrative Review of Multimodal Foundation Models, Cross-Scale Biomedical Representations, and Evidence Gaps in Pharmaceutical and Precision Medicine Applications

Michael Tan[1]*, Farah Aziz[2], Peter Nagy[3], Johan de Vries[4]

1. ***Department of Pharmaceutical AI and Precision Medicine, National University of Singapore, Singapore.***
2. ***Department of Multimodal Biomedical Data Science, Universiti Putra Malaysia, Serdang, Malaysia.***
3. ***Department of Pharmaceutical Data Science and Foundation Models, Corvinus University of Budapest, Budapest, Hungary.***
4. ***Department of Precision Medicine and Cross-Scale Biomedical Modeling, Erasmus University Rotterdam, Rotterdam, Netherlands.***

ARTICLE INFO

Received:
19 May 2026
Received in revised form:
25 September 2026
Accepted:
26 September 2026
Available online:
28 October 2026

Keywords: Multimodal foundation models, Molecular design, Multi-omics integration, Drug response prediction, Precision medicine, Cross-scale biomedical representations

ABSTRACT

Multimodal foundation models are expanding the range of biomedical information that artificial intelligence can represent, combining molecular structures, omic profiles, cellular states, medical images, and clinical records within increasingly general computational systems. However, the capacity to represent heterogeneous information does not establish that molecular design outputs can inform individualized treatment decisions. This integrative review examines the evidentiary connections between pharmaceutical discovery and precision-medicine applications, distinguishing computational representation, biological validation, clinical prediction, and treatment selection. The review combines molecular-design studies, biomedical foundation-model evaluations, multimodal integration research, pharmacological response models, and clinically oriented investigations. A bounded bibliographic search and structured evidence mapping inform the synthesis; complete database-wide screening and full-text inclusion denominators remain unresolved. Evidence is organized according to biomedical scale, representation type, validation setting, and the specific decision supported. The central contribution is a cross-scale interpretation that distinguishes directly evaluated connections from relationships inferred through adjacent studies and transitions that remain untested. Molecular and cellular models provide increasingly sophisticated representations, while clinical investigations demonstrate selected routes from patient profiling to treatment matching. Evidence connecting these capabilities within a single prospectively validated molecular-design-to-treatment system remains more limited. The review identifies the methodological and translational conditions under which multimodal representations could become clinically actionable, emphasizing biological context, independent validation, comparative treatment evidence, and preservation of uncertainty across successive scales.

*This is an **open-access** article distributed under the terms of the Creative Commons Attribution-Non Commercial-Share Alike 4.0 License, which allows others to remix, and build upon the work non commercially.*

To Cite This Article: Tan M, Aziz F, Nagy P, de Vries J. Connecting Molecular Design with Patient-Level Treatment Decisions: An Integrative Review of Multimodal Foundation Models, Cross-Scale Biomedical Representations, and Evidence Gaps in Pharmaceutical and Precision Medicine Applications. Pharmacophore. 2026;17(5):82-91. https://doi.org/10.51847/o6iDrAOuiq

Introduction

Artificial intelligence has altered pharmaceutical discovery through advances in molecular representation, structure prediction, candidate generation, and computational assessment. These developments have also intensified expectations that information generated during drug discovery might eventually support more individualized treatment decisions. Yet pharmaceutical development and clinical treatment selection address different questions. Molecular design investigates which chemical entities might possess useful properties under specified experimental conditions; patient-level decision-making concerns which available intervention is appropriate for a particular person, relative to alternatives, given disease state, comorbidities, exposure, tolerability, and preferences. An improvement in the first task does not automatically establish an improvement in

Corresponding Author: Michael Tan; Department of Pharmaceutical AI and Precision Medicine, National University of Singapore, Singapore. E-mail: michael.tan@nus.edu.sg.

the second. Contemporary assessments of AI-enabled discovery emphasize the importance of distinguishing computational progress from experimentally and clinically meaningful outcomes [1].

Foundation models offer a potential means of connecting information that has traditionally been analysed within separate biomedical domains. Broad pretraining followed by adaptation can support multiple downstream tasks and allow learned representations to be transferred between related settings. In medical applications, proposed generalist architectures combine language, imaging, and other clinical inputs, while emerging molecular and cellular approaches investigate relationships among different biological modalities [2]. The attraction is evident: a shared representation could, in principle, allow molecular evidence to be interpreted alongside biological context and patient-specific observations. Nevertheless, generality within a model's training distribution and transfer between evaluation tasks do not establish biological correspondence across every represented scale.

Multimodal foundation models in molecular cell biology extend this ambition toward the integration of heterogeneous cellular measurements and biological knowledge. Such approaches create opportunities to examine relationships between molecular perturbations, cellular states, and complex phenotypes [3]. Their relevance to pharmaceutical decisions depends on whether these relationships remain informative when the biological system changes, when an intervention is introduced, or when observations originate from a different patient population. A model may align molecular and cellular information while providing little direct evidence concerning drug exposure, treatment response, or comparative clinical effectiveness.

Generative molecular design illustrates the distinction between computational capability and therapeutic relevance. Models can explore chemical structures and properties within learned or constrained design spaces, generating candidates for subsequent investigation. However, molecular novelty, predicted activity, synthetic feasibility, and biological performance are separate considerations [4]. Even experimentally supported molecular activity represents an intermediate contribution unless the evidence also addresses the disease mechanism, relevant biological environment, and conditions under which an intervention could benefit patients.

