The Journal of Prevention of Alzheimer's Disease 13 (2026) 100417

Contents lists available at ScienceDirect

The Journal of Prevention of Alzheimer's Disease

journal homepage: www.elsevier.com/locate/tjpad

Special Article

The evolution of Alzheimer's target identification: Towards a fusion of artificial and cellular intelligence

Gayle Wittenberg [a], Fiona Elwood [b], Andrea Houghton [c], Tommaso Mansi [d], Bart Smets [e], Simon Lovestone [f,*]

[a] Neuroscience R&D, Johnson & Johnson Innovative Medicine, Titusville, NJ, USA
[b] Neuroscience R&D, Johnson & Johnson Innovative Medicine, Cambridge, MA, USA
[c] Neuroscience R&D, Johnson & Johnson Innovative Medicine, Spring House, PA, USA
[d] Data Science & Digital Health R&D, Johnson & Johnson Innovative Medicine, Titusville, NJ, USA
[e] Data Science & Digital Health R&D, Johnson & Johnson Innovative Medicine, Beerse, Belgium
[f] Neuroscience R&D, Johnson & Johnson Innovative Medicine, London, United Kingdom

ARTICLE INFO

Keywords:
Target identification
Artificial intelligence
Big Data
iPSCs

ABSTRACT

Decades of advances unfolding in parallel across diverse domains have delivered to science rapid rises in the scale of multiplexing, population-level cohort sizes, global computational capacity, massive-scale artificial intelligence (AI) models, and advanced human cellular modeling capabilities. These have generated unprecedented volumes of data, allowing researchers to explore Alzheimer's disease (AD) biology at a depth and scale never before possible. The explosion of multi-omics datasets and computational power heralds an era in which the complexity of AD can be meaningfully dissected and reconstructed leveraging AI. These can be applied to advance our understanding of the root causes of disease, fundamentally a forward problem, tracing how dysfunction emergence from interactions across genes, cells and environments over time. On the other hand, therapeutic discovery requires addressing the inverse problem, working back from the diseased state to pinpoint upstream interventions that restore health. Human induced pluripotent stem cells (iPSCs) and other human cell models play a pivotal role in this process, naturally computing the mapping from perturbation to phenotype at scale. By recreating human-relevant biology, this cellular intelligence enables validation of targets predicted by AI and testing of interventions that drive therapeutic progress. We look to the next horizon in Alzheimer's research as a collaboration, a convergence of three forms of intelligence: human, artificial and cellular. In unison, these complementary forces will shape a new frontier for AD research where scientific innovation and human ingenuity work together bringing hope for meaningful advances and new therapies.

1. Introduction

Biological systems are complex systems. Disease represents maladaptive perturbation to this system arising from genetic or environmental hits over time. Therapeutics seek to restore function by strategically modulating, typically, one or a small number of biological targets. Target identification is the nomination of such targets for therapeutic intervention; target validation is confirmation, through orthogonal approaches, that perturbation of the target restores the desired function. Together, a case could be made for target identification and target validation (TiTv) being the most critical hurdle in drug discovery.

Understanding the cause of disease, and how it emerges, is a *forward problem* - with an aim to understand the emergence of dysfunction from genetic variants or environmental exposures across the complexity of the 30 trillion cells in the human body, 20,000 genes, mRNAs, proteins, miRNAs, epigenetics, post-translational modifications, autoantibodies, and pathogen exposures. Target identification, by contrast, is best framed as an *inverse problem*: starting from an undesired system state - disease - we identify an upstream intervention, often involving one protein or pathway, that will restore healthy function (Fig. 1). Target identification demands not only biological insight but also a shift in perspective. This process aims to pinpoint molecular entities whose modulation can alter disease progression, providing a rational basis for

* Corresponding author.
E-mail address: SLovesto@its.jnj.com (S. Lovestone).

https://doi.org/10.1016/j.tjpad.2025.100417
Received 5 September 2025; Received in revised form 14 October 2025; Accepted 20 October 2025
Available online 1 December 2025

therapeutic intervention.

Historically, target identification relied on reductionism: isolating pathways, studying disease models, and gradually triangulating typically on a protein or gene believed to be central to pathology. Validation of a target could involve years of experimental work leveraging genetic knockouts, tool compounds, and animal models built around a carefully constructed mechanistic hypothesis. This approach has led to many notable successes in drug discovery and historically has been the driver of most therapeutic programs for Alzheimer's disease (AD) in the clinic and in clinical development today.

The emergence of big data and AI is enabling a new paradigm. With genome-wide association studies, whole exome and genome sequencing, and multi-omics data sets including emerging vast proteomics cohorts, the data landscape available for learning has exploded. This data explosion coupled with recent advances in artificial intelligence, such as generative AI, makes it possible to explore disease biology at unprecedented scale, depth and speed. These massive datasets and AI tools have broadened our view, enabling us to identify new targets, cluster patients by molecular subtype or project along different mechanistic dimensions of disease and uncover latent patterns driving progression.

In the face of optimism about these advances in technology, data, and AI, we must remember that these are tools to be applied judiciously. As our methods grow more powerful and our data more complex, it becomes even more critical to sharpen the questions we ask. Big data and AI are often used to search for biological causes of disease. It is easily assumed, without careful framing, that a causal gene will be a viable target. Dennis Noble reminds us that the concept of 'relativity' applies to biology, not just to physics [1]: in the reference frame of the drug developer, we are not searching for cause, but for cure – even if at times these may align. Below we outline the journey of TiTv in Alzheimer's disease, from empirical biochemistry, molecular and cellular biology through Big Data analytics to the potential of AI. We conclude that the potential of AI is enormous, but the role of the scientist remains paramount

2. TiTv before the age of 'big data'

Three key questions lie at the heart of selecting a novel target to treat disease: (1) which molecular target to modulate, (2) which therapeutic modality can be developed that best delivers the desired modulation of that target, and (3) which patients will benefit? While all three questions are important, the success of any discovery campaign relies on defining a strong link between a proposed target and disease. Before the advances in high-throughput omics and computational analytics, target identification relied primarily on a combination of human pathology, biochemical pathway elucidation, animal models, and genetic insights from rare familial disorders. These formed the bedrock of many successful drug development efforts.

Discovery and development of therapies typically have an increased chance of success when the biology translates well between preclinical species and man as it enables the drug discovery campaign significantly. In Alzheimer's disease (AD), translation between preclinical species and man is particularly challenging, given preclinical species do not succumb to the same pathophysiology. Therefore, it is unsurprising the earliest approved treatments modulated pharmacology known to be perturbed in disease: namely, acetylcholinesterase (AChE) inhibitors such as donepezil [2]. These treatments emerged from the observations of cholinergic deficits in the brains of patients with AD. Postmortem studies in the 1970s and 80 s revealed degeneration of cholinergic neurons in the basal forebrain, leading researchers to test therapeutics that could enhance acetylcholine signaling (see [3]). Furthermore, modulating acetyl choline transmission in rodents can improve cognition, building confidence in the approach, and enabling drug discovery programs. More specifically, the preclinical development of donepezil relied on *in vitro* assays - a rat brain homogenate assay of inhibition of AChE, a rat plasma assay of butyl cholinesterase – and *in vivo* assays of target modulation i.e. demonstration that donepezil inhibition of AChE in aged rat brains resulted in improved learning in rats. Much of this work also provided the basis for clinical development as the assays were translatable. For example, dose selection during clinical development was based on AChE inhibition in red blood cell membranes and plasma [4]. These treatments were successfully developed without molecular biomarkers or genetic stratification. This was possible because clinical trial duration for a symptomatic individual is relatively short (12 weeks), a sufficient percentage of the population responds to treatment, adverse events are monitorable, and patients with dementia of various

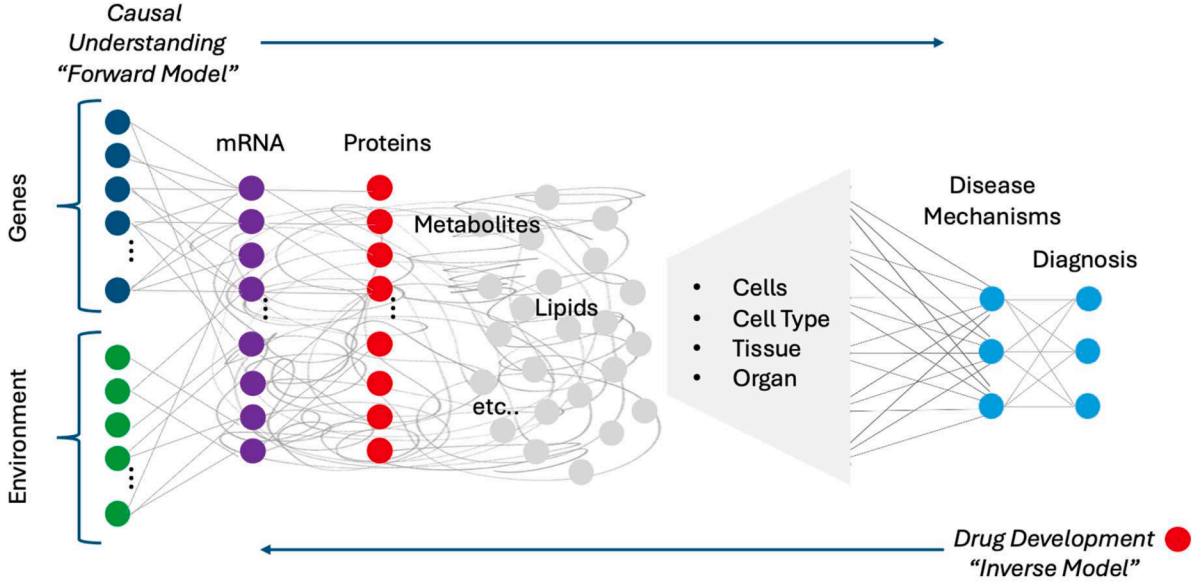

Fig. 1. Conceptual model of disease and treatment. Causal understanding of disease consists of tracing the biological cascade from genes and environmental inputs through the successive molecular, cellular and tissue-level networks towards disease mechanisms, and eventually clinical manifestations (e.g., ICD codes). This "forward model" reflects causal biological understanding – how perturbations in genes or environment drive disease. Drug development typically begins with the clinical labeling of disease and works backward, seeking to identify a target, often a protein, to manipulate to reverse or modulate disease associated signatures and disease.

etiologies respond to treatment with anti-AChE agents.

