



Unifying the odyssey: artificial intelligence for rare disease diagnosis and therapy

Mai-Lan Ho¹, Marinka Zitnik², Ronen Azachi³, Sanjay Basu⁴, Pranav Rajpurkar², Richard Sidlow⁵

¹Department of Radiology, University of Missouri, 1 Hospital Dr., Columbia, MO 65212, USA

²Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA

³Daliio Ltd., Tel Aviv, MS, Israel

⁴Waymark Care, San Francisco, CA, USA

⁵Department of Genetics, University of Missouri, Columbia, MO, USA

Abstract

Purpose: To summarize current challenges in rare disease (RD) diagnosis and therapy, highlight recent advances in artificial intelligence (AI) for RDs, and propose a model future state for RD patient care.

Methods: Multidisciplinary expert-led narrative review summarizing modern practical challenges and rate-limiting steps in RD patient care, citing key clinical and research considerations with respect to regulatory and economic constraints.

Results: Over 10,000 known RDs collectively affect 1 in 10 Americans, a total of over 30 million people. Annually, RDs account for over \$1 trillion of annual US healthcare expenditures. Despite advances in genomic medicine, it takes 5–8 years on average to obtain an accurate diagnosis, and less than 5% of RDs currently have FDA-approved therapies. In this article, we review the history of RD diagnosis and current healthcare gaps underlying the major failures in patient care. Next, we will highlight emerging advances in genomic medicine and AI that are

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Mai-Lan Ho, mailanho@gmail.com.

Pranav Rajpurkar and Richard Sidlow are co-senior authors.

Author contributions Mai-Lan Ho, Pranav Rajpurkar, and Richard Sidlow contributed to the study conception and design. Material preparation, data collection and analysis were performed by Mai-Lan Ho. The first draft of the manuscript was written by Mai-Lan Ho, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Competing interests Ronen Azachi is the founder of Daliio Ltd. Sanjay Basu is a paid employee of HealthRight360 and Waymark, and holds stock options in Waymark and Collective Health. Richard Sidlow is an advisory board member for Guide Genetics and Egetis Pharmaceuticals.

Declarations

Ethics approval This study did not involve human subjects data.

rapidly changing the RD landscape. Finally, we propose a target future state that integrates agentic AI for diagnosis and therapy with human-in-the-loop feedback.

Conclusions: The rare disease diagnostic and therapeutic odyssey represents healthcare's most persistent failure mode. Ongoing challenges for clinical implementation involve biological modeling, manufacturing bottlenecks, and clinical trial design. We propose strategies for artificial intelligence to restructure the traditional sequence of diagnosis-then-therapy into a proactive orchestrated system delivering personalized cures at scale.

Keywords

Artificial intelligence; Genetic; Genome; Omics; Orphan; Rare

1 Introduction

For the 30 million Americans living with a rare disease (RD), the journey to a diagnosis is often a multi-year “diagnostic odyssey” of clinical uncertainty, ineffective treatments, and profound frustration. This diagnostic delay, averaging 5–8 years and involving 8 or more specialists, is more than a logistical failure; it is a fundamental therapeutic barrier. With approved treatments available for less than 5% of the over 10,000 known RDs, the window for effective intervention frequently closes before a molecular cause is even identified [1]. The economic toll is staggering: RDs account for over \$1 trillion in annual U.S. healthcare expenditures, with the therapeutic market alone valued at \$240 billion in 2025 and projected to reach \$400 billion by 2030 [2].

The RD diagnostic and therapeutic odyssey represents healthcare's most persistent failure mode. For decades, the diagnostic and therapeutic pipelines have operated in a slow, disjointed sequence. We are now at the cusp of a paradigm shift, where artificial intelligence (AI) promises to collapse this sequence into a single, integrated engine, creating a powerful feedback loop where precise diagnosis directly informs and accelerates the creation of bespoke therapies. While this future is not yet fully realized, its core components are taking shape, offering a clear perspective on how we might finally address the long tail of human genetic disease [3].

We propose that AI can fundamentally restructure this paradigm by collapsing the traditional sequential pipeline of diagnosis-then-therapy into an orchestrated engine. This integration creates a powerful feedback loop: AI-driven population-scale screening identifies patients before irreversible damage occurs; multimodal models extract deep phenotypes that bridge the genotype–phenotype gap; and the same AI systems immediately pivot to design personalized therapies using in silico and in vitro pipelines. While AI has demonstrated remarkable capabilities in accelerating RD discovery, three non-algorithmic barriers prevent clinical deployment: biological models that fail to capture in vivo complexity, manufacturing bottlenecks that lag computational design by orders of magnitude, and absent implementation frameworks for N-of-1 clinical trials. We argue that the path forward requires not another breakthrough algorithm, but the systematic alignment of computational, biological, and economic infrastructure.

In this article, we review the history of RD diagnosis and current healthcare gaps underlying the major failures in patient care. Next, we will highlight emerging advances in genomic medicine and AI that are rapidly changing the RD landscape. Finally, we propose a target future state that integrates agentic AI for diagnosis and therapy with human-in-the-loop feedback. This orchestrated ecosystem would transform RD care from reactive, fragmented interventions to a proactive, closed-loop system delivering personalized cures at scale.

2 Clinical genomics: driven by cost

As genomic testing becomes more affordable, RDs are becoming increasingly reported with over 250 new monogenic disorders discovered per year [1]. At the same time, “common” disorders such as cardiovascular disease, neurodegeneration, and cancer are becoming stratified into “rare” genetic subtypes with distinct risk profiles and therapies [4]. Clinical genomic analyses are most frequently performed on blood or tissue samples, with access and payor considerations driving the choice of test. Commercial gene panels can be ordered for specific diseases or conditions, evaluating for the highest-probability causative mutations in relevant genes. If negative, patients can reflex to whole exome sequencing (WES), covering protein-coding genes and splice-site variations within 1.5% of the total genome; or whole genome sequencing (WGS), which additionally analyzes the intronic space, mitochondrial DNA, and nucleotide-repeat variations. Currently, second-generation sequencing is most commonly utilized for routine clinical care, given the rapid throughput and relatively low cost for massively parallel reading of short DNA fragments. However, the amount of data generated by WGS is enormous, with 200 gigabytes of raw data generated per human genome, and requiring ever more complex analytical methodologies. For healthcare payors to support more advanced testing, such technologies must demonstrate cost effectiveness by enabling earlier, more precise and comprehensive diagnosis that reduces downstream indirect costs [5].

3 Rare diseases: healthcare’s failure mode

Around 72% of all RD are thought to be genetic in etiology, of which 70% begin in childhood [6]. Several single-gene disorders have the potential to cause major morbidity and mortality in the neonatal period, so-called “devastating disorders of the newborn.” In the U.S., newborn screening (NBS) is legally mandated and evaluates for between 33 and 75 potentially treatable diseases based on state guidelines. NBS is technically limited, relying on biochemical screening of blood spots from newborn heel pricks. Furthermore, no state covers all 38 core and 26 secondary conditions listed in the Recommended Uniform Screening Panel (RUSP) from the U.S. Department of Health and Human Services (HHS) [7]. Many centers are starting to adopt rapid whole exome or genome sequencing (rWES, rWGS) for broader disease coverage using the same blood spot cards. However, due to the need for quick actionable decision-making in the neonatal ICU, rapid analysis algorithms currently prioritize turn-around time (hours) over depth of information, flagging only high-probability conditions in acutely ill patients. Agnostic use of rapid newborn sequencing is currently not performed, due to ethical issues regarding variants of unknown significance (VUS) that are identified incidentally, preclinically, and/or linked to diseases without known treatments [8].

In older children and adults, RD screening is dependent on primary care physicians, who already struggle with systemic workforce shortages and limited resources. Clinically, RD patients can present with nonspecific features, such as developmental delay and abnormal facies, that are subjective and difficult to distinguish from common “bread and butter” disorders. Without a high index of suspicion for RD, patients with multisystem disorders may be overlooked or referred to many different specialists who lack RD domain expertise, resulting in “false leads” with further confusion and delays in care. At the population level, RDs create disproportionate burdens on healthcare systems with excessive ordering of costly tests and treatments that fail to help patients [9].