Clinical precision-medicine research provides a complementary perspective. The I-PREDICT study investigated molecularly informed personalized combination treatment in patients with advanced malignancies [5]. The WINTHER trial evaluated genomic and transcriptomic profiling as approaches to treatment matching [6]. These investigations demonstrate that patient-specific biological measurements can inform therapeutic selection. They do not establish that the selected treatments were generated by an integrated foundation model or that a molecular-design system was validated throughout the development-to-prescribing pathway.

More recent multimodal clinical modelling has incorporated heterogeneous real-world information to investigate cancer-treatment outcomes [7]. This work expands the information available for clinical prediction but introduces a further distinction: estimating outcomes associated with observed care is different from demonstrating that model-directed treatment would improve outcomes relative to an alternative strategy. Treatment allocation, confounding, missing modalities, and changes in clinical practice can influence the interpretation of such predictions.

The development of the TNIK inhibitor rentosertib provides a particularly relevant connection between AI-supported molecular discovery and human investigation. A randomized phase 2a trial evaluated the candidate in patients with idiopathic pulmonary fibrosis [8]. This represents a clinically examined drug-development lineage, although it does not demonstrate individualized treatment allocation by an end-to-end multimodal foundation model. The distinction between a clinically tested AI-discovered molecule and a patient-specific prescribing system is central to the present review.

These different evidence streams create a scientific problem that cannot be resolved by treating multimodality as a single achievement. Molecular generation, omic integration, phenotype prediction, clinical risk estimation, and treatment matching involve different units of analysis, comparator structures, validation procedures, and consequences of error. The evidence supporting an association at one level may become less informative when transferred to another. Biological heterogeneity, measurement differences, and intervention-induced changes further complicate the interpretation of apparent continuity.

The present integrative review therefore examines where molecular-design and patient-level treatment evidence genuinely connect, where separate research programmes support only an inferred relationship, and where a required transition remains unevaluated. Its distinctive focus is the evidence linking biomedical scales, rather than molecular candidate advancement, target causality, dose optimization, or clinical implementation considered independently. The review also examines how validation requirements change as models move from describing biological systems toward informing consequential treatment choices.

Materials and Methods

This integrative review was organized around a cross-scale evidence question: what is directly demonstrated when computational representations move from molecular design toward biological response, clinical prediction, and patient-level treatment decisions? The unit of analysis was therefore defined at two levels. A publication report supplied information about a particular model, experiment, benchmark, cohort, or clinical investigation. An independent evidence unit represented the underlying study, evaluation, dataset, or intervention-development programme. Multiple reports from the same programme were to be linked without treating them as independent confirmations of efficacy.

The review used an evidence-mapping approach that accommodated heterogeneous biomedical research designs. Primary methodological studies, computational benchmarks, preclinical investigations, clinical studies, and directly relevant secondary

literature were considered for their distinct evidentiary contributions. Eligible topical evidence concerned at least one meaningful interface between molecular, omic, cellular, phenotypic, clinical, or treatment-decision information. Evidence was not excluded solely because its source journal lacked Q1 classification; the journal-quartile restriction applied to the separately specified final reference allocation.

Search and selection information was recorded using the transparency principles of PRISMA 2020 while retaining the declared integrative-review design [9]. Searches were conducted on 26 September 2026 using PubMed-indexed pages accessible through an indexed web-search interface.

The accessible search output contained 13 materialized indexed-result cards. These were considered alongside a previously assembled, DOI-identified bibliographic seed of 36 publications. The search interface provided a merged ranked result set rather than complete, source-specific database exports. Consequently, the recorded results represent a bounded observed corpus and cannot be interpreted as an exhaustive retrieval census.

The combined observed material comprised 49 records. Deduplication identified one repeated indexed-search card and two publications already represented in the bibliographic seed, leaving 46 unique publications. One non-peer-reviewed preprint was excluded at the publication-status stage. Three reporting or appraisal publications were separated from the topical evidence corpus and retained as methodological guidance. The remaining 42 publication reports were potentially eligible on the basis of accessible title, abstract, and bibliographic information. Complete full-text assessment, final eligibility decisions, and the independent included-evidence-unit denominator were not established. Accordingly, these 42 reports must not be described as definitively included studies.

The distinction between retrieved records and independent evidence units was particularly important for intervention-development programmes with multiple publications. Where a compound's preclinical development and clinical evaluation appeared in separate reports, the reports were linked at programme level while retaining their distinct experimental and clinical contributions. Similarly, an indexed preprint and its subsequent peer-reviewed publication were not treated as independent studies.

Figure 1 represents the observed selection arithmetic and explicitly separates completed record processing from unresolved full-text eligibility.