However, to move beyond symptomatic treatments, an understanding of disease pathology and underlying pathophysiology is required. The discovery that amyloid beta (Aß) was an integral part of plaques in postmortem brains from patients with AD in the mid 1980s [5] combined with the discovery that mutations in amyloid precursor protein (APP), presenilin 1 (PSEN1) and PSEN2 genes were associated with early onset familial Alzheimer's Disease (FAD), enabled scientists to focus on disease pathology in AD drug discovery programs. Importantly for drug discovery, knowledge of these genetic mutations enabled the development of preclinical models such as the Tg2576 mouse. This model has a Swedish FAD mutation (K670N/M671Lz), and the expression of the human APP is five-fold above the levels of endogenous APP; additionally, the expression of Aß1–40 and Aß1–42, and amyloid deposition increases with age, along with gliosis and dystrophic neuritis. Amyloid plaques appear in mice between 11 and 13 months of age. Furthermore, these mice also display spatial memory impairment by 9–10 months of age [6]. Having such models fueled investment in treatments for AD as there was viable platform to profile candidate molecules.

These models successfully recapitulated some pathological features but often failed to predict clinical efficacy in humans, contributing to the high attrition rate in Alzheimer's trials. For example, ß secretase (BACE) inhibitors were predicted to be efficacious based on transgenic mouse model data but failed in the clinic even though target modulation was demonstrated in preclinical species and in man (see Neuman et al., 2019 [7]). This is because these models do not recapitulate disease; rather, they model but one aspect of the biology. For AD research, most *in vivo* models are a model of amyloid pathology or a model of tau pathology. It is still unclear how to develop a mouse model in which these pathologies co-exist in a manner that more closely recapitulates the human disease state.

Thus, in recent years target identification and validation for drug discovery in AD has pivoted to a greater reliance on human data collected from living patients at different stages of disease. This has informed our understanding of disease evolution, led to new hypotheses to test (e.g. tau seeding) and provided data which we can exploit as we endeavor to build more human relevant models. Induced pluripotent stem cell (iPSC) technology has enabled the creation of human-based models for Alzheimer's disease (AD) drug discovery. These models, derived from patient cells, allow researchers to study disease mechanisms and test potential drug candidates in a human context, complementing traditional animal models. Human-induced pluripotent stem cell (hiPSC) models range substantially in complexity, from two-dimensional monolayers to three-dimensional organoids and from single purified cell types (e.g. neurons) to complex cultures with a mixture of cell types (e.g. neurons, microglia and astrocytes). Neurons can be derived from iPSCs of healthy control or AD-patient populations, and gene-editing tools such as CRISPR/Cas9 can be used to generate isogenic cell lines for the study of genetic variants associated with AD [8–10].

Researchers can leverage panels of hiPSCs derived from genetically diverse donors to study how genetic variants influence molecular and cellular phenotypes in multiple genetic backgrounds. These cellular-level association studies (e.g., eQTL or chromatin QTL mapping) complement large scale GWAS by helping to identify which variants have functional consequences in human cells. Epigenetic profiles can be generated for different cell types with single-cell resolution (e.g. single-cell ATAC-seq [11]). Sequencing can identify genes of relatively small effect size, provided the pools of *in vitro* samples are large enough. Emerging optical tools may further support pooled genetic screens with imaging and morphology-based phenotypes. When combined with CRISPR-based perturbation screens, these approaches can identify modifiers of disease-relevant cellular phenotypes in scalable target discovery efforts, while also providing compelling evidence to validate new targets in human cells.

Often, human biology used to identify and validate targets as well as enable drug discovery efforts can also be used as a translatable endpoint for clinical development. For example, CSF and plasma measures of phosphorylated tau have been used to monitor target engagement in anti-tau trials. Amyloid PET imaging has become a *de facto* standard for validating anti-amyloid therapies. Still, the disconnect between biomarker improvement and clinical benefit remains a central challenge in Alzheimer's drug development, underscoring the need for early validation of mechanistic links between target engagement and disease modification.

3. TiTv in the era of 'big data'

The advent of large-scale molecular datasets from human samples has transformed the landscape of target identification and validation (Fig. 2). Rather than relying largely on pre-defined hypotheses generated through painstaking empirical studies as described above, researchers can now interrogate high-dimensional data from genetics, transcriptomics, epigenomics, proteomics and other modalities to uncover novel putative targets. These data support a fundamentally different mode of discovery that is network-informed, context-sensitive, and population-aware, rather than linear or pathway-bound. It is also more successful. A target supported by genetic data is over twice as likely to yield a successful drug discovery programme, and a target supported by single cell sequencing expression data is similarly a better bet for success [12,13].

Nowhere has this shift been more profound than in neuroscience, where decades of research often failed to translate due to poor models and limited access to human brain tissue. Genome-wide association studies (GWAS) involving hundreds of thousands of individuals have now identified robust, replicable genetic risk loci across a range of neuropsychiatric and neurodegenerative disorders. Large-scale GWAS studies of late onset Alzheimer's disease have identified variants associated with altered risk near APOE, BIN1, CLU, and PICALM genes [14–16]. Beyond confirming the central role of amyloid processing, these studies highlighted the role of specific biological processes, such as neuroinflammation, lipid biosynthesis, and endocytosis in AD.

A primary challenge with GWAS data is translating statistical associations into a mechanistic understanding of disease and, ultimately, into viable drug targets. Many disease-associated single nucleotide polymorphisms (SNPs) reside in non-coding regions of the genome, making their functional consequences difficult to decipher [17,18].

To bridge this gap, GWAS data are being integrated with other omics layers to prioritize causal genes and pathways. A key strategy is the use of expression and protein quantitative trait loci (eQTLs and pQTLs), which link genetic variants to changes in gene expression and protein abundance, respectively. By co-localizing a GWAS risk signal with an eQTL or pQTL, researchers can form a hypothesis that the genetic variant influences disease risk by altering the expression of a specific gene or protein [19–21].

Statistical methods such as Mendelian randomization (MR) allow for more formal tests of causality. MR uses genetic variants as unconfounded proxies for an exposure (e.g., the level of a specific protein) to infer its causal effect on a disease outcome. Other powerful approaches include transcriptome-wide and proteome-wide association studies (TWAS/PWAS), which integrate GWAS summary statistics with gene expression or protein level data to identify genes whose expression levels are associated with disease risk.

The integration of eQTLs and pQTLs, derived from large-scale multi-tissue transcriptomics and proteomics datasets, with GWAS data using these approaches led to the prioritization of many genes and proteins causally implicated in AD and other neurological disorders [22–27]. These analyses are being further revolutionized by single-cell technologies, which allow for the investigation of the cell-type-specific effects of disease-associated variants [28].

These multimodal analyses are important because although the genome provides a static blueprint of inherited disease risk, it does not capture the dynamic changes that occur as a disease develops and

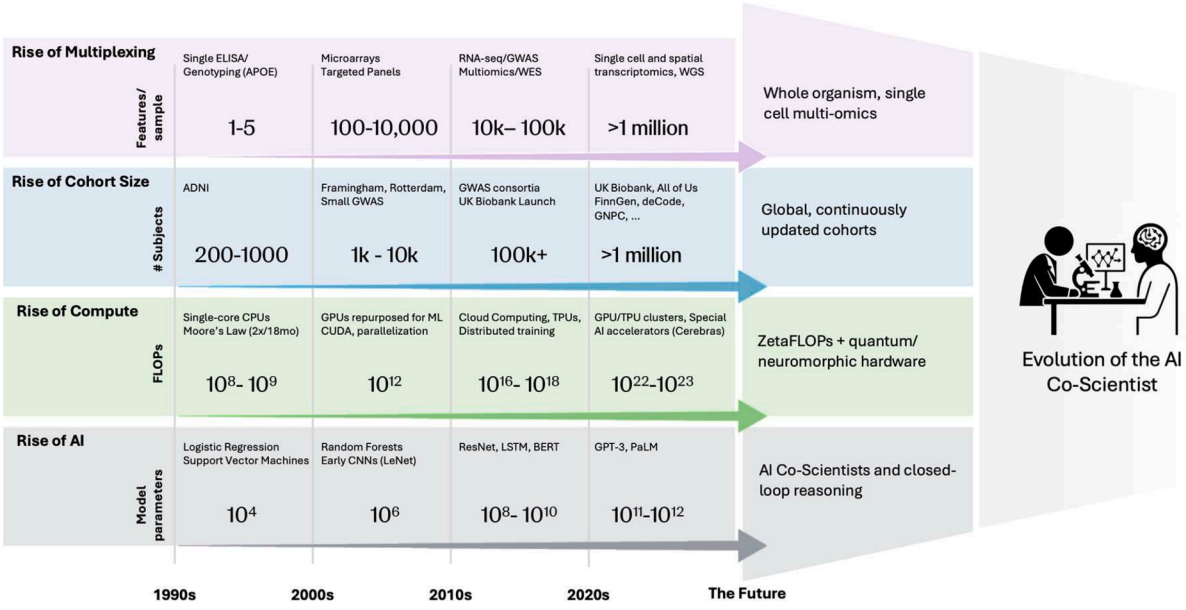

Fig. 2. Trends Enabling AI-Driven TiTv and the Evolution of the AI Co-Scientist. This figure illustrates four parallel trends that are transforming precision neuroscience and biomedical discovery. (1) The rise of multiplexing (purple): from single-analyte assays to massively multiplexed, spatially resolved single-cell multi-omics. (2) The rise of cohort size (blue): from hundreds of individuals in early disease studies to global biobanks with >1 million participants, moving towards continuously updated, real-world integrated cohorts bringing together medical records, clinical, biomarker and digital data. (3) The rise of compute (green): from single-core CPUs and Moore's Law to cloud-scale GPU/TPU clusters and specialized accelerators, with future prospects in zetaFLOPs and neuromorphic or quantum computing. (4) The rise of AI (grey): from classical machine learning to deep learning and foundational models, advancing towards agentic AI systems capable of reasoning, planning, and closed-loop experimentation. Together, these converging trajectories point towards the emergence and evolution of the AI Co-Scientist, partnering with human researchers to accelerate targeting discovery.

progresses. Transcriptomics, proteomics, metabolomics, and epigenomics offer a real-time snapshot of the molecular state of a cell or tissue, reflecting the integrated output of genetic risk, environmental influences and co-morbidities. Large-scale, spatiotemporal analyses of bulk brain transcriptomics data, utilizing resources from consortia such as the Accelerating Medicines Partnership-Alzheimer's Disease (AMP-AD), have facilitated the identification of co-expressed gene modules linked to AD pathology and cognitive decline [29,30]. These studies have uncovered both brain-region specific expression changes and co-ordinated alterations across different areas of the brain. While it is challenging to distinguish cause from consequence in transcriptomics data, the enrichment of top modules for AD genetic risk factors and the observation that many transcriptional changes occur early in the disease highlights potential pathways driving disease progression and points to possible targets for therapeutic intervention. Notably, these transcriptional modules are enriched for genes involved in processes such as nervous system development, axon growth, inflammation, and proteostasis.