4 Genomic pitfalls and investigational gaps

WGS results can vary greatly between laboratories due to variable quality and depth of sequencing, differing annotation databases for meaningful genetic variants, and parental genomic information. If the raw data is available, genomic scientists may elect to reprocess it using different or updated bioinformatic algorithms and databases. Continual literature reanalyses are crucial, since a single case report can result in a VUS being reclassified as pathogenic. Functional genomics data is also highly fragmented and variable in quality, requiring better access and standardization to accelerate clinical uptake. Furthermore, a negative clinical WGS does not exclude the presence of genetic disease. Current short-read techniques can miss repeat expansions, as seen in Huntington disease and Fragile X syndrome; and large chromosomal aberrations [10]. Though more expensive, third-generation (long-read) sequencing is preferred for characterizing complex genomic regions; is sensitive to single-base modifications such as native DNA methylation; and enables more accurate haplotype phasing for disease risk and inheritance patterns [11].

Multi-omic testing is still investigational and not covered by clinical payors, but can offer improved understanding of disease mechanisms by revealing interactions between different biological levels. The emerging fields of epigenomics, transcriptomics, proteomics, metabolomics, and metagenomics can provide complementary information regarding gene expression, transcription, translation, stability, and interactions. For example, DNA methylation and histone modification play major roles in cancer and inflammation. RNA spliceopathies are responsible for retinitis pigmentosa, myelodysplastic syndromes, and muscular dystrophies. Proteinopathies are implicated in the majority of neurodegenerative disorders. Dynamic interactions between the host genome, microbiome, and exposome continually influence the state of health and disease [12].

5 Clinical pitfalls in RD evaluation

Clinical evaluation of RDs is notoriously difficult with nonspecific history, subtle physical exam findings, and unique domain expertise required to order and interpret the appropriate diagnostic tests. Additional confusion can arise due to complex heritability and/or weak effect sizes with multiple pathway interactions, epigenetic, and environmental factors. Common genetic phenomena include variable penetrance (likelihood of expressing a visible phenotype), expressivity (degree of phenotypic expression), latent phenotypes (hidden but unmasked by certain conditions), blended phenotypes (mixed due to incomplete dominance),

polygenic inheritance (multiple genes contributing to a single trait), and/or pleiotropy (single gene influencing multiple unrelated traits). Often, difficult cases require multidisciplinary expert teams to come together and identify unique combinations of subtle phenotypic alterations across multiple domains. Ongoing work in the field involves “geno–pheno correlation”: relating genotype to phenotype via vertical integration of information from micro- to macroscale [13–18].

When the evidence is inconclusive, medical geneticists must make an informed decision based on multi-domain knowledge of genetic, embryologic, and developmental mechanisms, detailed phenotypic batteries, and ethical implications for patients and their families. The concept of endophenotype or intermediate phenotype, a heritable biomarker associated with a particular disease or trait, can help bridge the geno–pheno gap. Endophenotypes (intermediate phenotypes) reflect mesoscale processes that are related to genetic mechanisms, yet are more readily measurable than standard clinical phenotypes. Examples include quantitative risk scores based on electronic health records (EHR), laboratory tests, or medical imaging [19, 20]. Deep phenotyping is a broader concept that involves synthesizing advanced information across multiple dimensions, including clinical, molecular, and behavioral, to better characterize the complexities and nuances of individual phenotypes [21, 22].

6 Healthcare data and workforce limitations

To capture RD patients early, primary care physicians need to maintain a high index of suspicion with a low threshold for genetic consultation. However, it is both unrealistic and cost-prohibitive to refer every atypical primary care patient to a geneticist. In light of widespread human workforce shortages, there is a critical need for large-scale population screening of RD using automated, objective, and high-throughput technologies. Unfortunately, data for RD is highly fragmented, with variable-quality information in multiple domains disseminated across space and time. There are no systemic data warehouses for RD analysis, and data pooling is unrealistic given that each individual’s information resides in multiple different hospitals and laboratories. Various attempts are being made to consolidate multicenter data by federal, academic, and private organizations as well as patient foundations. However, progress in data sharing is greatly hindered by privacy, ethics, and cost considerations across the vast landscape of RDs. Even bioinformatic annotation and analysis methods vary widely, with no standardized international syntaxes for consistent data exchange [23, 24].

Recent AI innovations, particularly generative AI, can help overcome data and workforce challenges with powerful foundation models discovering novel patterns within unstructured multimodal EHR and genomic data. Although AI cannot replace human experts, it can help integrate, annotate, and update multicenter datasets to accelerate research translation, mechanistic understanding, and expert review of high-probability cases. In addition, AI can help provide standardized and quantitative metrics to augment human expertise including clinical risk scores, endophenotype discovery, and composite phenotyping based on multimodal data (clinical notes, laboratory values, audio/video recordings, medical images, genomic sequences) [25–27].

In light of these exciting technological advances, it is important to concurrently address the implementation challenges of AI screening in the EHR. Primary care physicians already suffer from “alert fatigue” with repeated exposure to low-value and false-positive notifications, leading to desensitization with delayed or inadequate responses to critical warnings. Therefore, the ultimate success of AI for RD clinical support will hinge on the ability to provide high-yield, low-volume alerts that are seamlessly integrated into the daily workflow [28].

7 RD market: growing yet constrained

The RD diagnostics market is valued at \$1.5 billion in 2026 and estimated to reach \$8 billion by 2030, for a compound annual growth rate (CAGR) of 14.23%. This includes startups and established companies, government consortia, academic research groups, and patient foundations. Patient-directed services include symptom matching, networking, education, and connections to clinical trials, centers of excellence, and physician specialists based on disease. Many companies are analyzing EHR data using various AI approaches (machine learning, deep learning, natural language processing, large language models) to accelerate identification of potential RDs from structured and unstructured medical records. Federal and academic research groups maintain a number of publicly available omics databases, atlases, bioinformatic models, and literature review tools. Genomic laboratories have applied AI to various use cases including faster base calling, improved variant interpretation, error correction, and omics data integration. AI tools are also available to analyze facial morphometry, audio/video recordings, ECG, EEG, fundoscopy, skin analysis, radiology, and pathology data [29]. Progress has been limited by insufficient clinical expert feedback, highlighting the need for humans-in-the loop to achieve successful model training, evaluation, and deployment [26].

The RD treatment market is valued at \$240 billion in 2026 and estimated to reach \$400 billion by 2030, for a CAGR of 11.93%. Several companies are leveraging AI for drug discovery and repurposing, protein structure and binding sites, and clinical trials optimization [30]. The Orphan Drug Act of 1983 and additional legislation have incentivized pharmaceutical development for RDs [31, 32]. Over the past five years, 143 drugs launched with orphan drug designation in the U.S., representing over half of new U.S. Food and Drug Administration (FDA) active substance approvals [33]. A high unmet need remains, with 95% of over 10,000 RDs lacking available treatments. Even when promising therapeutic candidates are developed, pharma companies experience major challenges enrolling RD patients early enough to show treatment efficacy. This situation further emphasizes the need to accelerate RD diagnosis and route candidates for appropriate clinical trials [1].

8 AI for RD diagnosis

8.1 From fragmentation to multimodal integration

The foundational challenge in RDs has always been the lack of information, both in terms of availability and quality standards. The diagnostic odyssey is a symptom of a healthcare system struggling to recognize faint signals within noisy and unstructured clinical data. Diverse AI algorithms are being applied for RD diagnosis by government, academia,

industry, and foundations, unfortunately without a coherent focus or direction. Progress is hindered by the lack of an open and interoperable ecosystem for data sharing and collaboration across multiple diseases, sectors, and populations [26].