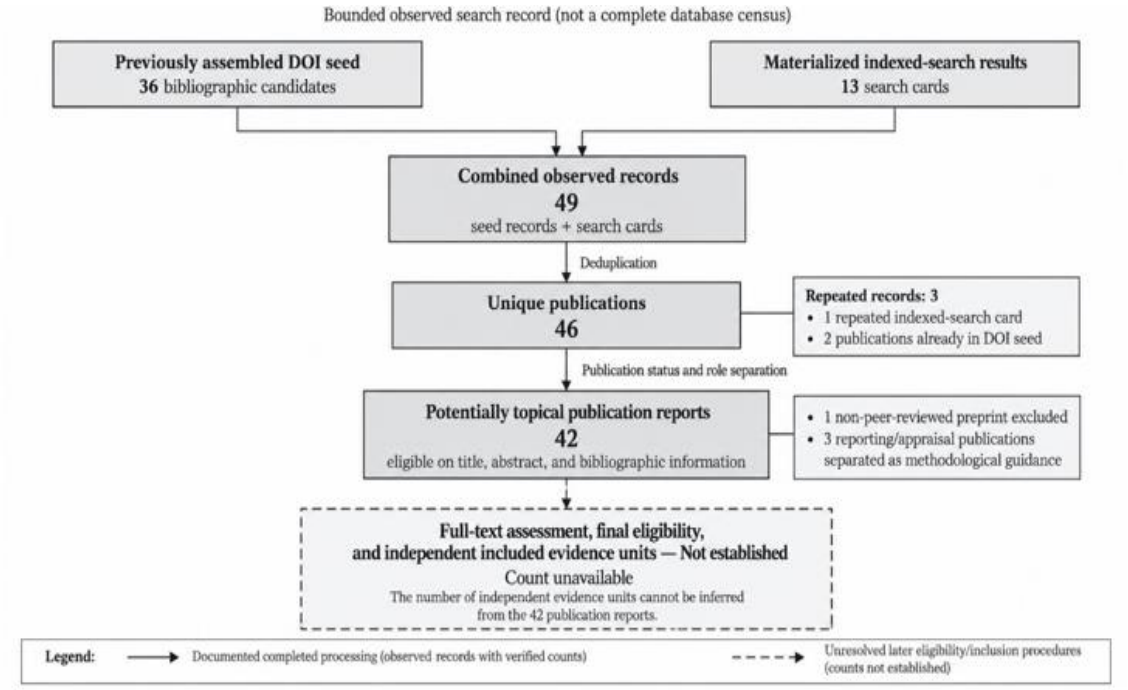


Figure 1. Observed record selection and unresolved eligibility in the cross-scale integrative review

The evidence-extraction structure was designed to retain the differences between computational performance and clinical applicability. For prediction-oriented studies, relevant reporting domains included model purpose, input representation, development data, validation population, comparator, performance measures, and intended use. TRIPOD+AI provides the reporting basis for distinguishing these elements in clinical prediction research [10]. The extraction framework additionally distinguished internal evaluation, external validation, calibration where applicable, biological validation, and evidence that a prediction informed an actual clinical decision.

Critical appraisal was design-sensitive. PROBAST+AI supplied relevant concepts for assessing risk of bias and applicability in prediction-model research [11]. Computational models also required attention to data leakage, biological or chemical distribution shift, benchmark construction, and comparison with appropriate simpler methods. Experimental investigations required consideration of assay relevance and independent confirmation, while clinical studies required attention to population,

intervention, comparator, endpoints, and treatment-allocation procedures. However, complete source-level appraisal had not been executed for the bounded corpus; no aggregate quality rating or unsupported appraisal result was assigned.

Cross-scale evidence extraction separated five principal representational levels: molecular information, omic measurements, cellular or phenotypic states, patient-level clinical information, and treatment decisions. For each reported model or study, the extraction framework identified the source and destination of a claimed relationship, the modalities used, the validation task, and whether the result established an observed association, an experimentally tested intervention, a predictive relationship, or a clinically evaluated action.

A model's ability to process multiple modalities was classified separately from its evidence of transfer between biological scales. Joint embeddings, molecular generation, drug-response prediction, and treatment allocation were therefore treated as different analytical operations, even when a study used the term foundation model for more than one application.

The available report types also differed in how directly they connected pharmaceutical and clinical information. Molecular-design studies addressed upstream chemical candidates; molecular-profiling investigations addressed patient-specific therapeutic matching; retrospective multimodal models addressed relationships within observed clinical data; and interventional studies evaluated specified treatments. **Table 1** distinguishes these report-level analytical functions without assigning unsupported included-study counts.

Table 1. Biomedical scales and decision functions represented in the documented candidate evidence

Evidence-report family	Principal biomedical scale	Representation or modality	Model or analytical approach	Application and evaluated output	Treatment-decision linkage
Pharmaceutical AI assessment	Molecular to translational	Chemical and biological evidence	Critical synthesis	Interpretation of discovery and validation claims	Conceptual; no independent treatment allocation
Generalist medical foundation-model research	Clinical and multimodal	Medical text, imaging, and clinical information	Broad pretraining and adaptation	Transfer across medical tasks	Task dependent; treatment effects not established by generality
Molecular-cellular multimodal research	Molecular, omic, cellular	Heterogeneous molecular and cellular measurements	Multimodal representation learning	Biological representation and transfer	Requires separate intervention and clinical validation
Generative molecular design	Molecular	Chemical structures and predicted properties	Generative machine learning	Candidate generation and molecular optimization	Upstream candidate evidence
Molecularly matched oncology treatment	Tumour molecular profile to patient	Molecular profiling and clinical information	Precision-treatment matching	Personalized treatment selection	Clinically examined treatment linkage
Genomic/transcriptomic treatment matching	Omic to patient	Genomic and transcriptomic profiles	Molecular matching strategy	Treatment selection in oncology	Clinically examined omic-to-treatment interface
Multimodal real-world oncology modelling	Patient and clinical outcome	Heterogeneous clinical data	Explainable multimodal AI	Prediction of observed treatment outcomes	Retrospective prediction; no demonstrated model-directed benefit
AI-discovered TNIK inhibitor programme	Molecular design to clinical intervention	Candidate-development and clinical-trial evidence	AI-supported discovery and clinical evaluation	Investigation of a specified therapeutic candidate	Clinical intervention evaluated; individualized AI treatment selection not demonstrated

Table note: Rows represent analytically distinct types of candidate evidence, not mutually exclusive counts of included studies. The table does not imply completion of the full-text eligibility assessment. Its classification distinguishes reported research functions from the review's interpretive grouping. Supporting publications are cited in the surrounding manuscript narrative; no Vancouver reference numbers are placed within the table.