A further leap in understanding has come from single-cell sequencing. Traditional bulk tissue analysis averages the molecular signals from millions of different cells, obscuring the contributions of rare cell types or specific cell states. Single-cell RNA sequencing (scRNA-seq) of post-mortem human brain tissue has enabled the deconvolution of this complexity, leading to the identification of specific disease-associated cell states, such as activated microglia (DAM), astrocytes (DAA), and inhibitory neuronal subtypes associated with resilience to AD pathology [31–34]. This level of resolution is critical for mapping genetic risk variants to the specific cell types in which they exert their effects and provides highly specific hypotheses for therapeutic intervention.

Recent advances in large-scale proteomics now bring the systems biology perspective of big data to the protein level, where most drug targets reside. High-throughput mass spectrometry and affinity-based platforms enable large-scale profiling of protein expression, post-translational modification and protein-protein interactions across tissues, cell types and disease states. In AD, proteomic analyses across multiple brain regions have uncovered disease-associated changes in protein co-expression networks, some of which were not observed at the RNA level, including a MAPK signaling module associated with cognition and a extracellular matrix proteins that showed a positive correlation with plaques and tangles [35–37].

The latter discovery underscores the complementarity of the different data layers and the critical importance of integrating multiple omics modalities to get the full picture of biology. By combining genomics, epigenomics, transcriptomics, proteomics and metabolomics from the same individuals, researchers are building more complete models of disease [38–40]. This will allow for the construction of a chain of evidence from a genetic risk variant to its functional consequence on the epigenome, transcriptome, proteome, and metabolome, providing a much richer understanding of disease mechanisms and a more solid foundation for target identification.

The immense scale of data required for robust omics-based target discovery necessitates a collaborative, open-science approach. Large consortia and public-private partnerships have been essential in generating and harmonizing the necessary multimodal datasets. Initiatives like the Accelerating Medicines Partnership for Alzheimer's Disease (AMP-AD), the Global Neurodegeneration Proteomics consortium [41], UK Biobank and FinnGen are creating invaluable resources that integrate deep clinical data with multi-omics profiles from thousands of participants. By making these data broadly available, these consortia are accelerating the pace of discovery and empowering researchers globally to identify and validate the next generation of therapeutic targets for Alzheimer's disease (Fig. 2).

Beyond identification of biologically effective targets, big data combined with advanced methods including AI is advancing the early identification of target-specific safety concerns. Human genetics and multi-omics integration now enable systematic prediction of on-target safety liabilities based on gene constraint metrics and natural loss-of-

function variation [42,43]. Complementary transcriptomic and single-cell atlases allow tissue-specific expression mapping to anticipate unintended effects in critical organs [44,45]. Moreover, AI-driven systems toxicology approaches are beginning to model cross-pathway perturbations and highlight mechanisms of target-associated risk earlier in discovery [46]. Collectively, modern data science pipelines are reshaping how safety evaluation is incorporated into target prioritization.

Following the identification and validation of a target, opportunities need to be prioritized based on druggability and the latest understanding of the therapeutic modalities available to the scientist. Quantitative assessments of tractability now integrate structural features, ligandability scores, and prior success across protein families [47,48]. Different modalities will inherently bring different benefits and liabilities, and these need to be matched to the profile of the target and needs of the patient [49,50]. An integrated view of the end-to-end drug discovery and development process, in addition to an intimate understanding of the patient's experience, can lead to improved decision-making even at the earliest stages of target selection.

Altogether, this era of big data has not only enhanced the precision of target discovery but also expanded its scope, bringing molecular insights with disease phenotype, stages and therapeutic response to drive forward more personalized and effective drug development strategies.

4. Analysis before the age of AI

Before big data entered the scene, AI remained underdeveloped. Data analysis in target discovery was primarily rooted in classical statistical models, expert intuition and manual exploration. Linear regression, logistic models, survival analysis, ANOVA and PCA formed the core toolkit. In Alzheimer's disease, early biomarker studies applied these methods to investigate cerebrospinal fluid (CSF) levels of Aβ42, tau and phosphorylated tau. These methods prioritized transparency, interpretability, and hypothesis-driven reasoning; this produced many durable insights.

Researchers manually explored gene lists, drew networks, annotated pathways, and traced findings back to curated knowledge using tools like Gene Ontologies [51,52], KEGG [53], and Ingenuity Pathway Analysis [54]. These approaches lent interpretability and structure to transcriptomic studies in AD, for example, by highlighting disruptions in mitochondrial function, immune signaling or synaptic plasticity in postmortem brain tissue.

The rise of omics datasets exposed the limitations of these approaches. One of the most pervasive challenges was the 'curse of dimensionality': the number of features (e.g. genes, transcripts, proteins) far exceeded the number of samples, creating sparsity and instability in traditional statistical frameworks. This was especially stark in AD, where precious human biosamples were limited in availability and heterogeneity across individuals was high. In high-dimensional, low sample-size (HDLSS) regimes, traditional statistics became fragile – models overfit, false positives abounded, and biologically relevant signals could be easily missed [55].

To cope with complexity, analysts extended classical approaches using network models, weighted gene co-expression networks (WGCNA) [56], Bayesian frameworks [57], and dimensionality reduction techniques [58]. These methods edged closer to machine learning but remained grounded in static, human-interpretable paradigms. In AD, such methods helped uncover co-regulated gene modules associated with neuroinflammation and amyloid pathology. While useful, these approaches struggled to Nscale across conditions, datasets and modalities, often failing to capture nonlinear relationships, context-specific interactions or multivariate patterns that characterize neurodegenerative disease processes. This was particularly true in transcriptomics and proteomics, where multiple testing correction and arbitrary thresholds could obscure emergent biological signals.

Compounding these challenges was the heterogeneity of the data itself. Omics datasets are generated from a wide array of platforms and protocols, each with its own formats, distributions and artifacts. Batch effects, systemic, non-biological variations introduced during sample processing or sequencing, frequently masqueraded as biological signal. A lack of standardized pipelines and shared frameworks made it difficult to combine insights across studies or derive generalizable biological understanding.

Even when technical hurdles were overcome, interpretation often stopped short of biological meaning. Observational omics data could highlight correlations but rarely offered a clear path to causality or therapeutic intervention. The leap from statistical signal to functional relevance required time-intensive experimental validation; many promising findings failed to replicate. As datasets grew larger and more complex, the ability to reason through them manually, or to visualize and explore them directly reached a point beyond the limit of human intelligence, The sheer volume, diversity and complexity of modern datasets pushed conventional methods to their limits. Analyses remained constrained by what we knew how to model, which introduced biases and limited what we could discover. Human reasoning, while essential for grounding interpretation, could not keep pace with the full dimensionality of the data. These challenges set the stage for a transformation – a transition into the age of AI.

5. Analysis in the era of AI

Artificial intelligence (AI) has been undergoing a dramatic acceleration since the "revival" of neural networks and deep learning in the early 2010′s, powered by the exponential growth of data, through internet and compute through the use of GPU for AI (Fig. 2). Launched in the fields of vision, language and audio, AI approaches were quickly translated to the fields of biology and physics, making AI a foundational force in biomedical research. This rise has been propelled by several key enablers: the proliferation of high-throughput, high-dimensional biological data sets described above, advances in machine learning architectures [59], modern training approaches (like self-supervision), and unprecedented access to advanced hardware like GPUs or TPUs through cloud providers. Together, these advances have unlocked the potential to analyze not only large volumes of unstructured data, but also highly heterogeneous and noisy data at scale.

One of the most important paradigm-shifts in AI is the advent of self-supervised learning and resulting foundation models in the late 2010′s [60]. Traditionally, machine learning systems are trained from scratch from well-curated, small databases. Methods were tailored to extract insights from these isolated databases and studies or combined through meta-analyses. Self-supervised learning changed the approach [61]. Enabled by the proliferation of data, deep neural networks are pre-trained on a very large amount of unlabeled, heterogeneous datasets, leading to a so-called foundation model (FM). Intuitively speaking, the models are trained to recover perturbations made to the original data, enabling training at scale. Foundation models are then used as back-bones to more specific, task focused predictive models [62] or generative AI models [63]. By design, a FM thrives on heterogeneity, pulling strength from patterns distributed across vast, diverse corpora. Moreover, FMs enable principled ways of integrating multiple modalities [64], like text and images, learning across modalities and across studies.

For complex diseases like Alzheimer's, which defy simple causal models and often suffer from fragmented evidence across data types and cohorts, this shift in the power of AI, coupled with the release of large cohort data, holds transformational potential. For example, foundational models trained on single-cell transcriptomic data across development and disease stages have revealed cell-state transition in microglia, astrocytes, and neurons that mark early divergence from healthy aging [65,66]. These insights, elusive in smaller datasets, suggest new axes of stratification and windows for therapeutic intervention.

FM are able to integrate data across omics modalities and generate

latent representations of biological structure that have the potential to capture the relationships between genes, proteins, cell types, pathways and disease states [67,68]. These methods enable not only higher accuracy, but a different kind of *reasoning* that is less dependent on predefined hypotheses and more able to surface unexpected emergent patterns. For instance, genetic variation, transcript abundance, protein levels, post-translational modifications, and epigenetic markers can be analyzed jointly, creating a layered understanding of disease, in a hypothesis-generating setting and purely driven by data. This multi-modal integration has the potential to enable mechanistic tracing from inherited risk alleles through cellular dysfunction to clinical outcomes, and can illuminate synergistic or antagonistic interactions that would remain invisible if each data type were analyzed in isolation.

When foundation models are used in their generative form, they add an additional capability: the ability to simulate new data instances that reflect learned distribution from the training datasets. In practice, this means AI can generate synthetic omics profiles within the manifold of the disease states observed in the training data, potentially offering new hypotheses about disease progression or subtypes, hypotheses that would still need to be verified experimentally of course. In Alzheimer's research this capacity is particularly valuable in the preclinical space, where early disease signatures are subtle and underpowered in any single dataset. Generative models can also propose novel peptides or molecules optimized for binding to pathologically relevant targets, including tau aggregates or synaptic receptors, accelerating iterative loop of design, synthesis and testing.