With decreasing costs of WGS, massive amounts of patient genomic data are becoming widely available. However, the greater challenge lies in functional interpretation—linking a variant of unknown significance (VUS) to a subtle constellation of clinical phenotypes scattered across unstructured EHRs, imaging reports, and specialist notes. This geno–pheno gap represents the central chasm in RD care: a space where human cognition struggles to synthesize disparate, longitudinal data points into a coherent diagnosis [5].

Though currently limited for RD applications, generative AI (GAI) may help address longstanding challenges with scarce, unstructured, and low-quality data in the field. Multiple groups have demonstrated the feasibility of pretrained foundation models to improve detection of common diseases. In addition, generalist models can be fine-tuned on RD datasets in order to extract unique phenotypic signatures [34]. Large language models (LLMs) tested on diverse retrospective datasets show the potential to augment RD diagnosis through conversational patient engagement, improved physician ontologies, and population EHR screening [35–37]. Vision language models (VLMs) are also being used to analyze image, video, and audio data in general medicine and subspecialties such as radiology, pathology, ophthalmology, and surgery [38–41].

Despite these promising preliminary results, prospective GAI deployment in the clinical setting requires expert human oversight at critical decision points to provide real-world feedback; define appropriate performance metrics; and mitigate failure modes including hallucination and bias. For example, LLMs struggle with complex patient cases that involve multi-step reasoning or long patient trajectories. Genomic data is also sparse and less structured in essential content compared to human language, so LLM tokenization may yield inaccurate and nonsensical results [42, 43].

In the long run, GAI algorithms trained on multimodal RD data have the potential to detect unique patterns systematically across thousands of patients. An innovative primary care paradigm would leverage AI to proactively screen entire health system populations as a first-line approach. Foundation models could be trained to sift through years of unstructured multimodal data—visit notes, laboratory values, electrophysiology recordings, and medical images—to discover new quantitative biomarkers of RDs. By integrating multiple historically siloed data sources, GAI models would construct holistic deep phenotypes to bridge the geno–pheno gap with unprecedented fidelity [44, 45].

8.2 Cost-effective screening

As modern healthcare costs skyrocket, it becomes essential to optimize cost-effective approaches to RDs. In particular, indiscriminate genomic sequencing should be avoided, as it becomes prohibitively expensive to scale and generates massive amounts of data with negligible clinical value. Furthermore, the results are challenging to analyze and can elevate psychological anxiety, while also risking false-positive diagnoses and unnecessary

treatments. A cost-effective paradigm for RD care should leverage the highest-value technologies to identify at-risk patients early, selectively apply high-probability tests, deliver effective therapies, and minimize downstream costs. If designed correctly, these workflows would cost a fraction of the current \$1 trillion annual health care expenditures for RDs, enabling early detection and treatment with minimal downstream costs (Fig. 1).

Since the majority of RDs are pediatric-onset and potentially devastating if undiagnosed, young children should be screened early in life: at birth, or even prenatally for high-risk pregnancies (using maternal cell-free DNA). As genomic testing becomes cheaper, rapid WGS should supplant NBS for broader and more consistent disease screening in newborns. Furthermore, WGS specimens and raw data can be biobanked for future reference and potential reanalysis using more comprehensive algorithms, should the patient develop symptoms later in life [8].

For older children and adults, AI screening of available EHR data should be the first step to identify and recommend high-risk individuals for genetic evaluation. AI can generate objective and reproducible thresholds for population screening using basic medical information reported by patients and/or general practitioners. This process will lift the burden from primary care physicians and ensure that limited genetics resources are directed toward the highest-probability patients. Low-access areas can be supported by telemedicine programs using smartphones and/or wearables to enable remote phenotyping of facial morphology, body proportions, voice characteristics, eye tracking, and movement patterns. AI can also augment information extraction and provide unique insights into disease mechanisms, such as new symptom clusters or subtle features that deviate from population norms [46].

AI can also help genomicists interrogate variations in data quality and bioinformatic approaches. To date, researchers have focused on computational extremes to identify genetic loci: linkage analyses to identify heritable regions with large effect sizes, and genome-wide association studies to detect population-level associations between common variants and traits. There is a critical need for standardized and publicly available omics toolkits that can perform rapid “one-stop-shop” quality assessment, genomic annotation, molecular modeling, variant curation, phenotype prediction, and personalized treatment [10, 11]. GAI agents are ideally suited for this purpose, as they can continually retrieve and learn from multiple data sources including research literature, public databases, clinical trials, and expert reviews [34].

For the most complex unsolved cases, advanced “4D multi-omic” testing may be necessary, collecting spatial, temporal, and molecular data across multiple biological scales, with AI essential for integrating these massive datasets into actionable insights. This would collect relevant spatial, temporal, and molecular data, for example the emerging fields of interactomics (total set of interactions among molecules) and exposomics (total lifetime environmental exposures) [12]. The key objective is to find patients before the critical point at which irreversible organ damage occurs, thereby closing the therapeutic window. By moving from a reactive symptom hunt to a proactive, population-scale screening process,

AI can generate a stream of accurately diagnosed patients who are ready for the next step: treatment.

9 AI for RD treatment

9.1 Many barriers, few successes

Over the past decades, the Orphan Drug Act and FDA special designations for unmet medical needs have spurred the pharmaceutical industry's interest in developing RD therapies [31–33]. However, current initiatives remain insufficient to address increasing patient needs. Bringing a pharmaceutical to market averages 10–15 years and over \$1 billion of investments in research & development, clinical trial, regulatory, and manufacturing costs [47]. Furthermore, around 90% of new drugs fail in clinical trials, with half due to human toxicities that were not predicted by preclinical studies [48].

To facilitate bench-to-bedside translation, the FDA recently established the plausible mechanism pathway to approve personalized RD therapies, for which traditional clinical trials are nearly impossible. This allows for conditional approval based on a small number of patients with strong biological evidence and meaningful clinical improvement [49]. Furthermore, both the FDA [50] and National Institutes of Health (NIH) [51] have announced prioritization of human-based research technologies and reduction of animal experimentation, based on cost, ethics, and accuracy concerns [52].

Attempts to accelerate drug development pipelines should be coupled with strong governance frameworks to ensure that AI systems are deployed ethically, responsibly, and fairly. Many controversies around RDs are attributed to poor understanding and ineffective therapies, resulting in insurance/employment discrimination and social stigmatization. For example, polygenic risk scores estimate the likelihood of developing certain diseases relative to the population, but can influence providers to overdiagnose preclinical diseases and overtreat secondary findings. At the other extreme, positive testing for delayed-onset and intractable conditions (such as Huntington disease or amyotrophic lateral sclerosis) can yield severe psychological distress with negligible likelihood of cure [53].

9.2 Intelligent therapeutic design

The ultimate potential of AI in RDs lies in coupling the aforementioned diagnostic engine to a novel therapeutic development workflow. The current linear progression of molecular target discovery, preclinical animal models, and large-scale human clinical trials is ill-suited to ultrarare and nanorare disorders affecting fewer than 30–100 patients worldwide. Rather, the future of AI-designed therapeutics lies in a rapid, iterative, and closed-loop system based on patient-specific biology. After RD diagnostic resolution, therapy begins by identifying the biological target (causative molecular defect) and aligning it with an appropriate therapeutic modality. For drug design, target properties define the route of administration. Loss-of-function targets with preserved reading frame and accessible tissues point to gene augmentation or mRNA therapy. Gain-of-function or dominant-negative targets prioritize allele-selective silencing, RNA interference, or editing. Splice defects can be treated with antisense oligonucleotides. Extracellular or secreted proteins favor antibodies or protein

therapeutics. Enzyme deficiencies suggest replacement or chaperones. Cell loss or circuit dysfunction motivates cell therapy approaches. Network-level dysregulation with multiple interrelated effects often suits small molecules or rational combinations [54].