The resulting synthesis strategy compares the evidentiary meaning of each scale transition before considering any broader integration. Quantitative pooling was not undertaken: the mapped evidence spans different targets, populations, modalities, interventions, comparators, and outcome definitions, and the completed full-text extraction needed to determine meta-analytic feasibility is unavailable. The subsequent synthesis consequently distinguishes observed within-task validation from cross-scale transfer, independent biological confirmation, and prospective clinical utility. These categories permit different findings to be interpreted without treating performance on an upstream benchmark as evidence of downstream treatment benefit.

Cross-Scale Evidence Synthesis

The mapped evidence indicates that biomedical foundation models connect different kinds of information at different levels of biological organization. Their outputs become relevant to treatment decisions only when the measured relationships remain valid across the particular transitions required by that decision. Three distinctions are especially consequential: chemical representation versus biological activity, cellular response versus patient response, and clinical prediction versus comparative treatment benefit.

Protein structure prediction provides an upstream example. AlphaFold established advances in predicting protein structures from sequence information, expanding access to structural representations of potential therapeutic targets [12]. AlphaFold 3 extends the modelling task to selected interactions among proteins, nucleic acids, ligands, and other molecular components [13]. These achievements support molecular interpretation and hypothesis generation. However, predicted structural correspondence does not directly measure intracellular target engagement, pharmacological exposure, or therapeutic benefit. The distinction becomes especially important when biological activity depends on conformational changes, tissue conditions, or interactions not represented in the evaluation setting.

Protein design introduces additional requirements concerning sequence, structure, and function. Functional protein-design research demonstrates how machine learning can support these connected but analytically distinct objectives [14]. At the small-molecule level, bidirectional generation of molecular structures and properties illustrates how a common model can operate across related representations [15]. Language-model evaluations further extend computational property-prediction approaches [16]. These studies establish capabilities within molecular modelling; they do not demonstrate that the resulting candidates produce clinically preferable treatment outcomes. A model may generate chemically plausible structures while leaving efficacy, exposure, toxicity, and patient-specific suitability unresolved.

The omic evidence introduces a different representational problem. Single-cell generative modelling allows transcriptional variation to be represented while accounting for properties of the observed data [17]. MOFA+ uses latent-factor modelling to integrate heterogeneous single-cell measurements [18]. TotalVI jointly models paired single-cell RNA and protein observations [19]. The latter distinction matters because genuinely paired measurements provide a different basis for cross-modal interpretation from measurements assembled across unrelated samples. Even paired molecular observations remain conditional on biological source, platform characteristics, and the state of the sampled cells.

Foundation-model approaches extend this representational capacity through large-scale pretraining. scGPT investigates generative pretraining and adaptation across single-cell multi-omic tasks [20]. Large-scale transcriptomic foundation modelling similarly examines transfer across cellular analysis settings [21]. Network-biology work demonstrates how transfer learning can support predictions involving biological relationships and selected perturbation contexts [22]. The evidence therefore extends beyond simple reconstruction of observed measurements, but the relevant experimental unit remains a cell, biological system, or defined perturbation rather than an individual treatment decision.

Independent evaluation complicates claims of general superiority. Reassessment of a single-cell foundation model found that simpler alternatives were competitive on the examined tasks [23]. This does not invalidate foundation modelling as a research strategy; it establishes that model scale and pretraining alone cannot substitute for task-specific comparison. A more general representation may offer value in some settings while contributing little under others.

Drug-response prediction creates a more direct connection between biological representations and intervention-related outcomes. The scDrugMap benchmark distinguishes pooled-data, cross-data, and zero-shot evaluation settings and reports context-dependent differences in model performance [24]. Research on unseen cell types investigates whether pretrained representations can transfer drug-response information to cellular contexts not represented during training [25]. Such evaluation addresses an important translational problem, but unseen cell types and previously unobserved patients are not interchangeable units. Patient-specific treatment involves additional pharmacological, physiological, and clinical variables.

Earlier multimodal response models remain informative comparators. MOLI combines multiple omic data types for drug-response prediction [26]. Deep learning approaches incorporating cellular genotype and drug-related features have also investigated response and synergy in human cancer-cell systems [27]. These investigations connect drug-related information to observed or predicted cellular outcomes. Their clinical relevance depends on whether cellular response remains predictive under patient-specific tumour heterogeneity, achievable exposure, combination treatment, and other biological conditions. A predicted response or synergy signal cannot independently establish the clinical benefit of assigning a treatment.

Image-based research expands the available phenotype information. Multimodal pathology foundation models investigate relationships among whole-slide images, associated textual information, and molecular measurements [28]. Whole-slide foundation modelling also evaluates image–text representations across downstream pathology tasks [29]. These studies show how morphology and related biomedical information can be represented jointly. Nevertheless, pathology-image alignment is a different evidentiary operation from showing that a molecularly generated intervention changes the disease process or improves treatment outcomes.