It seems likely that these advances in AI will increasingly dominate in our analysis of very large, and very complex datasets such as the multilayered bio-ome, for insights leading to novel TiTv. However, the form of AI that has so stunned the world in the last few years has been large language models (LLMs), and there are many ways in which this form of generative AI will impact on TiTv. Recent developments in LLM architectures, particularly those incorporating test-time compute strategies and reinforcement learning fine-tuning methodologies [69] have demonstrated enhanced reasoning capabilities in what are now characterized as large reasoning models. Emerging systems integrate tools or models together with retrieval-augmented generation frameworks; these systems exhibit capacity for executing complex tasks. Such approaches, often referred to as "AI co-scientists", are being explored to augment biological analysis and TiTv [70–72], with similar platforms designed to accelerate hypothesis generation (Fig. 2). These systems typically employ sophisticated orchestration mechanisms integrating literature retrieval, automated peer review protocols, and ranking systems, collectively contributing to expedited discovery workflows.

While the transformative potential of deep learning and LLMs is evident, their application to Alzheimer's disease and related neurodegenerative conditions remains constrained by data and domain limitations. Model performance scales with data volume, yet biological data sets are historically small, heterogeneous and context-dependent, limiting generalizability [73,74]. Moreover, phenomena such as "hallucinations" or spurious correlations can compromise interpretability and reliability when models are applied to clinical hypotheses [75,76]. Emerging Large Reasoning Models (LRMs) also exhibit inconsistency in producing accurate and coherent reasoning traces, with performance declining as task complexity increases [77,78]. Overcoming these challenges may require hybrid AI frameworks that integrate data-driven learning with causal and symbolic reasoning ground in biological and mechanistic knowledge [79,80].

The primary value proposition of these computational frameworks lies in their potential to not only accelerate discovery, but also to mitigate the cognitive biases and subjective preferences inherent in traditional scientific methodologies, while simultaneously identifying previously unconsidered research directions and providing more comprehensive analytical assessments. Concurrent developments in multi-modal agentic systems have enabled the integration of specialized models, such as large-scale single-cell analysis platforms, which can be dynamically activated by LLM orchestrators to execute targeted computational tasks with automated result interpretation and presentation. Finally, "vibe coding", where users describe what they want in natural language and the system automatically generates appropriate source code, is gaining traction in the world of computational biology, with the promise to bring complex data analysis to the fingertips of biologists, accelerating the hypothesis – experiment – analyze cycle for TiTv.

These tools, taken together, expand not just the scale but the scope of what can be discovered and the ease and accessibility of use. Ultimately, what distinguishes analysis in the age of AI is not only its power or precision, but its ability to tackle more and more complex data and scientific questions. Where traditional methods were optimized for curated datasets and well-defined questions, recent developments in AI are enabling more robust data quantification, accurate prediction, and more sophisticated hypothesis-generating experiments to discover novel biology (Fig. 3). Through the power of foundation models, AI allows weak signals across noisy datasets to reinforce one another. In doing so, it creates new opportunities for inference. This is particularly advantageous in diseases like AD where causality is diffuse, clinical manifestations are delayed, and success has long been hindered by the mismatch between mechanism and measurement.

6. Current limitations of AI and big data

Despite the rapid evolution and promise of AI in biomedical research, significant limitations remain. Most limitations are not rooted in the algorithms themselves, but in the data, systems, and human structures that surround them. These limitations are particularly salient in neuroscience and other complex therapeutic areas where disease heterogeneity, sparse signals and incomplete understanding of biology compound the inherent challenges.

One of the most commonly cited limitations is the quality and nature of the underlying data. Even with vast quantities of omic data, much of it is noisy, incomplete, inconsistently annotated, and riddled with batch effects. In such environments, AI systems are susceptible to the classic problem of 'garbage in, garbage out'. Quantity does not equal quality, and poor-quality data can mislead even the most sophisticated models. These issues are amplified when data are drawn from different platforms, time points or patient populations.

Adding to this challenge is the scarcity of data *for specific problems*. While the overall volume of biological data is increasing, well-annotated, disease-specific, or subgroup-specific datasets remain limited. This is especially true in fields like neurodegeneration, where relevant patient cohorts may be small, diverse and difficult to access – and where longitudinal data, paired with imaging and other biomarkers may be needed to facilitate interpretation, yet take years to generate. AI models trained on narrowly scoped or homogeneous data risk learning features that fail to generalize to broader populations, or worse, perpetuate biases embedded in the training data. This risk is acute when datasets underrepresent certain demographics, genotypes or disease stages raising both scientific and ethical concerns [81–83].

Even when data are sufficient in volume and quality, the models themselves may introduce their own limitations. Many of the most powerful AI systems, deep neural networks and large foundational models, operate as black boxes. They generate highly accurate predictions, but the logic behind those predictions often remains opaque. This lack of interpretability poses a challenge for clinicians and regulators in high-stakes domains like drug development where decisions must be both evidence-based and explainable [84]. Without understanding *why* a model has arrived at a given conclusion, it becomes difficult to validate or trust its output, particularly when it is used to justify experimental therapeutics or diagnostic decisions.

This limits biological insight. In drug discovery, it is often not enough to know that something works. We must understand why it works in order to improve it, avoid side effects and design next-generation

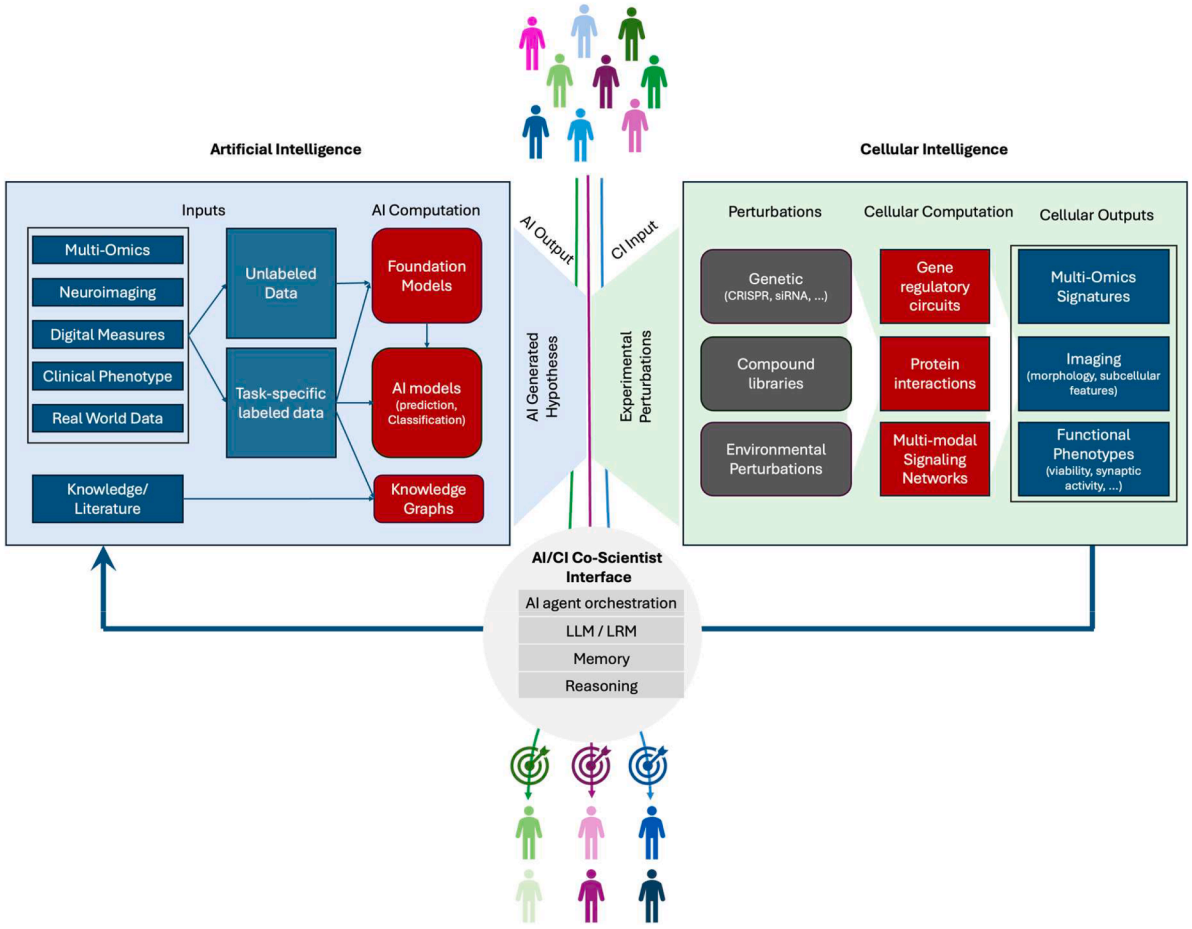

Fig. 3. Interplay of Artificial and Cellular Intelligence in target identification and validation. Artificial intelligence (left) integrates diverse inputs – including multi-omics, neuroimaging, digital and clinical measures, real-world data, and literature into broader foundation models, task-specific predictive models, and knowledge graphs. These models generate putative target hypotheses. These are encoded as perturbations (e.g. CRISPR edits, drug or environmental challenges). Cellular Intelligence (right) executes these perturbations through intrinsic biological computation (gene regulatory networks, protein interactions, and signaling networks) producing measurable outputs such as multi-omics signatures, morphological and imaging features, and functional phenotypes. These experimental results feed back into AI systems, refining the next cycle of hypothesis generation and refinement. Together this loop defines and AI-cell co-scientist paradigm, where artificial and cellular computation operate in tandem to accelerate discovery, and identify the patient population most likely to benefit from a novel intervention.

interventions. Models that obscure biological reasoning may succeed in pattern recognition but fail to advance mechanism-based understanding or rational design.

Reproducibility and generalizability further complicate the landscape. AI models can overfit to training data, particularly when data are sparse or biased, leading to performance drops on new or external datasets. Combined with the inherent variability of biological experiments, this contributes to the reproducibility crisis that has long plagued the biomedical sciences. Without shared benchmarks, rigorous standards and transparency around methods and results it is difficult to compare models or build on prior work.