With the target and modality selected, drug design and optimization become a search over a shared intervention space. AI can make pharmaceutical routing data-driven; we envision that multimodal AI learners will fuse genomics, proteomics, imaging, and clinical trajectories to infer target class, quantify druggability, and anticipate delivery constraints such as tissue access or immune context. Structure-aware and atom-scale AI models can evaluate whether a particular molecular feature pocket or splice pattern is amenable to a given modality [55]. Knowledge-graph reasoning aligns the target with pathways and compensatory circuits [56]. Drug repurposing [57] and *de novo* design [58] use patient-conditioned priors and scoring functions. Retrieval-augmented and agentic systems surface similar cases, mechanistic evidence, and preclinical results to match the target and modality [59]. Generative [60, 61] and quantum computing [62] models can propose sequence edits, chemotypes, delivery vectors, or protein scaffolds consistent with biophysical and exposure constraints. Combinatorial optimizers explore multi-drug regimens and drug-plus-novel-compound strategies under interaction and exposure limits and simulate dose and schedule [63]. Multi-objective optimization trades off benefit, risk, and time to treatment, producing a ranked slate of drugs or drug candidates rather than a single option [64]. Each AI-generated drug proposal carries a structured rationale, calibrated uncertainty, and counterfactuals that reveal which observations or design changes would shift rank. This target-to-modality decision can fix the intervention design space and hand off a target-conditioned plan to candidate generation. A policy-aware model can additionally integrate benefit, safety, feasibility, and time to treatment and recommend hybrid strategies, for example an existing drug paired with a novel compound in order to balance efficacy and safety [65].

To address current challenges in RD therapeutics research, we propose that progression from in silico proposals to treatment should pass through key decision gates that prioritize practical utility. **Gate 1** evaluates expected net benefit at the patient level, anchored to minimal clinically important differences for outcomes that matter. **Gate 2** tests concordance with human-derived biologic models such as induced cells, organoids, or engineered tissues using a battery that integrates molecular, morphological, and functional readouts. In the future, AI can align digital twin/avatar responses with patient features to estimate transport from bench to bedside and minimize the need for animal testing [66]. **Gate 3** verifies efficacy and safety margins in context, including dose and route feasibility, interaction risk for combinations, and operational readiness in the clinical setting. Drug candidates that clear the gates will move to treat-now pathways. Candidates that do not clear a gate require additional targeted experiments to optimize pathways or design iterations to lift the limiting criteria. This approach will enable continual recalibration, rapid escalation when net benefit exceeds threshold, and de-escalation when validity erodes (Fig. 2).

9.3 Real-world bottlenecks

The powerful vision for integrated RD theranostics is presently impeded by three stubborn chokepoints [23, 67, 68]. First, current biological models, both in vitro and in silico, still miss too much real-world complexity. The most advanced “virtual cell” models couple metabolism and signaling pathways, but do not include emergent higher-order features such as the organization of phase-separated compartments or the dynamics of mitochondrial heteroplasmy, leading to fragile dose-response predictions [69]. In the lab, patient-derived organoids fail to capture the multi-system crosstalk that critically impacts drug response, including cell maturation, extracellular matrix, and integrated neural, vascular, and hematopoietic networks [70]. In light of this fidelity gap, many AI-nominated hits that succeed in theory will continue to fail when they enter the complex environment of a living organism. Until synthetic models can better recapitulate the full spectrum of biologic complexity from nano- to macroscale, AI search engines will keep reporting answers that biology later vetoes.

Second, the throughput of Good Manufacturing Practices (GMP)-compliant facilities currently constrains the treatment calendar. Although an AI algorithm can design a personalized ASO or adeno-associated virus (AAV) vector in hours, actual physical production of a clinical-grade batch may take months or quarters. Individualized therapies must pass through the contract development and manufacturing (CDMO) sector, which is already running short on skilled staff and clean-room slots. The market for therapeutic oligonucleotides is projected to triple by 2029, a growth rate that far outpaces facility expansion. There is a frustrating disconnect with the discovery process being orders of magnitude faster than the production process, resulting in delayed access to potentially life-saving treatments. In future, AI could help address these physical limitations by enabling robotic automation, integrated process monitoring, predictive maintenance, formulation optimization, quality control, supply chain management, patient distribution, and regulatory intelligence [71].

Third, the economic model for RD therapies remains unsolved. The question of who pays for expensive and personalized therapies in this ultra-niche market is a critical barrier to widespread adoption. Philanthropic models like the n-Lorem Foundation have been pioneers, treating dozens of “nano-rare” patients with personalized ASO technologies and proving the technical feasibility of the N-of-1 approach [72]. However, charity alone cannot support the hundreds already approved for help, let alone the future demands. AI may make the discovery of a cure astonishingly cheap, but if the final product remains prohibitively expensive and unreimbursed, it will remain inaccessible to patients. Therefore, close cooperation between government, foundation, and industry partners is sorely needed to create a sustainable infrastructure for research, development, commercialization, and distribution of RD technologies. Until actuarial approaches and reimbursement policies are established for therapies that cannot achieve the statistical power of a traditional clinical trial, mainstream health systems and payors cannot budget for bespoke treatments [73] (Fig. 3).

10 AI orchestration: end-to-end care

Within the booming RD market, numerous research and commercial AI tools are emerging for each stage of care: population screening and deep phenotyping, genomic interpretation and integrated diagnosis, target identification and therapy design, model validation and iteration, manufacturing and deployment. These initial prototypes can be further developed into autonomous AI agents that are fine-tuned to specific domains, perform complex multi-step reasoning, update knowledge from database/literature searches, and dynamically learn from feedback. The final layer involves an AI orchestrator agent that connects modular domain-specific agents and integrates expert human supervision at critical decision points. This multi-agent system can be utilized to manage the RD diagnostic and therapeutic odyssey from end-to-end, including reinforcement learning, experimentation, and collaborative decision making [74–77]. Orchestrated AI represents the ideal paradigm for RD care: seamless connection of evolving data, research, and testing capabilities to find a personalized cure for each individual [78].

11 Conclusion

Despite advances in genomic technologies, our current healthcare system is fragmented and fails to correctly diagnose and treat RD patients in a timely manner. AI shows transformative potential in accelerating RD diagnosis and treatment, yet significant implementation challenges remain. The path forward requires a focused effort on the longstanding barriers that constrain deployment: improving the biological fidelity of models, scaling distributed and modular manufacturing, and developing novel payor reimbursement frameworks. The future of RD care will not be delivered by a single breakthrough algorithm, but by the systematic orchestration of these computational, biological, and economic solutions. By building this unified ecosystem, we can transition from celebrating one-off miracles to delivering personalized medicine as the standard of care.

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

References

1. Rare Diseases. National Institutes of Health. <https://www.nih.gov/about-nih/nih-turning-discovery-into-health/promise-precisionmedicine/rare-diseases>. Accessed 3 Aug 2025.
2. The National Economic Burden of Rare Disease Study. Every-Life Foundation. <https://everylifefoundation.org/burden-landing/>. Accessed 3 Aug 2025.
3. Shen T, Lee A, Shen C, et al. The long tail and rare disease research: the impact of next-generation sequencing for rare Mendelian disorders. *Genet Res (Camb)*. 2015;97:e15. [PubMed: 26365496]
4. Reinholdt L, Chesler E, Pera M, Rosenthal N. The rare-to-common disease journey: a winding road to new therapies. *Trends Genet*. 2025 Jun 2:S0168–9525(25)00110–6. 10.1016/j.tig.2025.05.003. Epub ahead of print.
5. Xi Q, Jin S, Morris S. Economic evaluations of predictive genetic testing: A scoping review. *PLoS One*. 2023;18(8):e0276572. 10.1371/journal.pone.0276572. [PubMed: 37531363]
6. eClinicalMedicine. Raising the voice for rare diseases: under the spotlight for equity. *EClinicalMedicine*. 2023;57:101941. 10.1016/j.eclinm.2023.101941. [PubMed: 37064509]