At the clinical scale, pretrained electronic health record representations provide another form of transferable biomedical information. Med-BERT applies contextualized representations to structured records for disease prediction [30]. Multicentre research has examined the adaptability of shared EHR foundation models across institutions [31]. These investigations address clinical-data representation and transportability. They also expose a fundamental difference between predicting recorded disease states and recommending treatment: records contain observations produced by previous clinical decisions, local coding practices, and treatment-selection processes.

Clinician-oriented drug repurposing offers a further connection between biomedical knowledge and therapeutic hypotheses. A foundation model for clinician-centred drug repurposing investigates how accumulated medical knowledge can support candidate identification [32]. Repurposing, however, concerns the evaluation of existing therapeutic agents for additional indications. Its evidentiary pathway should remain distinct from de novo molecular generation and from prospective validation of patient-specific prescribing.

Across these studies, apparent integration takes several forms: shared molecular representations, paired multi-omic measurements, perturbation-response modelling, image–molecular alignment, longitudinal clinical representations, and drug–disease knowledge integration. Their outputs cannot be placed on one evidentiary scale without preserving the task and population under which each relationship was evaluated.

Table 2 therefore distinguishes the measured alignment operation from the additional evidence needed before its output could inform an individualized treatment decision.

Table 2. Representation alignment and validation boundaries across molecular, biological, and clinical evidence

Representational interface	Information aligned	Evaluated analytical operation	Directly supported inference	Additional evidence required for treatment use
Protein sequence–structure	Amino-acid sequence and structural representation	Structure prediction	Molecular structural hypothesis	Functional activity, target engagement, exposure
Molecular structure–property	Chemical representation and predicted attributes	Property prediction or conditional generation	Candidate-level computational properties	Experimental activity, selectivity, developability
RNA–protein	Paired single-cell measurements	Joint probabilistic modelling	Within-cell cross-modal representation	Perturbation response and biological applicability
Multi-omic latent factors	Heterogeneous molecular measurements	Shared-factor integration	Statistical relationships among modalities	Independent biological interpretation
Cellular state–drug perturbation	Cell features and drug-related information	Response or synergy prediction	Response in the evaluated cellular context	Human exposure and patient-level validation
Pathology–molecular information	Whole-slide images, text, and molecular measurements	Multimodal representation and downstream prediction	Image-linked biomedical associations	Treatment-specific predictive and comparative evidence
EHR–clinical state	Recorded diagnoses and clinical features	Disease prediction or model adaptation	Clinical prediction within evaluated settings	Intervention comparison and decision-impact assessment
Knowledge graph–drug indication	Biomedical entities and therapeutic relationships	Candidate prioritization	Repurposing hypotheses	Independent intervention and treatment-outcome evaluation

Table note: The classifications describe the analytical operations represented in the mapped candidate literature. Additional evidence requirements are the review's proposed synthesis, not outcomes established by the cited studies. A shared embedding does not, by itself, establish biological causation, treatment-response validity, or clinical utility.

Integrative Framework

The evidence supports an interpretation organized around the status of the connection between two biomedical scales. A connection is directly evaluated when a study tests a specified relationship using an identifiable source, target, and validation task. It is indirectly inferred when findings from separate investigations are combined to suggest a plausible transition that has not itself been evaluated. It remains untested when the mapped evidence does not establish the proposed relationship.

This distinction clarifies what can be concluded from the TNIK inhibitor programme. The earlier development report describes AI-supported candidate discovery, preclinical investigation, and initial human evaluation.[33] Its relationship to the subsequent phase 2a trial concerns the same therapeutic development lineage. The reports therefore provide evidence extending from molecular discovery to human clinical investigation, without representing independent demonstrations of two successful compounds. They also do not establish an AI system that selects different newly designed molecules for individual patients.

A second type of connectivity concerns the relationship between molecular compounds and experimentally observed cellular function. Alignment of high-content screening resources can support transitive prediction and experimentally evaluated relationships among small molecules and biological phenotypes.[34] Such evidence extends beyond chemical similarity alone because the evaluated output concerns cellular function. Nevertheless, a cellular phenotype cannot be interpreted as an individual patient's treatment response without additional evidence concerning biological context, exposure, and clinical applicability.

The oncology treatment-matching investigations provide a contrasting downstream route. Their clinical contribution begins with patient molecular profiling and proceeds toward treatment selection, without requiring de novo molecular generation.[5] Related research using genomic and transcriptomic information likewise examines a clinically situated connection between biological measurements and therapeutic decisions.[6] These routes establish that meaningful translation can occur through patient-specific biological characterization even when an upstream generative molecular model is absent.

The mapped real-world multimodal research establishes another distinct route: heterogeneous clinical information can support prediction of observed treatment outcomes [7]. However, treatment-outcome prediction becomes an intervention-selection claim only when the decision procedure, relevant alternatives, and consequences of following the model are evaluated. Predictive associations and causal treatment effects must therefore remain distinguishable.

The proposed integrative interpretation treats these research programmes as partially connected evidence regions rather than components of one demonstrated system. Each region has its own validation unit. Molecular studies evaluate chemical or structural properties; biological studies evaluate cellular measurements or intervention responses; clinical investigations examine patients, treatment assignments, or observed outcomes. Their possible integration depends on whether a measured relationship is preserved when the unit of analysis changes.