Finally, technical and infrastructure-related challenges should not be underestimated. Integrating multimodal data across genomics, imaging, clinical records and other modalities remains a daunting task. Standards for data formatting, normalization and metadata are often inconsistent. Training advanced AI models requires significant computational resources: cloud computing, GPUs, and expert teams spanning biology, data science and engineering.

While AI offers unprecedented power, its impact today is bounded by the limitations of the data it consumes, the transparency of its operations and the structure within which it is deployed. Addressing these limitations is essential if we are to realize the full potential of AI in neuroscience drug discovery.

Many complementary efforts are addressing challenges that remain outside the reach of AI. Advances in human stem-cell-derived organoids and micro-physiological systems are providing experimentally tractable models that better capture both cellular and circuit-level context not yet represented *in silico* [85]. Large-scale longitudinal data initiatives like UK Biobank [86] and AllofUs [87] are improving diversity and generalizability while mitigating bias inherent in training data sets. In parallel, hybrid approaches that integrate mechanistic modeling with data-driven inference are emerging to bridge causal understanding with predictive power [88,89]. We focus here on the power of coupling AI with computations performed by the biological system, which we call "cellular intelligence".

7. Cellular intelligence

Data mining of very large and increasingly multi-modal molecular datasets ('multi-omics'), although demonstrably successful, has limitations, as noted above, which might be mitigated in part through a combined use of AI and innovation in experimental design.

Two examples to illustrate the point include the problem of the dependent variable or outcome, and the challenge of interpretation or how to use the results. In this section we discuss both, with some examples of how they might be addressed including with advanced analytics such as AI and ML. Importantly, the point is that simply using ever larger datasets combined with AI, whilst enormously valuable, is not

enough. The scientist is still an important actor in this play.

In observational studies it is the disease itself that is the dependent variable or outcome, and in the case of molecular data mining, it is the omics that is the independent variable being measured. This is challenging when it comes to AD and other neurodegenerative disease. AD is a common disorder of the elderly with a long preclinical phase and because of this, unaffected individuals are a less-than-optimal control or comparison group. Many elderly people will have disease pathology even if apparently unaffected and others will already be on course to do so. In case-control studies, the controls may not be so different from the cases. Various approaches can be used to circumvent this limitation; the dependent variable can be switched to age of onset [90], or some other clinical phenotype of interest, for example, comparing people with AD with slow versus rapid decline [91,92]. As biomarkers for pathology become available, the pathology itself can be used as the dependent variable [93]. The latter is highly attractive when it comes to target identification for drug discovery as it provides the potential to launch a precision intervention accompanied by biomarkers. Taking a precision approach further, sub-groups of AD might be identified such as those relatively resilient or vulnerable to pathology. As examples of precision sub-group creation, using GNPC and other data-sources, Oh et al. [94] identify markers of cognitive resilience; using a very deeply phenotyped cohort, Ng et al. [95] identify, and then validate *in vitro*, sub-groups of people with AD relatively resilient or vulnerable to amyloid pathology. Dolan et al. show that *in vitro* iPSC derived microglia have a validated disease transcriptional phenotype when challenged [96]. Combining these experimental approaches that go beyond the AD case / age matched control together with AI/ML in very large datasets seems a promising approach for future target identification. Especially when, as in some of the examples given, it seems possible to replicate *in vitro* disease relevant phenotypic response to challenges.

A second limitation of large-scale data mining comes after the successful delivery of results—how to interpret these results? Typically, the outcome of the data mining will come in the form of a list of potential targets ranked in some way to reflect their contribution to the differentiation of the dependent variable. It is in the nature of biology that this ranking is subtle – the difference between the top of the list and the middle of the list might be relatively small. Furthermore, it is in the nature of the analysis that repeating the exercise with the same data and the same analytical approach often yields a different list with a different ranking. This is to be expected; if a number of variables contribute equally to the differentiation, unless prevented, the model will represent all these variables with a single one. And perhaps a different one on repetition. In effect, the ranking of the list of variables contributing to the outcome is by itself a poor identifier of targets. Taking this into account, bioinformaticians will frequently represent the outcome list with a pathway nomination using GO terms or some similar approach to interpretation. Whilst this can be helpfully suggestive, all such bioinformatics tools have their limitations [97]. Alternatively, lists of targets can be parsed using a druggability assessment or validation from existing literature, both approaches being made easier using large language models to analyze the scientific literature.

An alternative mitigation for the challenge of interpretation is to not interpret but to instead to use the outcome of the analysis *in its entirety*. To use all of the omics signature detected, rather than trying to pick out targets. This was the in effect the driver behind the impactful NIH funded connectivity map (cMap) generated by the Broad Institute [98]. In the first iteration of this, the transcript map of cells perturbed with each of some 1500 compounds was generated and made available on a platform together with analytical software designed to allow researchers to compare expression signatures from disease to those generated by drugs. A number of studies have used the cMAP Gene Set Enrichment Analysis and other tools, to compare lists of genes differentially expressed in AD to the effect of compounds, seeking a signature counter-match (i.e. similar genes ranking but in opposite direction) as part of drug repurposing efforts [99–101].

Given that the targets of these compounds, representing those in clinical use, are known [102], then such approaches can be used to support novel compound discovery programs, as much as for repurposing. The cMAP and its successor, the LINCS program [103] have been followed by another Broad/MIT led program, The Joint Undertaking in Morphology and Cell Painting (JUMP-CP) in collaboration with a number of pharma, that instead of using expression analysis used morphological profiling. Essentially generating an image of cells perturbed by compounds, this public-private initiative was very high throughput and generated, and has made available, data on over 100k compounds [104]. Given the power of generative AI to analyze imaging data, this is an obviously rich source for deep learning on the effects of compounds on cells and the identification of targets. However, in contrast to the cMAP, there is no ready source of cell morphology data to compare to the compound perturbation. But clearly opportunities exist to develop such data – cell lines could be engineered to carry AD related genetic variants to identify morphological signatures which in turn could be matched to compound signatures. Such an *in silico* agnostic phenotypic screen might generate useful packages for drug discovery including compounds for target deconvolution and which might be used as tools for validation, as starting chemical matter for discovery programs or for repurposing efforts (Fig. 3).

There are many approaches to target identification using large data sets that could be enhanced using AI methods. One such is the use of real-world clinical data (RWD) either to identify or to validate targets. Whether from administrative data or from electronic medical records, the amount of RWD is steadily increasing as is the community of scientists using it; most obviously as represented in the Observational Health Data and Informatics initiative (https://www.ohdsi.org/). In neurodegenerative diseases such data was used to validate targets for Parkinson's disease identified through a screening program [105] although varying results from replication RWD studies demonstrate that interpreting such findings can be complicated [106,107]. Combining real world clinical data with genomic and other molecular data might add confidence in the findings, with an example from a public-private consortium study being the nomination of JAK-STAT signaling participants as targets for AD using a combination of real world clinical evidence, GWAS, expression data together with experimental data from preclinical models [108]. Given that so much real-world data is contained in text and given the explosive advances in the ability of AI to derive information from language inputs it seems very likely that using AI together with ML will significantly enhance the combined use of very large clinical and molecular data including electronic medical records (EMRs) augmented by biology.

8. The future of drug discovery

If drug discovery for Alzheimer's disease started with the cholinergic hypothesis and the identification of the protein forming the core of plaque pathology approximately a half-century ago, then it has to be acknowledged that the paradigm for TiTv in use for most of that time has been pretty successful. Using post-mortem studies of human brain together with hypothesis driven cell and molecular biology and biochemistry, a generation of scientists have identified targets that have fed today's rich and diverse drug discovery portfolio [109]. Now supplemented by ever larger datasets of the layers of biology from genomes through transcriptomes to proteomes and not forgetting lipidomes, metabolomes, microbiomes and so on and so forth, the cell and molecular biologist, the physiologist and biochemist have become increasingly adept in utilizing advanced analytics to derive information from such data and use this for precision neuroscience target identification and validation.

However, in the last few years with the sudden arrival – an instant success built on several lifetimes of work – of foundational models, large language models and generative AI, it seems as if the world of TiTv has just shifted, or if not yet, will soon. To a large degree, this will be an

incremental shift. AI is predictably going to make data mining of the increasingly large datasets more interesting and more informative, especially when those datasets become truly multi-modal across all layers of the biome and include imaging, complex real world data and adjacent data such as the environome. Massive datasets combined with AI analytics will rapidly eclipse current methodologies. Interpreting the results of such datasets for TiTv will also be facilitated by AI which is already today a more effective reader and user of the scientific literature than most of us human scientists.

The goal of target identification for Alzheimer's disease in the age of AI and CI is still to generate improved hypotheses that can lead to compounds that can be tested in human clinical trials (Fig. 3). The percentage of programs that are successful in clinical development are so low that any improvement in the accuracy, speed and variety of targets identified and validated will make a significant impact on the number and variety of clinical development programs. With faster, more comprehensive ways to analyze massive multi-omics datasets there is an opportunity to address the *inverse problem of drug discovery*, complementing approaches focused on characterizing the *forward pathways of disease* (Fig. 1).

Nonetheless, it seems likely that the scientist will also remain an essential part of the target identification process. As well as being a more effective data-miner, the AI combined with the scientist can be a smarter data-miner or user of the data. We have discussed here the ways in which AI could be used beyond data-mining, for example in compound signature matching, in identification of sub-groups of disease for precision intervention, in combining highly disparate types of data. There will be many others.

Experimentation is still required, both to generate signatures to perturb the analytical models, as well as to validate analytically determined hypotheses. Here the scientist is faced with a challenge: if the experiment gives a negative result is that because the model does not replicate the human disease to the same extent as the human-data analytical model, or is it because the human-data analytical model generated an incorrect hypothesis? Still, existing models, even when imperfect, provide the scaffolding for hypothesis generation and falsification, enabling continual learning across *in silico* and *in vitro* domains. The future of AI-driven discovery will be shaped as much by this disciplined cycle of use, evaluation and improvement as by the eventual realization of fully predictive disease models.

Although preclinical models of neurodegeneration remain insufficiently predictive and human validation is still limited, the trajectory is unmistakable. We are moving towards a future in which scientists partner with AI systems, not to replace insight, but to refine it, linking model-trained networks with biological intuition to identify, test and validate therapeutic targets in AD and other neurodegenerative disorders. Humans could not be more "in the loop" as patients, their families and caregivers become the ultimate beneficiaries of this dramatic progress.

9. Glossary

Agentic AI: AI systems capable of autonomously planning, reasoning, and executing sequences of actions toward a defined scientific or analytic goal. In biomedical research, agentic AI refers to "AI co-scientists" that can orchestrate data analysis, hypothesis generation and experiment design through iterative, self-directed workflows.