7. Newborn Screening in Your State. <https://newbornscreening.hrsa.gov/your-state>. Accessed 3 Aug 2025.
8. Yang L, Chen J, Shen B. Newborn Screening in the Era of Precision Medicine. *Adv Exp Med Biol*. 2017;1005:47–61. 10.1007/978-981-10-5717-5_3. [PubMed: 28916928]
9. Baynam G, Hartman AL, Letinturier MCV, Bolz-Johnson M, Carrion P, Grady AC, Dong X, Doms M, Dreyer L, Graessner H, Granados A, Groza T, Houwink E, Jamuar SS, Vasquez-Loarte T, Tumiene B, Wiafe SA, Bjornson-Pennell H, Groft S. Global health for rare diseases through primary care. *Lancet Glob Health*. 2024;12(7):e1192–9. 10.1016/S2214-109X(24)00134-7. [PubMed: 38876765]
10. Austin-Tse CA, Jobanputra V, Perry DL, Bick D, Taft RJ, Venner E, Gibbs RA, Young T, Barnett S, Belmont JW, Boczek N, Chowdhury S, Ellsworth KA, Guha S, Kulkarni S, Marcou C, Meng L, Murdock DR, Rehman AU, Spiteri E, Thomas-Wilson A, Kearney HM, Rehm HL. Medical genome Initiative*. Best practices for the interpretation and reporting of clinical whole genome sequencing. *NPJ Genom Med*. 2022;7(1):27. 10.1038/s41525-022-00295-z. [PubMed: 35395838]
11. Scarano C, Veneruso I, De Simone RR, Di Bonito G, Secondino A, D'Argenio V. The Third-Generation sequencing challenge: novel insights for the omic sciences. *Biomolecules*. 2024;14(5):568. 10.3390/biom14050568. [PubMed: 38785975]
12. Ali SS, Li Q, Agrawal PB. Implementation of multi-omics in diagnosis of pediatric rare diseases. *Pediatr Res*. 2025;97(4):1337–44. 10.1038/s41390-024-03728-w. Epub 2024 Nov 19. [PubMed: 39562738]
13. Cooper DN, Krawczak M, Polychronakos C, Tyler-Smith C, Kehrer-Sawatzki H. Where genotype is not predictive of phenotype: towards an Understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum Genet*. 2013;132(10):1077–130. 10.1007/s00439-013-1331-2. Epub 2013 Jul 3. [PubMed: 23820649]
14. Payne JL, Wagner A. Latent phenotypes pervade gene regulatory circuits. *BMC Syst Biol*. 2014;8:64. 10.1186/1752-0509-8-64. [PubMed: 24884746]
15. Rosina E, Pezzani L, Pezzoli L, Marchetti D, Bellini M, Pilotta A, Calabrese O, Nicastro E, Cirillo F, Cereda A, Scatigno A, Milani D, Iacone M. Atypical, Composite, or blended phenotypes: how different molecular mechanisms could associate in Double-Diagnosed patients. *Genes (Basel)*. 2022;13(7):1275. 10.3390/genes13071275. [PubMed: 35886058]
16. Crouch DJM, Bodmer WF. Polygenic inheritance, GWAS, polygenic risk scores, and the search for functional variants. *Proc Natl Acad Sci U S A*. 2020;117(32):18924–33. 10.1073/pnas.2005634117. Epub 2020 Aug 4. [PubMed: 32753378]
17. Watanabe K, Stringer S, Frei O, Umičević Mirkov M, de Leeuw C, Polderman TJC, van der Sluis S, Andreassen OA, Neale BM, Posthuma D. A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet*. 2019;51(9):1339–1348. <https://doi.org/10.1038/s41588-019-0481-0>. Epub 2019 Aug 19. Erratum in: *Nat Genet*. 2020;52(3):353. doi: 10.1038/s41588-019-0571-z. [PubMed: 31427789]
18. Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. *Cell*. 2017;169(7):1177–86. 10.1016/j.cell.2017.05.038. [PubMed: 28622505]
19. Liu C, Gershon ES. Endophenotype 2.0: updated definitions and criteria for endophenotypes of psychiatric disorders, incorporating new technologies and findings. *Transl Psychiatry*. 2024;14(1):502. 10.1038/s41398-024-03195-1. [PubMed: 39719446]
20. Nava-Gonzalez EJ, Gallegos-Cabriaes EC, Leal-Berumen I, Bastarrachea RA. Mini-Review: the contribution of intermediate phenotypes to GxE effects on disorders of body composition in the new OMICS era. *Int J Environ Res Public Health*. 2017;14(9):1079. 10.3390/ijerph14091079. [PubMed: 28926971]
21. DeBoever C, Tanigawa Y, Aguirre M, McInnes G, Lavertu A, Rivas MA. Assessing digital phenotyping to enhance genetic studies of human diseases. *Am J Hum Genet*. 2020;106(5):611–22. 10.1016/j.ajhg.2020.03.007. Epub 2020 Apr 9. [PubMed: 32275883]
22. Son JH, Xie G, Yuan C, Ena L, Li Z, Goldstein A, Huang L, Wang L, Shen F, Liu H, Mehl K, Groopman EE, Marasa M, Kiryluk K, Gharavi AG, Chung WK, Hripacsak G, Friedman C, Weng C, Wang K. Deep phenotyping on electronic health records facilitates genetic diagnosis by clinical exomes. *Am J Hum Genet*. 2018;103(1):58–73. Epub 2018 Jun 28. [PubMed: 29961570]

23. McMurry J, Sizer A, Allaway R et al. Critical bottlenecks in rare disease research and care: a community perspective. <https://zenodo.org/records/14907384>. Accessed August 3, 2025.
24. Taruscio D, Gahl WA. Rare diseases: challenges and opportunities for research and public health. *Nat Rev Dis Primers*. 2024;10(1):13. 10.1038/s41572-024-00505-1. [PubMed: 38424095]
25. Ou P, Wen R, Shi L, Wang J, Liu C. Artificial intelligence empowering rare diseases: a bibliometric perspective over the last two decades. *Orphanet J Rare Dis*. 2024;19(1):345. 10.1186/s13023-024-03352-1. [PubMed: 39272071]
26. Karako K. Artificial intelligence applications in rare and intractable diseases: Advances, challenges, and future directions. *Intractable Rare Dis Res*. 2025;14(2):88–92. 10.5582/irdr.2025.01030. [PubMed: 40485885]
27. Nishat SMH, Shahid Tanweer A, Alshamsi B, Shaheen MH, Shahid Tanveer A, Nishat A, Alharbat Y, Alaboud A, Almazrouei M, Ali-Mohamed RA. Artificial intelligence: A new frontier in rare disease early diagnosis. *Cureus*. 2025;17(2):e79487. 10.7759/cureus.79487. [PubMed: 40135033]
28. Gani I, Litchfield I, Shukla D, Delanerolle G, Cockburn N, Pathmanathan A. Understanding alert fatigue in primary care: qualitative systematic review of general practitioners attitudes and experiences of clinical alerts, Prompts, and reminders. *J Med Internet Res*. 2025;27:e62763. 10.2196/62763. [PubMed: 39918864]
29. Grand View Research. Rare Disease Genetic Testing Market. (2025–2030). <https://www.grandviewresearch.com/industry-analysis/rare-disease-genetic-testing-market>. Accessed August 3, 2025.
30. Rare Diseases Treatment Market. (2025–2030). <https://www.grandviewresearch.com/industry-analysis/rare-diseases-treatment-market-report>. Accessed August 3, 2025.
31. Mikami K. Orphans in the market: the history of orphan drug policy. *Soc Hist Med*. 2019;32(3):609–30. 10.1093/shm/hkx098. Epub 2017 Nov 27. [PubMed: 31384102]
32. Michaeli DT, Michaeli T, Albers S, Boch T, Michaeli JC. Special FDA designations for drug development: orphan, fast track, accelerated approval, priority review, and breakthrough therapy. *Eur J Health Econ*. 2024;25(6):979–97. 10.1007/s10198-023-01639-x. Epub 2023 Nov 14. [PubMed: 37962724]
33. Fermaglich LJ, Miller KL. A comprehensive study of the rare diseases and conditions targeted by orphan drug designations and approvals over the Forty years of the orphan drug act. *Orphanet J Rare Dis*. 2023;18(1):163. 10.1186/s13023-023-02790-7. [PubMed: 37353796]
34. Liu X, Liu H, Yang G, Jiang Z, Cui S, Zhang Z, Wang H, Tao L, Sun Y, Song Z, Hong T, Yang J, Gao T, Zhang J, Li X, Zhang J, Sang Y, Yang Z, Xue K, Wu S, Zhang P, Yang J, Song C, Wang G. A generalist medical Language model for disease diagnosis assistance. *Nat Med*. 2025;31(3):932–42. 10.1038/s41591-024-03416-6. Epub 2025 Jan 8. [PubMed: 39779927]
35. Ao G, Chen M, Li J, Nie H, Zhang L, Chen Z. Comparative analysis of large Language models on rare disease identification. *Orphanet J Rare Dis*. 2025;20(1):150. 10.1186/s13023-025-03656-w. [PubMed: 40165285]
36. Zhong W, Sun M, Yao S, Liu Y, Peng D, Liu Y, Yang K, Gao H, Yan H, Hao W, Yan Y, Yin C. Enhancing the accuracy of human phenotype ontology identification: comparative evaluation of multimodal large Language models. *J Med Internet Res*. 2025;27:e73233. 10.2196/73233. [PubMed: 40456109]
37. Chimirri L, Caufield JH, Bridges Y, Matentzoglou N, Gargano M, Cazalla M, Chen S, Danis D, Dingemans AJM, Gehle K, Gehle P, Graefe ASL, Gu W, Ladewig MS, Lapunzina P, Nevado J, Niyonkuru E, Ogishima S, Seelow D, Tenorio Castaño JA, Turnovec M, de Vries BBA, Wang K, Wissink K, Yüksel Z, Zucca G, Haendel MA, Mungall CJ, Reese J, Robinson PN. Consistent performance of large Language models in rare disease diagnosis across ten Languages and 4917 cases. *EBioMedicine*. 2025;121:105957. 10.1016/j.ebiom.2025.105957. Epub 2025 Oct 14. [PubMed: 41092581]
38. Zhang K, Zhou R, Adhikarla E, Yan Z, Liu Y, Yu J, Liu Z, Chen X, Davison BD, Ren H, Huang J, Chen C, Zhou Y, Fu S, Liu W, Liu T, Li X, Chen Y, He L, Zou J, Li Q, Liu H, Sun L. A generalist vision-language foundation model for diverse biomedical tasks. *Nat Med*. 2024;30(11):3129–41. 10.1038/s41591-024-03185-2. Epub 2024 Aug 7. [PubMed: 39112796]