The relationship represented by each cross-scale connection determines which claim can legitimately be carried forward. **Table 3** applies this distinction to the observed research routes and identifies the additional information needed to interpret them as patient-level treatment evidence.

Table 3. Evidentiary status of biomedical scale transitions relevant to individualized pharmaceutical decisions

Scale transition	Evidence status in mapped sources	Evaluated unit or relationship	Remaining transition
Sequence to molecular structure	Directly evaluated	Structural prediction	Experimental activity and pharmacological relevance
Molecular representation to candidate properties	Directly evaluated	Computational molecular-property tasks	Biological performance under relevant conditions
Paired omic measurements to cell-state representation	Directly evaluated	Joint molecular measurements	Intervention-specific response
Cellular state and drug features to response	Directly evaluated within defined settings	Computational or experimental cellular response	Human exposure and patient-specific validation
Pathology and molecular information to representation	Directly evaluated	Image-linked molecular or diagnostic tasks	Treatment-specific predictive validity
Patient molecular profile to treatment matching	Clinically examined	Molecularly informed treatment selection	Applicability to newly designed interventions
AI-supported candidate discovery to clinical testing	Documented in a linked programme	TNIK inhibitor development and clinical investigation	Individualized treatment selection
Multimodal clinical information to outcome prediction	Directly evaluated as prediction	Observational patient-level outcome modelling	Comparative benefit from model-directed treatment
Newly designed molecule to foundation-model-directed individualized therapy	Not established in the mapped evidence	No verified joint evaluation	Connected experimental and clinical validation

Table note: Evidence-status assignments concern the specified research tasks and the bounded mapped literature. “Not established” does not mean that a relationship is biologically impossible or absent from all published research. The table's cross-programme classification is the review's proposed synthesis.

Results and Discussion

The mapped literature supports increasing representational breadth across pharmaceutical and clinical applications, but the strength of evidence varies substantially between biological scales. An especially important limitation arises when a model evaluated on a molecular, cellular, or retrospective clinical task is subsequently interpreted as supporting treatment selection. The transition changes the target of inference, the relevant comparator, and the consequences of error.

Model evaluation must therefore reflect the decision that the model is expected to support. Research examining large language models in clinical decision-making demonstrates limitations that conventional knowledge-oriented evaluation may fail to capture [35]. For a cross-scale pharmaceutical system, appropriate assessment would require more than recovering a plausible association between a molecule and a disease. The system would need to distinguish available therapeutic alternatives, incorporate relevant patient characteristics, and identify circumstances in which its recommendation should not be followed.

Robustness is an additional concern. Evaluations of frontier models in health applications examine how model behaviour changes under challenging input conditions [36]. Such findings matter for multimodal pharmaceutical systems because missing measurements, inconsistent clinical documentation, altered biological distributions, and conflicting evidence may affect the interpretation of a recommendation. Robustness demonstrated for one model task cannot automatically be transferred to an untested drug-development or prescribing application.

The single-cell evidence introduces a parallel concern regarding generalization. Model performance can depend on the evaluation task, data distribution, and comparison method [23]. Drug-response benchmarks further distinguish performance under pooled, cross-dataset, and zero-shot conditions [24]. This variation suggests that evidence should be reported at the level of the actual transfer being tested. Successful reconstruction of a cellular representation cannot serve as an indirect substitute for evaluating response to an intervention in a biologically different population.

The clinical implications of these differences become particularly important when evidence is assembled across independent programmes. A molecular model and a clinical outcome model may share disease terminology while relying on different

populations, experimental conditions, treatment definitions, and endpoints. Combining their outputs could create an apparently coherent recommendation without establishing that the underlying quantities are compatible.

This review consequently proposes that future investigations preserve traceability from the original molecular hypothesis through the specific biological measurements and clinical observations used to support a treatment recommendation. Such traceability would permit investigators to identify whether an unexpected result reflects molecular-model error, failed biological transfer, population differences, measurement uncertainty, or an incorrect treatment-effect assumption.

A corresponding validation programme should include appropriate molecular and experimental comparators, biological evaluation in disease-relevant settings, independent testing across relevant patient populations, and prospective assessment of decisions informed by the model. Where comparative treatment claims are made, evaluation must distinguish the observed outcome under the selected intervention from the outcome expected under a relevant alternative.

The methodological limitation of the present review must also be preserved. Its bounded search record supports structured comparison of the identified literature but does not establish an exhaustive evidence census. Complete full-text eligibility assessment and source-level critical appraisal remain unresolved. Accordingly, the absence of a demonstrated end-to-end system from the mapped evidence cannot be interpreted as proof that no such system exists. The review's contribution is the distinction among documented capabilities, separately demonstrated clinical routes, and cross-scale claims requiring additional validation.

Table 4 translates these limitations into research questions whose answers would determine whether an apparent cross-scale connection can support a treatment decision.

Table 4. Validation priorities for molecular-design-to-treatment evidence

Unresolved transition	Required evaluation	Critical comparator or boundary	Claim that becomes assessable
Predicted molecular property to biological activity	Independent target-specific experiments	Alternative molecular candidates and assay conditions	Experimentally supported molecular function
Molecular activity to cellular response	Disease-relevant perturbation studies	Cell state, exposure, and biological context	Transfer to a defined cellular response
Cellular response to patient response	Patient-linked biological validation	Tissue, population, and treatment heterogeneity	Clinical applicability of response predictions
Multimodal alignment to treatment prediction	Independent treatment-specific testing	Relevant unimodal and simpler models	Incremental predictive contribution
Clinical prediction to treatment choice	Prospective decision evaluation	Standard care and competing interventions	Consequences of following model recommendations
Foundation-model adaptation to new settings	External and distribution-shift evaluation	Population, platform, and institutional changes	Transportability of a defined prediction task
AI-designed candidate to individualized use	Linked pharmaceutical and clinical investigation	Prespecified treatment-selection procedure	Evidence for the complete proposed application

Table note: These validation priorities constitute proposed research directions derived from the review's cross-scale evidence synthesis. They are not completed experiments, validated thresholds, or claims of demonstrated clinical benefit.