Artificial Intelligence (AI): Computational systems designed to perform tasks that typically require human intelligence, such as pattern recognition, prediction and reasoning.

Big Data: Extremely large and complex data sets requiring advanced computational tools for storage, integration and analysis.

Cellular Intelligence (CI): The capacity of human-derived cell systems to provide biological grounded insights that complement computational and human analytical approaches reading perturbations through signal transduction cascades, and producing observable or measurable changes in phenotype.

Foundational Model (FM): A large, pre-trained AI model developed on diverse, multimodal data that can be adapted (find-tuned) to specific biomedical tasks.

Generative AI: A class of AI models capable of producing new data, such as text, images, or synthetic omics profiles, based on patterns learned from training datasets.

Genome-Wide Association Study (GWAS): A statistical approach to identify genetic variants associated with disease risk by scanning the entire genome in large populations.

Induced pluripotent step cells (iPSCs): Stem cells reprogrammed from adult comatic cells that can differentiate into multiple cell types enabling disease modeling and drug testing.

Large Language Model (LLM): A neural network trained on massive text corpora to perform language-based reasoning, summarization, and data synthesis; increasingly applied in biomedical research.

Large Reasoning Model (LRM): An AI system integrating structured reasoning and retrieval mechanisms to support hypothesis generation and interpretation across multimodal data.

Multi-omics: Integrated biological data combining multiple "omics" layers such as genetic (DNA), transcriptomics (mRNA), proteomics and metabolomics to provide a holistic view of disease biology.

Population cohort: A large group of individuals followed over time in a research study to assess biological, clinical or genetic factors related to disease risk or progression

Real-World Data (RWD): Data derived from sources outside traditional clinical trials, such as electronic medical records, insurance claims, disease registries, and digital health platforms, increasingly used to complement experimental data.

Self-Supervised Learning: An AI training approach in which models learn patterns or representations from unlabeled data by predicting hidden or missing parts of the input, often used to build large foundation models.

Target Identification and Target Validation (TiTv): The process of discovering, prioritizing and experimentally confirming molecular entities that can be modulated to achieve clinical benefit in disease treatment.

Quantitative Trait Locus (QTL): A genomic region associated with variation in a measurable trait, such as gene expression (eQTL) or protein abundance (pQTL).

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT and Claude in order to assist with initial content ideation within human-defined subsections. After using this tool/service, the author(s) reviewed and edited the content heavily and take(s) full responsibility for the content of the publication.

CRediT authorship contribution statement

Gayle Wittenberg: Writing – review & editing, Writing – original draft, Visualization, Conceptualization. **Fiona Elwood:** Writing – review & editing, Writing – original draft. **Andrea Houghton:** Writing – review & editing, Writing – original draft. **Tommaso Mansi:** Writing – review & editing, Writing – original draft. **Bart Smets:** Writing – review & editing, Writing – original draft, Visualization. **Simon Lovestone:** Writing – review & editing, Writing – original draft, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Gayle Wittenberg reports financial support was provided by Johnson & Johnson Innovative Medicines. Fiona Elwood reports financial support

was provided by Johnson & Johnson Innovative Medicines. Andrea Houghton reports financial support was provided by Johnson & Johnson Innovative Medicines. Tommaso Mansi reports financial support was provided by Johnson & Johnson Innovative Medicines. Bart Smets reports financial support was provided by Johnson & Johnson Innovative Medicines. Simon Lovestone reports financial support was provided by Johnson & Johnson Innovative Medicines. Gayle Wittenberg reports a relationship with Johnson & Johnson that includes: employment and equity or stocks. Fiona Elwood reports a relationship with Johnson & Johnson that includes: employment and equity or stocks. Andrea Houghton reports a relationship with Johnson & Johnson that includes: employment and equity or stocks. Tommaso Mansi reports a relationship with Johnson & Johnson that includes: employment and equity or stocks. Bart Smets reports a relationship with Johnson & Johnson that includes: employment and equity or stocks. Simon Lovestone reports a relationship with Johnson & Johnson that includes: employment and equity or stocks. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Noble D. A theory of biological relativity: no privileged level of causation. Interface Focus 2012;2:55–64. https://doi.org/10.1098/rsfs.2011.0067.
[2] Sugimoto H, Ogura H, Arai Y, Iimura Y, Yamanishi Y. Research and development of Donepezil hydrochloride, a new type of acetylcholinesterase inhibitor. Jpn J Pharmacol 2002;89:7–20. https://doi.org/10.1254/jjp.89.7.
[3] Bartus RT. On neurodegenerative diseases, models, and treatment strategies: lessons learned and lessons forgotten a generation following the cholinergic hypothesis. Exp Neurol 2000;163:495–529. https://doi.org/10.1006/exnr.2000.7397.
[4] Tiseo, Rogers, Friedhoff. Pharmacokinetic and pharmacodynamic profile of donepezil HCl following evening administration. Br J Clin Pharmacol 1998;46:13–8. https://doi.org/10.1046/j.1365-2125.1998.0460s1013.x.
[5] Weggen S, Beher D. Molecular consequences of amyloid precursor protein and presenilin mutations causing autosomal-dominant Alzheimer's disease. Alzheimers Res Ther 2012;4:9. https://doi.org/10.1186/alzrt107.
[6] Hsiao K, Chapman P, Nilsen S, Eckman C, Harigaya Y, Younkin S, et al. Correlative memory deficits, aβ elevation, and amyloid plaques in transgenic mice. Science 1996;274:99–103. https://doi.org/10.1126/science.274.5284.99.
[7] Neumann U, Machauer R, Shimshek DR. The β-secretase (BACE) inhibitor NB-360 in preclinical models: from amyloid-β reduction to downstream disease-relevant effects. Br J Pharmacol 2019;176:3435–46. https://doi.org/10.1111/bph.14582.
[8] Barak M, Fedorova V, Pospisilova V, Raska J, Vochyanova S, Sedmik J, et al. Human iPSC-derived neural models for studying Alzheimer's disease: from neural stem cells to cerebral organoids. Stem Cell Rev Rep 2022;18:792–820. https://doi.org/10.1007/s12015-021-10254-3.
[9] Sen T, Thummer RP. CRISPR and iPSCs: recent developments and future perspectives in neurodegenerative disease modelling, research, and therapeutics. Neurotox Res 2022;40:1597–623. https://doi.org/10.1007/s12640-022-00564-w.
[10] Ramos DM, Skarnes WC, Singleton AB, Cookson MR, Ward ME. Tackling neurodegenerative diseases with genomic engineering: a new stem cell initiative from the NIH. Neuron 2021;109:1080–3. https://doi.org/10.1016/j.neuron.2021.03.022.
[11] Quaid K, Xing X, Chen Y-H, Miao Y, Neilson A, Selvamani V, et al. iPSCs and iPSC-derived cells as a model of human genetic and epigenetic variation. Nat Commun 2025;16:1750. https://doi.org/10.1038/s41467-025-56569-4.
[12] Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, et al. The support of human genetic evidence for approved drug indications. Nat Genet 2015;47:856–60. https://doi.org/10.1038/ng.3314.
[13] Van de Sande B, Lee JS, Mutasa-Gottgens E, Naughton B, Bacon W, Manning J, et al. Applications of single-cell RNA sequencing in drug discovery and development. Nat Rev Drug Discov 2023;22:496–520. https://doi.org/10.1038/s41573-023-00688-4.
[14] Bellenguez C, Küçükali F, Jansen IE, Kleineidam L, Moreno-Grau S, Amin N, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. Nat Genet 2022;54:412–36. https://doi.org/10.1038/s41588-022-01024-z.
[15] Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. Nat Genet 2019;51:404–13. https://doi.org/10.1038/s41588-018-0311-9.
[16] Marioni RE, Harris SE, Zhang Q, McRae AF, Hagenaars SP, Hill WD, et al. GWAS on family history of Alzheimer's disease. Transl Psychiatry 2018;8. https://doi.org/10.1038/s41398-018-0150-6.
[17] Gallagher MD, AS Chen-Plotkin. The Post-GWAS era: from association to function. Am J Hum Genet 2018;102:717–30. https://doi.org/10.1016/j.ajhg.2018.04.002.
[18] Sierksma A, Escott-Price V, De Strooper B. Translating genetic risk of Alzheimer's disease into mechanistic insight and drug targets. Science 2020;370:61–6. https://doi.org/10.1126/science.abb8575.
[19] Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet 2016;48:481–7. https://doi.org/10.1038/ng.3538.
[20] Li B, Ritchie MD. From GWAS to gene: transcriptome-wide association studies and other methods to functionally understand GWAS discoveries. Front Genet 2021;12. https://doi.org/10.3389/fgene.2021.713230.
[21] Sun BB, Chiou J, Traylor M, Benner C, Hsu Y-H, Richardson TG, et al. Plasma proteomic associations with genetics and health in the UK Biobank. Nature 2023;622:329–38. https://doi.org/10.1038/s41586-023-06592-6.
[22] Yang C, Farias FHG, Ibanez L, Suhy A, Sadler B, Fernandez MV, et al. Genomic atlas of the proteome from brain, CSF and plasma prioritizes proteins implicated in neurological disorders. Nat Neurosci 2021;24:1302–12. https://doi.org/10.1038/s41593-021-00886-6.
[23] Yang C, Fagan AM, Perrin RJ, Rhinn H, Harari O, Cruchaga C. Mendelian randomization and genetic colocalization infer the effects of the multi-tissue proteome on 211 complex disease-related phenotypes. Genome Med 2022;14:140. https://doi.org/10.1186/s13073-022-01140-9.
[24] Western D, Timsina J, Wang L, Wang C, Yang C, Phillips B, et al. Proteogenomic analysis of human cerebrospinal fluid identifies neurologically relevant regulation and implicates causal proteins for Alzheimer's disease. Nat Genet 2024;56:2672–84. https://doi.org/10.1038/s41588-024-01972-8.
[25] Wingo AP, Liu Y, Gerasimov ES, Gockley J, Logsdon BA, Duong DM, et al. Integrating human brain proteomes with genome-wide association data implicates new proteins in Alzheimer's disease pathogenesis. Nat Genet 2021;53:143–6. https://doi.org/10.1038/s41588-020-00773-z.
[26] Hu T, Liu Q, Dai Q, Buchman AS, Bennett DA, Tasaki S, et al. Proteome-wide association studies using summary pQTL data of brain, CSF, and plasma identify 30 risk genes of Alzheimer's disease dementia. Alzheimers Res Ther 2025;17:135. https://doi.org/10.1186/s13195-025-01774-y.
[27] Sun Y, Zhu J, Zhou D, Canchi S, Wu C, Cox NJ, et al. A transcriptome-wide association study of Alzheimer's disease using prediction models of relevant tissues identifies novel candidate susceptibility genes. Genome Med 2021;13:141. https://doi.org/10.1186/s13073-021-00959-y.
[28] Liu S, Cho M, Huang Y-N, Park T, Chaudhuri S, Rosewood TJ, et al. Multi-omics analysis for identifying cell-type-specific and bulk-level druggable targets in Alzheimer's disease. J Transl Med 2025;23:788. https://doi.org/10.1186/s12967-025-06739-1.
[29] Wan Y-W, Al-Ouran R, Mangleburg CG, Perumal TM, Lee TV, Allison K, et al. Meta-analysis of the Alzheimer's Disease Human brain transcriptome and functional dissection in mouse models. Cell Rep 2020;32:107908. https://doi.org/10.1016/j.celrep.2020.107908.
[30] Wang M, Roussos P, McKenzie A, Zhou X, Kajiwara Y, Brennand KJ, et al. Integrative network analysis of nineteen brain regions identifies molecular signatures and networks underlying selective regional vulnerability to Alzheimer's disease. Genome Med 2016;8:104. https://doi.org/10.1186/s13073-016-0355-3.
[31] Mathys H, Adaikkan C, Gao F, Young JZ, Manet E, Hemberg M, et al. Temporal tracking of microglia activation in neurodegeneration at single-cell resolution. Cell Rep 2017;21:366–80. https://doi.org/10.1016/j.celrep.2017.09.039.
[32] Olah M, Menon V, Habib N, Taga MF, Ma Y, Yung CJ, et al. Single cell RNA sequencing of human microglia uncovers a subset associated with Alzheimer's disease. Nat Commun 2020;11:6129. https://doi.org/10.1038/s41467-020-19737-2.
[33] Habib N, McCabe C, Medina S, Varshavsky M, Kitsberg D, Dvir-Szternfeld R, et al. Disease-associated astrocytes in Alzheimer's disease and aging. Nat Neurosci 2020;23:701–6. https://doi.org/10.1038/s41593-020-0624-8.
[34] Mathys H, Peng Z, Boix CA, Victor MB, Leary N, Babu S, et al. Single-cell atlas reveals correlates of high cognitive function, dementia, and resilience to Alzheimer's disease pathology. Cell 2023;186:4365–85. https://doi.org/10.1016/j.cell.2023.08.039. e27.
[35] Swarup V, Chang TS, Duong DM, Dammer EB, Dai J, Lah JJ, et al. Identification of conserved proteomic networks in neurodegenerative dementia. Cell Rep 2020;31:107807. https://doi.org/10.1016/j.celrep.2020.107807.
[36] Johnson ECB, Carter EK, Dammer EB, Duong DM, Gerasimov ES, Liu Y, et al. Large-scale deep multi-layer analysis of Alzheimer's disease brain reveals strong proteomic disease-related changes not observed at the RNA level. Nat Neurosci 2022;25:213–25. https://doi.org/10.1038/s41593-021-00999-y.
[37] Pichet Binette A, Gaiteri C, Wennström M, Kumar A, Hristovska I, Spotorno N, et al. Proteomic changes in Alzheimer's disease associated with progressive Aβ plaque and tau tangle pathologies. Nat Neurosci 2024;27:1880–91. https://doi.org/10.1038/s41593-024-01737-w.
[38] Xu J, Bankov G, Kim M, Wretlind A, Lord J, Green R, et al. Integrated lipidomics and proteomics network analysis highlights lipid and immunity pathways associated with Alzheimer's disease. Transl Neurodegener 2020;9:36. https://doi.org/10.1186/s40035-020-00215-0.
[39] Nativio R, Lan Y, Donahue G, Sidoli S, Berson A, Srinivasan AR, et al. An integrated multi-omics approach identifies epigenetic alterations associated with Alzheimer's disease. Nat Genet 2020;52:1024–35. https://doi.org/10.1038/s41588-020-0696-0.
[40] Shi L, Xu J, Green R, Wretlind A, Homann J, Buckley NJ, et al. Multiomics profiling of human plasma and cerebrospinal fluid reveals ATN-derived networks and highlights causal links in Alzheimer's disease. Alzheimers Dement J Alzheimers Assoc 2023;19:3350–64. https://doi.org/10.1002/alz.12961.