39. Ferber D, El Nahhas OSM, Wölflein G, Wiest IC, Clusmann J, Leßmann ME, Foersch S, Lammert J, Tschochohei M, Jäger D, Salto-Tellez M, Schultz N, Truhn D, Kathner JN. Development and validation of an autonomous artificial intelligence agent for clinical decision-making in oncology. *Nat Cancer*. 2025;6(8):1337–49. 10.1038/s43018-025-00991-6. Epub 2025 Jun 6. [PubMed: 40481323]
40. Ding T, Wagner SJ, Song AH, Chen RJ, Lu MY, Zhang A, Vaidya AJ, Jaume G, Shaban M, Kim A, Williamson DFK, Robertson H, Chen B, Almagro-Pérez C, Doucet P, Sahai S, Chen C, Chen CS, Komura D, Kawabe A, Ochi M, Sato S, Yokose T, Miyagi Y, Ishikawa S, Gerber G, Peng T, Le LP, Mahmood F. A multimodal whole-slide foundation model for pathology. *Nat Med*. 2025;31(11):3749–61. 10.1038/s41591-025-03982-3. Epub 2025 Nov 5. [PubMed: 41193692]
41. Shi D, Zhang W, Yang J, Huang S, Chen X, Xu P, Jin K, Lin S, Wei J, Yusufu M, Liu S, Zhang Q, Ge Z, Xu X, He M. A multimodal visual-language foundation model for computational ophthalmology. *NPJ Digit Med*. 2025;8(1):381. 10.1038/s41746-025-01772-2. [PubMed: 40542189]
42. Yao G, Zhang W, Zhu Y, Wong UK, Zhang Y, Yang C, Shen G, Li Z, Gao H. Comparing the accuracy of large Language models and prompt engineering in diagnosing Realworld cases. *Int J Med Inf*. 2025;203:106026. Epub 2025 Jun 25.
43. Kruse M, Hu S, Derby N, Wu Y, Stonbraker S, Yao B, Wang D, Goldberg E, Gao Y. Large Language models with Temporal reasoning for longitudinal clinical summarization and prediction. *Find ACL EMNLP*. 2025;2025:20715–35. 10.18653/v1/2025.findings-emnlp.1128. [PubMed: 41399802]
44. Jacobsen JOB, Kelly C, Cipriani V, Research Consortium GE, Mungall CJ, Reese J, Danis D, Robinson PN, Smedley D. Phenotype-driven approaches to enhance variant prioritization and diagnosis of rare disease. *Hum Mutat*. 2022;43(8):1071–81. 10.1002/humu.24380. Epub 2022 Apr 27. [PubMed: 35391505]
45. Hou YC, Yu HC, Martin R, Cirulli ET, Schenker-Ahmed NM, Hicks M, Cohen IV, Jönsson TJ, Heister R, Napier L, Swisher CL, Dominguez S, Tang H, Li W, Perkins BA, Barea J, Rybak C, Smith E, Duchicela K, Doney M, Brar P, Hernandez N, Kirkness EF, Kahn AM, Venter JC, Karow DS, Caskey CT. Precision medicine integrating whole-genome sequencing, comprehensive metabolomics, and advanced imaging. *Proc Natl Acad Sci U S A*. 2020;117(6):3053–62. 10.1073/pnas.1909378117. Epub 2020 Jan 24. [PubMed: 31980526]
46. Chen FH, Hartman AL, Letinturier MCV, Antoniadou V, Baynam G, Bloom L, Crimi M, Della Rocca MG, Didato G, Houge SD, Jonker A, Kawome M, Mueller F, O'Brien J, Puri RD, Ryan N, Thong MK, Tumiené B, Parisi MA. Telehealth for rare disease care, research, and education across the globe: A review of the literature by the IRDiRC telehealth task force. *Eur J Med Genet*. 2024;72:104977. 10.1016/j.ejmg.2024.104977. Epub 2024 Oct 5. [PubMed: 39374775]
47. Wouters OJ, McKee M, Luyten J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009–2018. *JAMA*. 2020;323(9):844–853. doi: 10.1001/jama.2020.1166. [PubMed: 32125404] Erratum in: *JAMA*. 2022;328(11):1110. doi: 10.1001/jama.2022.14317. Erratum in: *JAMA*. 2022;328(11):1111. <https://doi.org/10.1001/jama.2022.14733>.
48. 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;12(7):3049–62. 10.1016/j.apsb.2022.02.002. Epub 2022 Feb 11. [PubMed: 35865092]
49. Prasad V, Makary MA, FDA's New Plausible Mechanism Pathway. 10.1056/NEJMs2512695. *N Engl J Med*. 2025;393(23):2365–7. Epub 2025 Nov 12 [PubMed: 41223362]
50. FDA Announces Plan to Phase Out Animal Testing Requirement for Monoclonal Antibodies and Other Drugs. <https://www.fda.gov/newsevents/press-announcements/fda-announces-plan-phase-out-animal-testing-requirementmonoclonal-antibodies-and-other-drugs>
51. NIH to prioritize human-based research technologies. <https://www.nih.gov/news-events/news-releases/nih-prioritize-human-based-research-technologies>
52. Hartung T. The (misleading) role of animal models in drug development. *Front Drug Discov*. 2024;4:1355044. 10.3389/fddsv.2024.1355044.
53. Kaceti J, Marešová P, Maskuriy R, Selamat A. Ethical questions linked to rare diseases and orphan Drugs - A systematic review. *Risk Manag Healthc Policy*. 2020;13:2125–48. [PubMed: 33116992]