A further research need concerns the appropriate role of model complexity. When a simpler method performs comparably on a defined task, the clinical or pharmaceutical value of additional pretraining and multimodal integration should be established through an identifiable improvement in relevant evidence, reliability, or decision consequences. Conversely, complex representations may be useful where their distinctive information can be shown to contribute beyond appropriate comparators. This question must be investigated separately at each biomedical interface.

The same principle applies to unsuccessful or contradictory results. A failure of prediction under a new cellular distribution may identify an applicability limitation rather than invalidate the underlying molecular representation. Similarly, a clinically ineffective candidate does not necessarily establish that the upstream structural or biochemical model was incorrect. Recording where the evidence ceases to transfer is essential for interpreting both successful and unsuccessful translational programmes.

The immediate research challenge is therefore to establish which relationships remain valid when molecular, biological, and clinical information are connected within the same evaluable programme. Its resolution requires explicit intervention definitions, independently validated transitions, clinically meaningful comparators, and evidence that the resulting information changes decisions in ways that benefit patients.

Conclusion

Multimodal foundation models provide increasingly capable representations of molecular, cellular, phenotypic, and clinical information, but their ability to combine modalities does not independently establish clinically actionable cross-scale integration. The mapped literature demonstrates important capabilities within individual biomedical tasks and identifies distinct routes connecting pharmaceutical discovery to clinical investigation and molecular profiling to treatment selection. These routes should not be treated as evidence of one validated end-to-end system.

The relationship between molecular design and individualized treatment decisions becomes defensible when the relevant biological transitions are independently evaluated, the clinical application is explicitly defined, and the consequences of model-informed treatment are assessed against appropriate alternatives. Until those conditions are established, cross-scale connectivity remains an evidence-dependent proposition rather than an inherent property of multimodal modelling.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Bender A, Thomas MC, Scannell JW, Shaywitz DA, Ghiandoni GM, Greener JG, et al. Artificial intelligence in drug discovery—what it is, where we stand and the path forward. Nat Rev Drug Discov. 2026. doi:10.1038/s41573-026-01496-2
2. Moor M, Banerjee O, Abad ZSH, Krumholz HM, Leskovec J, Topol EJ, et al. Foundation models for generalist medical artificial intelligence. Nature. 2023;616(7956):259-65. doi:10.1038/s41586-023-05881-4
3. Cui H, Tejada-Lapuerta A, Brbić M, Saez-Rodriguez J, Cristea S, Goodarzi H, et al. Towards multimodal foundation models in molecular cell biology. Nature. 2025;640(8059):623-33. doi:10.1038/s41586-025-08710-y
4. Du Y, Jamasb AR, Guo J, Fu T, Harris C, Wang Y, et al. Machine learning-aided generative molecular design. Nat Mach Intell. 2024;6(6):589-604. doi:10.1038/s42256-024-00843-5
5. Sicklick JK, Kato S, Okamura R, Schwaederle M, Hahn ME, Williams CB, et al. Molecular profiling of cancer patients enables personalized combination therapy: The I-PREDICT study. Nat Med. 2019;25(5):744-50. doi:10.1038/s41591-019-0407-5
6. Rodon J, Soria JC, Berger R, Miller WH, Rubin E, Kugel A, et al. Genomic and transcriptomic profiling expands precision cancer medicine: The WINTHER trial. Nat Med. 2019;25(5):751-8. doi:10.1038/s41591-019-0424-4
7. Keyl J, Keyl P, Montavon G, Hosch R, Brehmer A, Mochmann L, et al. Decoding pan-cancer treatment outcomes using multimodal real-world data and explainable artificial intelligence. Nat Cancer. 2025;6(2):307-22. doi:10.1038/s43018-024-00891-1
8. Xu Z, Ren F, Wang P, Cao J, Tan C, Ma D, et al. A generative AI-discovered TNIK inhibitor for idiopathic pulmonary fibrosis: A randomized phase 2a trial. Nat Med. 2025;31(8):2602-10. doi:10.1038/s41591-025-03743-2
9. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71
10. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. doi:10.1136/bmj-2023-078378
11. Moons KGM, Damen JAA, Kaul T, Hooft L, Andaur Navarro C, Dhiman P, et al. PROBAST+AI: An updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ. 2025;388:e082505. doi:10.1136/bmj-2024-082505
12. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-9. doi:10.1038/s41586-021-03819-2
13. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 2024;630(8016):493-500. doi:10.1038/s41586-024-07487-w
14. Notin P, Rollins N, Gal Y, Sander C, Marks D. Machine learning for functional protein design. Nat Biotechnol. 2024;42(2):216-28. doi:10.1038/s41587-024-02127-0
15. Chang J, Ye JC. Bidirectional generation of structure and properties through a single molecular foundation model. Nat Commun. 2024;15(1):2323. doi:10.1038/s41467-024-46440-3
16. Zheng Y, Koh HY, Ju J, Nguyen ATN, May LT, Webb GI, et al. Large language models for scientific discovery in molecular property prediction. Nat Mach Intell. 2025;7(3):437-47. doi:10.1038/s42256-025-00994-z
17. Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell transcriptomics. Nat Methods. 2018;15(12):1053-8. doi:10.1038/s41592-018-0229-2
18. Argelaguet R, Arnol D, Bredikhin D, Deloro Y, Velten B, Marioni JC, et al. MOFA+: A statistical framework for comprehensive integration of multi-modal single-cell data. Genome Biol. 2020;21(1):111. doi:10.1186/s13059-020-02015-1
19. Gayoso A, Steier Z, Lopez R, Regier J, Nazor KL, Streets A, et al. Joint probabilistic modeling of single-cell multi-omic data with totalVI. Nat Methods. 2021;18(3):272-82. doi:10.1038/s41592-020-01050-x