[41] Imam F, Saloner R, Vogel JW, Krish V, Abdel-Azim G, Ali M, et al. The Global Neurodegeneration Proteomics Consortium: biomarker and drug target discovery for common neurodegenerative diseases and aging. Nat Med 2025. https://doi.org/10.1038/s41591-025-03834-0.
[42] Minikel EV, Karczewski KJ, Martin HC, Cummings BB, Whiffin N, Rhodes D, et al. Evaluating drug targets through human loss-of-function genetic variation. Nature 2020;581:459–64. https://doi.org/10.1038/s41586-020-2267-z.
[43] Carss KJ, Deaton AM, Del Rio-Espinola A, Diogo D, Fielden M, Kulkarni DA, et al. Using human genetics to improve safety assessment of therapeutics. Nat Rev Drug Discov 2023;22:145–62. https://doi.org/10.1038/s41573-022-00561-w.
[44] Han X, Zhou Z, Fei L, Sun H, Wang R, Chen Y, et al. Construction of a human cell landscape at single-cell level. Nature 2020;581:303–9. https://doi.org/10.1038/s41586-020-2157-4.
[45] Chen J, Wu J, Bai Y, Yang C, Wang J. Recent advances of single-cell RNA sequencing in toxicology research: insight into hepatotoxicity and nephrotoxicity. Curr Opin Toxicol 2024;37:100462. https://doi.org/10.1016/j.cotox.2024.100462.
[46] Li T, Chen X, Tong W. Bridging organ transcriptomics for advancing multiple organ toxicity assessment with a generative AI approach. Npj Digit Med 2024;7:310. https://doi.org/10.1038/s41746-024-01317-z.
[47] Hopkins AL, Groom CR. The druggable genome. Nat Rev Drug Discov 2002;1:727–30. https://doi.org/10.1038/nrd892.
[48] Finan C, Gaulton A, Kruger FA, Lumbers RT, Shah T, Engmann J, et al. The druggable genome and support for target identification and validation in drug development. Sci Transl Med 2017;9:eaag1166. https://doi.org/10.1126/scitranslmed.aag1166.
[49] Schneider M, Radoux CJ, Hercules A, Ochoa D, Dunham I, Zalmas L-P, et al. The PROTACtable genome. Nat Rev Drug Discov 2021;20:789–97. https://doi.org/10.1038/s41573-021-00245-x.
[50] Whitehead KA, Langer R, Anderson DG. Knocking down barriers: advances in siRNA delivery. Nat Rev Drug Discov 2009;8:129–38. https://doi.org/10.1038/nrd2742.
[51] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. Nat Genet 2000;25:25–9. https://doi.org/10.1038/75556.
[52] Consortium The Gene Ontology, SA Aleksander, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, et al. The gene ontology knowledgebase in 2023. GENETICS 2023;224:iyad031. https://doi.org/10.1093/genetics/iyad031.
[53] Kanehisa MKEGG. Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 2000;28:27–30. https://doi.org/10.1093/nar/28.1.27.
[54] Krämer A, Green J, Pollard J, Tugendreich S. Causal analysis approaches in ingenuity pathway analysis. Bioinforma Oxf Engl 2014;30:523–30. https://doi.org/10.1093/bioinformatics/btt703.
[55] Clarke R, Ressom HW, Wang A, Xuan J, Liu MC, Gehan EA, et al. The properties of high-dimensional data spaces: implications for exploring gene and protein expression data. Nat Rev Cancer 2008;8:37–49. https://doi.org/10.1038/nrc2294.
[56] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 2008;9:559. https://doi.org/10.1186/1471-2105-9-559.
[57] Jiménez-Jiménez V, Martí-Gómez C, Del Pozo MÁ, Lara-Pezzi E, Sánchez-Cabo F. Bayesian inference of gene expression. In: Helder IN, editor. Bioinformatics. Brisbane (AU): Exon Publications; 2021.
[58] Huang H, Wang Y, Rudin C, Browne EP. Towards a comprehensive evaluation of dimension reduction methods for transcriptomic data visualization. Commun Biol 2022;5:719. https://doi.org/10.1038/s42003-022-03628-x.
[59] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A, et al. Attention is all you need. Adv. Neural Inf. Process. Syst. 2017;30.
[60] Bommasani R., Hudson D.A., Adeli E., Altman R., Arora S., Arx S von, et al. On the opportunities and risks of foundation models 2022. https://doi.org/10.48550/arXiv.2108.07258.
[61] Gui J, Chen T, Zhang J, Cao Q, Sun Z, Luo H, et al. A survey on self-supervised learning: algorithms, applications, and future trends. IEEE Trans Pattern Anal Mach Intell 2024;46:9052–71. https://doi.org/10.1109/TPAMI.2024.3415112.
[62] Juan Ramon A, Parmar C, Carrasco-Zevallos OM, Csiszer C, Yip SSF, Raciti P, et al. Development and deployment of a histopathology-based deep learning algorithm for patient prescreening in a clinical trial. Nat Commun 2024;15:4690. https://doi.org/10.1038/s41467-024-49153-9.
[63] Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, et al. Language models are few-shot learners. In: Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H, editors. Adv. neural inf. process. syst. Adv. neural inf. process. syst, 33. Curran Associates, Inc.; 2020. p. 1877–901.
[64] Radford A., Kim J.W., Hallacy C., Ramesh A., Goh G., Agarwal S., et al. Learning transferable visual models from natural language supervision 2021. https://doi.org/10.48550/ARXIV.2103.00020.
[65] 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;21:1470–80. https://doi.org/10.1038/s41592-024-02201-0.
[66] Zeng Y, Xie J, Shangguan N, Wei Z, Li W, Su Y, et al. CellFM: a large-scale foundation model pre-trained on transcriptomics of 100 million human cells. Nat Commun 2025;16:4679. https://doi.org/10.1038/s41467-025-59926-5.
[67] Avsec Ž., Latysheva N., Cheng J., Novati G., Taylor K.R., Ward T., et al. AlphaGenome: advancing regulatory variant effect prediction with a unified DNA sequence model 2025. https://doi.org/10.1101/2025.06.25.661532.
[68] Cui T., Xu S.-J., Moskalev A., Li S., Mansi T., Prakash M., et al. InfoSEM: a deep generative model with informative priors for gene regulatory network inference 2025. https://doi.org/10.48550/ARXIV.2503.04483.
[69] Snell C., Lee J., Xu K., Kumar A. Scaling LLM test-time compute optimally can be more effective than Scaling model parameters 2024. https://doi.org/10.48550/ARXIV.2408.03314.
[70] Gottweis J., Weng W.-H., Daryin A., Tu T., Palepu A., Sirkovic P., et al. Towards an AI co-scientist 2025. https://doi.org/10.48550/ARXIV.2502.18864.
[71] Huang K, Zhang S, Wang H, Qu Y, Lu Y, Roohani Y, et al. Biomni: a general-purpose biomedical AI agent. BioRxiv Prepr Serv Biol 2025;2025:05. https://doi.org/10.1101/2025.05.30.656746. 30.656746.
[72] Wang H., He Y., Coelho P.P., Bucci M., Nazir A., Chen B., et al. SpatialAgent: an autonomous AI agent for spatial biology 2025. https://doi.org/10.1101/2025.04.03.646459.
[73] Kaplan J., McCandlish S., Henighan T., Brown T.B., Chess B., Child R., et al. Scaling laws for neural language models 2020. https://doi.org/10.48550/arXiv.2001.08361.
[74] Bourached A, Bonkhoff AK, Schirmer MD, Regenhardt RW, Bretzner M, Hong S, et al. Scaling behaviours of deep learning and linear algorithms for the prediction of stroke severity. Brain Commun 2024;6:fcae007. https://doi.org/10.1093/braincomms/fcae007.
[75] Huang L, Yu W, Ma W, Zhong W, Feng Z, Wang H, et al. A survey on hallucination in large language models: principles, taxonomy, challenges, and open questions. ACM Trans Inf Syst 2025;43:1–55. https://doi.org/10.1145/3703155.
[76] Asgari E, Montaña-Brown N, Dubois M, Khalil S, Balloch J, Yeung JA, et al. A framework to assess clinical safety and hallucination rates of LLMs for medical text summarisation. Npj Digit Med 2025;8:274. https://doi.org/10.1038/s41746-025-01670-7.
[77] Valmeekam K., Olmo A., Sreedharan S., Kambhampati S. Large language models still can't plan (A Benchmark for LLMs on Planning and Reasoning about Change) n.d.
[78] Shojaee P., Mirzadeh I., Alizadeh K., Horton M., Bengio S., Farajtabar M. The illusion of thinking: understanding the strengths and limitations of reasoning models via the lens of problem complexity 2025. https://doi.org/10.48550/arXiv.2506.06941.
[79] d'Avila Garcez A, LC Lamb. Neurosymbolic AI: the 3rd wave. Artif Intell Rev 2023;56:12387–406. https://doi.org/10.1007/s10462-023-10448-w.
[80] Berrevoets J, Kacprzyk K, Qian Z, der Schaar M van. Causal deep learning: encouraging impact on real-world problems through causality. Found Trends® Signal Process 2024;18:200–309. https://doi.org/10.1561/2000000123.
[81] Sirugo G, Williams SM, Tishkoff SA. The missing diversity in Human genetic studies. Cell 2019;177:26–31. https://doi.org/10.1016/j.cell.2019.02.048.
[82] Chen IY, Joshi S, Ghassemi M. Treating health disparities with artificial intelligence. Nat Med 2020;26:16–7. https://doi.org/10.1038/s41591-019-0649-2.
[83] Aisen PS, Cummings J, Jack CR, Morris JC, Sperling R, Frölich L, et al. On the path to 2025: understanding the Alzheimer's disease continuum. Alzheimers Res Ther 2017;9:60. https://doi.org/10.1186/s13195-017-0283-5.
[84] Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 2019;1:206–15. https://doi.org/10.1038/s42256-019-0048-x.
[85] Smirnova L, Hartung T. The promise and potential of brain organoids. Adv Healthc Mater 2024;13:2302745. https://doi.org/10.1002/adhm.202302745.
[86] Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLOS Med 2015;12:e1001779. https://doi.org/10.1371/journal.pmed.1001779.
[87] The All of us research Program investigators. The "all of us" research program. N Engl J Med 2019;381:668–76. https://doi.org/10.1056/NEJMsr1809937.
[88] Noordijk B, Garcia Gomez ML, ten Tusscher KHWJ, de Ridder D, van Dijk ADJ, Smith RW. The rise of scientific machine learning: a perspective on combining mechanistic modelling with machine learning for systems biology. Front Syst Biol 2024;4. https://doi.org/10.3389/fsysb.2024.1407994.
[89] Procopio A, Cesarelli G, Donisi L, Merola A, Amato F, Cosentino C. Combined mechanistic modeling and machine-learning approaches in systems biology – A systematic literature review. Comput Methods Programs Biomed 2023;240:107681. https://doi.org/10.1016/j.cmpb.2023.107681.
[90] Kamboh MI, Barmada MM, Demirci FY, Minster RL, Carrasquillo MM, Pankratz VS, et al. Genome-wide association analysis of age-at-onset in Alzheimer's disease. Mol Psychiatry 2012;17:1340–6. https://doi.org/10.1038/mp.2011.135.
[91] Thambisetty M, An Y, Kinsey A, Koka D, Saleem M, Güntert A, et al. Plasma clusterin concentration is associated with longitudinal brain atrophy in mild cognitive impairment. NeuroImage 2012;59:212–7. https://doi.org/10.1016/j.neuroimage.2011.07.056.
[92] Cruchaga C, Kauwe JSK, Mayo K, Spiegel N, Bertelsen S, Nowotny P, et al. SNPs associated with cerebrospinal fluid phospho-tau levels influence rate of decline in Alzheimer's disease. PLoS Genet 2010;6:e1001101. https://doi.org/10.1371/journal.pgen.1001101.
[93] Saari TT, Palviainen T, Hiltunen M, Herukka S-K, Kokkola T, Kärkkäinen S, et al. Cross-sectional study of plasma phosphorylated tau 217 in persons without dementia. Alzheimers Dement Amst Neth 2025;17:e70107. https://doi.org/10.1002/dad2.70107.
[94] Oh HS-H, Urey DY, Karlsson L, Zhu Z, Shen Y, Farinas A, et al. A cerebrospinal fluid synaptic protein biomarker for prediction of cognitive resilience versus