54. Gerasimavicius L, Livesey BJ, Marsh JA. Loss-of-function, gain-of-function and dominant-negative mutations have profoundly different effects on protein structure. *Nat Commun.* 2022;13(1):3895. [PubMed: 35794153]
55. Liao Q, Zhang Y, Chu Y, et al. Application of artificial intelligence in Drug-target interactions prediction: A review. *Npj Biomed Innov.* 2025;2:1.
56. Sudhahar S, Ozer B, Chang J, et al. An experimentally validated approach to automated biological evidence generation in drug discovery using knowledge graphs. *Nat Commun.* 2024;15:5703. [PubMed: 38977662]
57. Wu Y, Ma L, Li X, et al. The role of artificial intelligence in drug screening, drug design, and clinical trials. *Front Pharmacol.* 2024;15:1459954. [PubMed: 39679365]
58. Mouchlis VD, Afantitis A, Serra A, et al. Advances in de Novo drug design: from conventional to machine learning methods. *Int J Mol Sci.* 2021;22(4):1676. [PubMed: 33562347]
59. Lakhan SE. The agentic era: why biopharma must embrace artificial intelligence that Acts, not just informs. *Cureus.* 2025;17(5):e83390. [PubMed: 40322603]
60. Huang K, Chandak P, Wang Q, Havaladar S, Vaid A, Leskovec J, Nadkarni GN, Glicksberg BS, Gehlenborg N, Zitnik M. A foundation model for clinician-centered drug repurposing. *Nat Med.* 2024;30(12):3601–13. 10.1038/s41591-024-03233-x. Epub 2024 Sep 25. [PubMed: 39322717]
61. Gangwal A, Lavecchia A. Unleashing the power of generative AI in drug discovery. *Drug Discov Today.* 2024;29(6):103992. 10.1016/j.drudis.2024.103992. Epub 2024 Apr 23. [PubMed: 38663579]
62. Thangamalai MS, Desai D, Selvaraj C. Revolutionizing structural biology: AI-driven protein structure prediction from alphafold to next-generation innovations. *Adv Protein Chem Struct Biol.* 2025;147:1–19. [PubMed: 40973394]
63. Ambreen S, Umar M, Noor A, et al. Advanced AI and ML frameworks for transforming drug discovery and optimization: with innovative insights in polypharmacology, drug repurposing, combination therapy and nanomedicine. *Eur J Med Chem.* 2025;284:117164. [PubMed: 39721292]
64. Luukkonen S, van den Maagdenberg HW, Emmerich MTM, et al. Artificial intelligence in multi-objective drug design. *Curr Opin Struct Biol.* 2023;79:102537. [PubMed: 36774727]
65. Fu L, Jia G, Liu Z, et al. The applications and advances of artificial intelligence in drug regulation: A global perspective. *Acta Pharm Sin B.* 2025;15(1):1–14. [PubMed: 40041903]
66. Pammi M, Shah PS, Yang LK, Hagan J, Aghaeepour N, Neu J. Digital twins, synthetic patient data, and in-silico trials: can they empower paediatric clinical trials? *Lancet Digit Health.* 2025;7(5):100851. 10.1016/j.landig.2025.01.007. Epub 2025 May 13. [PubMed: 40360351]
67. Barrett D, Cannon PM, Mingozzi F, Porteus M, Rivière I, Flotte TR. Overcoming barriers to commercially pre-viable gene and cell therapies for rare and ultra-rare diseases. *Mol Ther.* 2025;33(11):5316–26. 10.1016/j.ymthe.2025.09.049. Epub 2025 Oct 4. [PubMed: 41047628]
68. Fox TA, Booth C. Improving access to gene therapy for rare diseases. *Dis Model Mech.* 2024;17(6):dmm050623. 10.1242/dmm.050623. Epub 2024 Apr 19. [PubMed: 38639083]
69. Seifirad S, Haghpahan V. Inappropriate modeling of chronic and complex disorders: how to reconsider the approach in the context of predictive, preventive and personalized medicine, and translational medicine. *EPMA J.* 2019;10(3):195–209. [PubMed: 31462938]
70. Zhu Z, Cheng Y, Liu X, et al. Advances in the development and application of human organoids: Techniques, applications, and future perspectives. *Cell Transpl.* 2025;34:9636897241303272.
71. Huanbutta K, Burapapadh K, Kraisit P, et al. Artificial intelligence-driven pharmaceutical industry: A paradigm shift in drug discovery, formulation development, manufacturing, quality control, and post-market surveillance. *Eur J Pharm Sci.* 2024;203:106938. [PubMed: 39419129]
72. Jonker AH, Tataru EA, Graessner H, et al. The state-of-the-art of N-of-1 therapies and the IRDiRC N-of-1 development roadmap. *Nat Rev Drug Discov.* 2025;24(1):40–56. [PubMed: 39496921]
73. Koleva-Kolarova R, Buchanan J, Vellekoop H, Huygens S, Versteegh M, Mölken MR, Szilberhorn L, Zelei T, Nagy B, Wordsworth S, Tsiachristas A, HEcoPerMed Consortium. Financing and reimbursement models for personalised medicine: A systematic review to identify current models and future options. *Appl Health Econ Health Policy.* 2022;20(4):501–24. 10.1007/s40258-021-00714-9. Epub 2022 Apr 4. [PubMed: 35368231]

74. Chen X, Yi H, You M, Liu W, Wang L, Li H, Zhang X, Guo Y, Fan L, Chen G, Lao Q, Fu W, Li K, Li J. Enhancing diagnostic capability with multi-agents conversational large Language models. *NPJ Digit Med.* 2025;8(1):159. 10.1038/s41746-025-01550-0. [PubMed: 40082662]
75. Chen X, Jin Y, Mao X, Wang L, Zhang S. Ting Chen. RareAgents: Autonomous Multi-disciplinary Team for Rare Disease Diagnosis and Treatment. arXiv:2412.12475.
76. Zhao W, Wu C, Zhang YFX, Qiu P, Sun Y, Wang XZY, Zhang XSY. Yongguo Yu, Kun Sun, Weidi Xie. An agentic system for rare disease diagnosis with traceable reasoning. arXiv:2506.20430.
77. Kaiwen Zuo Z, Zhong P, Huang S, Tang Y, Chen Y Jiang, . HEAL-KGGen: a hierarchical multi-agent LLM framework with knowledge graph enhancement for genetic biomarker-based medical , diagnosis. bioRxiv 10.1101/2025.06.03.657521
78. Jonker AH, Tataru EA, Dimmock DP, et al. From roadmap to a sustainable end-to-end individualized therapy pathway. *Ther Adv Rare Dis.* 2025;6:26330040251339204. [PubMed: 40438698]