20. Cui H, Wang C, Maan H, Pang K, Luo F, Duan N, et al. scGPT: Toward building a foundation model for single-cell multi-omics using generative AI. Nat Methods. 2024;21(8):1470-80. doi:10.1038/s41592-024-02201-0
21. Hao M, Gong J, Zeng X, Liu C, Guo Y, Cheng X, et al. Large-scale foundation model on single-cell transcriptomics. Nat Methods. 2024;21(8):1481-91. doi:10.1038/s41592-024-02305-7
22. Theodoris CV, Xiao L, Chopra A, Chaffin MD, Al Sayed ZR, Hill MC, et al. Transfer learning enables predictions in network biology. Nature. 2023;618(7965):616-24. doi:10.1038/s41586-023-06139-9
23. Boiarsky R, Singh NM, Buendia A, Amini AP, Getz G, Sontag D. Deeper evaluation of a single-cell foundation model. Nat Mach Intell. 2024;6(12):1443-6. doi:10.1038/s42256-024-00949-w
24. Wang Q, Pan Y, Zhou M, Tang Z, Wang Y, Wang G, et al. scDrugMap: Benchmarking large foundation models for drug response prediction. Nat Commun. 2026;17(1):730. doi:10.1038/s41467-025-67481-2
25. Wang Y, Liu X, Fan Y, Xie B, Cheng J, Wong KC, et al. Predicting drug responses of unseen cell types through transfer learning with foundation models. Nat Comput Sci. 2026;6(1):39-52. doi:10.1038/s43588-025-00887-6
26. Sharifi-Noghabi H, Zolotareva O, Collins CC, Ester M. MOLI: Multi-omics late integration with deep neural networks for drug response prediction. Bioinformatics. 2019;35(14):i501-9. doi:10.1093/bioinformatics/btz318
27. Kuenzi BM, Park J, Fong SH, Sanchez KS, Lee J, Kreisberg JF, et al. Predicting drug response and synergy using a deep learning model of human cancer cells. Cancer Cell. 2020;38(5):672-84.e6. doi:10.1016/j.ccell.2020.09.014
28. Xu Y, Wang Y, Zhou F, Ma J, Jin C, Yang S, et al. A multimodal knowledge-enhanced whole-slide pathology foundation model. Nat Commun. 2025;16(1):11406. doi:10.1038/s41467-025-66220-x
29. Ding T, Wagner SJ, Song AH, Chen RJ, Lu MY, Zhang A, et al. A multimodal whole-slide foundation model for pathology. Nat Med. 2025;31(11):3749-61. doi:10.1038/s41591-025-03982-3
30. Rasmy L, Xiang Y, Xie Z, Tao C, Zhi D. Med-BERT: Pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. NPJ Digit Med. 2021;4(1):86. doi:10.1038/s41746-021-00455-y
31. Guo LL, Fries J, Steinberg E, Fleming SL, Morse K, Aftandilian C, et al. A multi-center study on the adaptability of a shared foundation model for electronic health records. NPJ Digit Med. 2024;7(1):171. doi:10.1038/s41746-024-01166-w
32. Huang K, Chandak P, Wang Q, Havaldar S, Vaid A, Leskovec J, et al. A foundation model for clinician-centered drug repurposing. Nat Med. 2024;30(12):3601-13. doi:10.1038/s41591-024-03233-x
33. Ren F, Aliper A, Chen J, Zhao H, Rao S, Kuppe C, et al. A small-molecule TNIK inhibitor targets fibrosis in preclinical and clinical models. Nat Biotechnol. 2025;43(1):63-75. doi:10.1038/s41587-024-02143-0
34. Bao F, Li L, Hammerlindl H, Shen SQ, Hammerlindl S, Altschuler SJ, et al. Transitive prediction of small-molecule function through alignment of high-content screening resources. Nat Biotechnol. 2026;44(5):862-72. doi:10.1038/s41587-025-02729-2
35. Hager P, Jungmann F, Holland R, Bhagat K, Hubrecht I, Knauer M, et al. Evaluation and mitigation of the limitations of large language models in clinical decision-making. Nat Med. 2024;30(9):2613-22. doi:10.1038/s41591-024-03097-1
36. Gu Y, Fu J, Liu X, Valanarasu JMJ, Codella NCF, Tan R, et al. Evaluating the robustness and readiness of large frontier models in health AI applications. Nat Med. 2026;32(8):3023-31. doi:10.1038/s41591-026-04501-8