decline in Alzheimer's disease. Nat Med 2025;31:1592–603. https://doi.org/10.1038/s41591-025-03565-2.
[95] Ng B, Rowland HA, Wei T, Arunasalam K, Hayes EM, Koychev I, et al. Neurons derived from individual early Alzheimer's disease patients reflect their clinical vulnerability. Brain Commun 2022;4:fcac267. https://doi.org/10.1093/braincomms/fcac267.
[96] Dolan M-J, Therrien M, Jereb S, Kamath T, Gazestani V, Atkeson T, et al. Exposure of iPSC-derived human microglia to brain substrates enables the generation and manipulation of diverse transcriptional states *in vitro*. Nat Immunol 2023;24:1382–90. https://doi.org/10.1038/s41590-023-01558-2.
[97] Khatri P, Sirota M, Butte AJ. Ten years of pathway analysis: current approaches and outstanding challenges. PLoS Comput Biol 2012;8:e1002375. https://doi.org/10.1371/journal.pcbi.1002375.
[98] Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science 2006;313:1929–35. https://doi.org/10.1126/science.1132939.
[99] Bellur O, Kastenmuller G, Requena F, Kaddurah-Daouk RF, Krumsiek J, Arnold M. In silico prioritiziation of drug repositioning candidates for Alzheimer's disease using signature search meta-analysis. Alzheimers Dement 2023;19. https://doi.org/10.1002/alz.076844.
[100] Williams G, Gatt A, Clarke E, Corcoran J, Doherty P, Chambers D, et al. Drug repurposing for Alzheimer's disease based on transcriptional profiling of human iPSC-derived cortical neurons. Transl Psychiatry 2019;9:220. https://doi.org/10.1038/s41398-019-0555-x.
[101] Xu Y, Kong J, Hu P. Computational drug repurposing for Alzheimer's disease using risk genes from GWAS and single-cell RNA sequencing studies. Front Pharmacol 2021;12:617537. https://doi.org/10.3389/fphar.2021.617537.
[102] Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, et al. A comprehensive map of molecular drug targets. Nat Rev Drug Discov 2017;16:19–34. https://doi.org/10.1038/nrd.2016.230.
[103] Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A next generation connectivity map: L1000 platform and the first 1000,000 profiles. Cell 2017;171:1437–52. https://doi.org/10.1016/j.cell.2017.10.049. e17.
[104] Chandrasekaran SN, Cimini BA, Goodale A, Miller L, Kost-Alimova M, Jamali N, et al. Three million images and morphological profiles of cells treated with matched chemical and genetic perturbations. Nat Methods 2024;21:1114–21. https://doi.org/10.1038/s41592-024-02241-6.
[105] Mittal S, Bjørnevik K, Im DS, Flierl A, Dong X, Locascio JJ, et al. β2-Adrenoreceptor is a regulator of the α-synuclein gene driving risk of Parkinson's disease. Science 2017;357:891–8. https://doi.org/10.1126/science.aaf3934.
[106] Giorgianni F, Ernst P, Dell'Aniello S, Suissa S, Renoux C. β 2-agonists and the incidence of Parkinson Disease. Am J Epidemiol 2020;189:801–10. https://doi.org/10.1093/aje/kwaa012.
[107] Searles Nielsen S, Gross A, Camacho-Soto A, Willis AW, Racette BA. β2-adrenoreceptor medications and risk of Parkinson disease. Ann Neurol 2018;84:683–93. https://doi.org/10.1002/ana.25341.
[108] Nevado-Holgado AJ, Ribe E, Thei L, Furlong L, Mayer M-A, Quan J, et al. Genetic and real-world clinical data, combined with empirical validation, nominate Jak-stat signaling as a target for Alzheimer's disease therapeutic development. Cells 2019;8:425. https://doi.org/10.3390/cells8050425.
[109] Cummings J, Zhou Y, Lee G, Zhong K, Fonseca J, Cheng F. Alzheimer's disease drug development pipeline: 2024. Alzheimers Dement Transl Res Clin Interv 2024;10:e12465. https://doi.org/10.1002/trc2.12465.