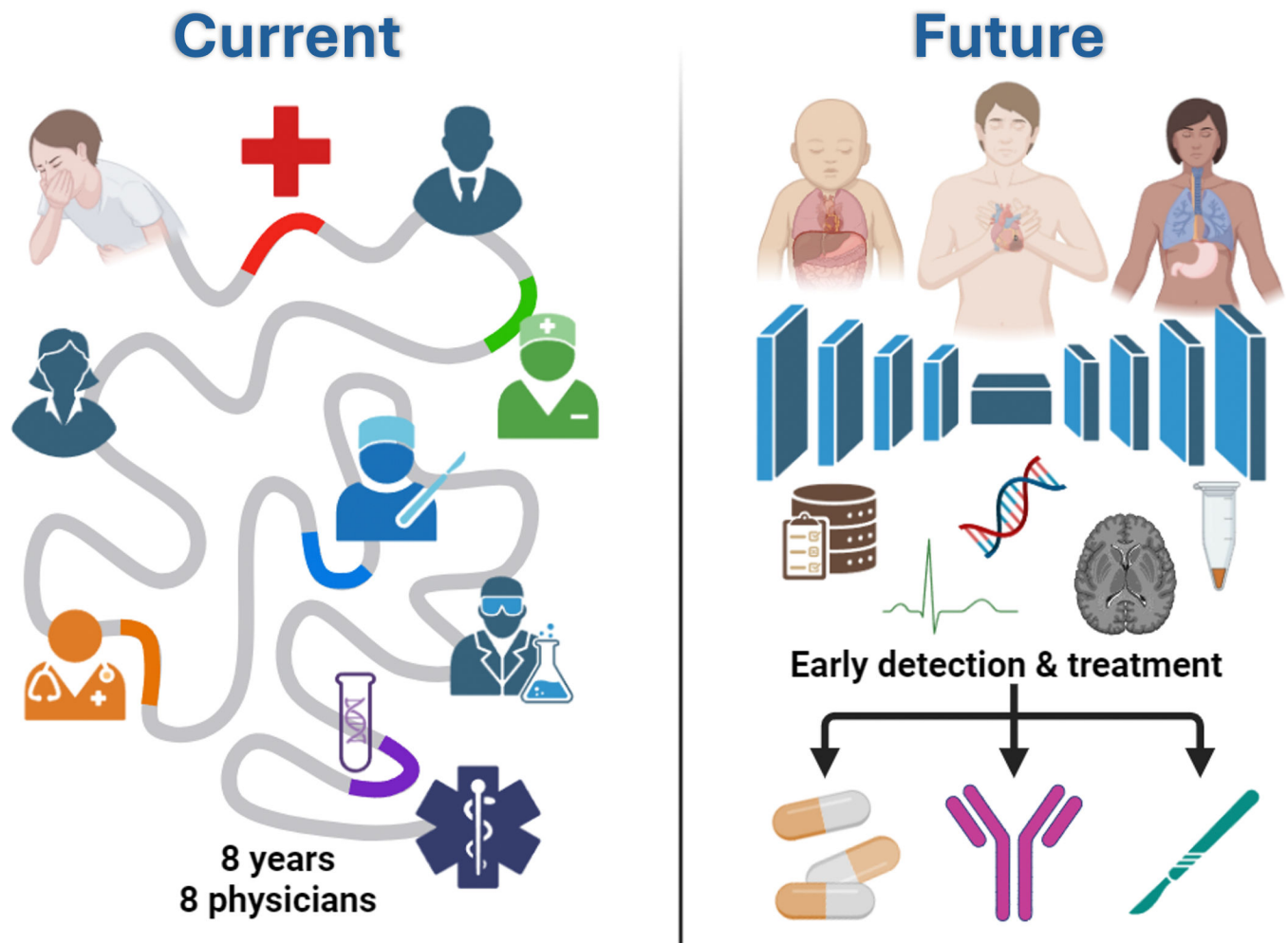


Fig. 1.

The diagnostic odyssey transformation. **Current:** Traditional pathway for rare diseases with multiple specialists, dead ends, and delayed diagnoses that often preclude effective treatment. **Future:** Proactive individualized AI screening of multimodal patient data, enabling early detection and initiation of appropriate therapies. Created in [BioRender.com](https://www.biorender.com)

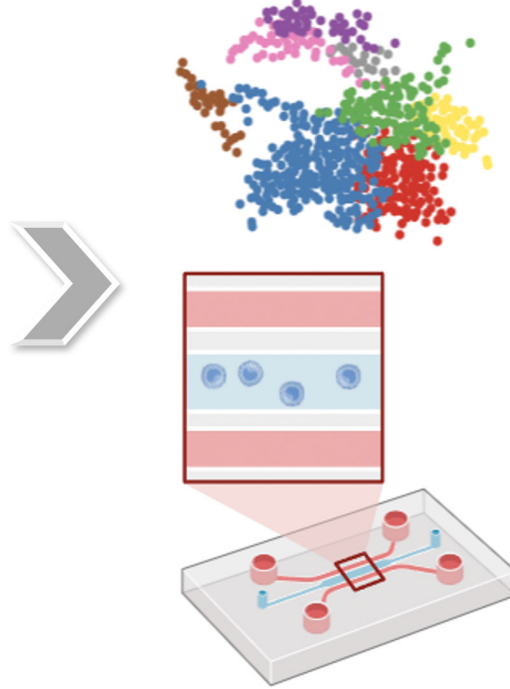
Gate 1

Clinical outcomes



Gate 2

Model predictions



Gate 3

Treatment feasibility



Fig. 2.

Decision gates for therapeutic development. **Gate 1** evaluates anticipated patient benefit based on clinically meaningful outcomes. **Gate 2** tests concordance with patient-derived in vitro or in silico models. **Gate 3** verifies treatment feasibility in context including dose, route, and interactions. Created in [BioRender.com](https://www.biorender.com)

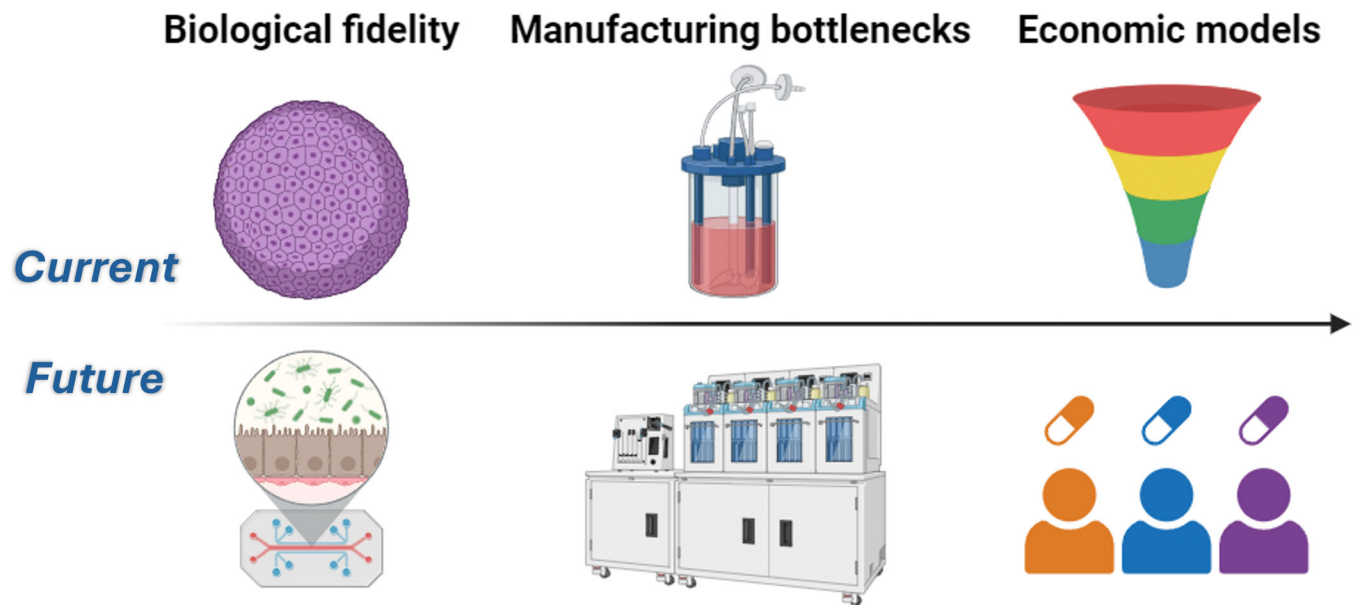


Fig. 3.

Real-world chokepoints that can be addressed using AI. **Biological fidelity:** Laboratory organoids currently lack biological complexity, requiring improved multisystem modeling for real-world fidelity. **Manufacturing bottlenecks:** Facilities limited by clinical batch production can be accelerated using robotic automation with continuous supply chain management. **Economic models:** Reimbursement policies favor common medications with tiered formularies, whereas rare diseases necessitate commercialization of N-of-1 frameworks. Created in [BioRender.com](https://www.biorender.com)