Proceedings of CONF-MLA 2026 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2026.LD39440

Machine Learning for Transcription Factor Target Discovery and Drug Intervention: Progress, Challenges and Outlook

Zichang Liu

School of Marine Sciences, Ningbo University, Ningbo, China
liuzichang@nbu.edu.cn

Abstract. Machine learning now reaches every stage of transcription factor (TF) drug discovery, but its contribution falls sharply along the evidence hierarchy. This review works through five tiers of evidence, from sequence and transcriptome through protein structure, cellular phenotype and chemical matter to clinical validation, and examines recent progress and the strength of evidence at each. Unified models can now predict binding, accessibility and expression within a single framework, and the definition of a ligandable site has widened from a static cavity to a transient conformational state, so disordered regions are no longer excluded a priori. The downstream tiers are far thinner. Molecular generation keeps improving geometric plausibility, but a unified benchmark shows that such metrics do not track the ability to produce active molecules, and no clinical intervention to date can be traced to substantive machine learning input. Missing trivial baselines and information leakage also make retrospective evaluation systematically optimistic. What limits machine learning at each tier appears to be the supply of verifiable labels more than model capacity. TFs sit upstream of regulation, so a single successful intervention may correct an entire downstream expression programme, which keeps them attractive for diseases that still lack usable targets and demanding as a test of algorithmic prediction. On this reading, further gains for TFs and other undruggable targets depend more on what each tier can be made to measure than on new architectures.

Keywords: Transcription Factor, Machine Learning, Deep Learning, Undruggable Target, Computer-Aided Drug Design

1. Introduction

Transcription factors (TFs) act upstream of gene expression and determine which genes are read in which cellular context [1]. A substantial share of TFs is linked directly to disease phenotypes [2], and in many diseases the abnormality lies in the whole transcriptional programme rather than in one defective gene [3], so intervening at a single TF offers a chance to correct an entire downstream expression abnormality. Yet TFs have long been considered undruggable. Their activation domains are rich in intrinsically disordered regions and recruit the transcriptional machinery through phase-separated condensates, so their function is spread across dynamic multivalent interactions. They have neither a catalytic site nor a stable small-molecule binding pocket, which leaves structure-guided drug design without an obvious foothold [2, 4]. Disordered structure, dynamic interaction

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.LD39440

and network-level organisation also compound one another, so TFs test computational methods harder than most target classes do.

Work on TF tractability has recently widened from the single protein to the higher-order organisation of the transcriptional regulatory network, and strategies that modulate autoinhibitory conformations, that use proteolysis-targeting chimeras and that address disordered regions directly have appeared in turn [5, 6]. More intervention points bring longer target lists, wider annotation gaps and greater network complexity, which is the kind of problem machine learning is brought in to handle. Models now infer regulation from sequence, identify ligandable sites from structure, predict cellular responses to perturbation and generate candidate molecules directly [7]. How much they contribute at each step is a separate question, because the supervisory signals and the means of validation differ greatly from one step to the next. This review follows five tiers of evidence, from sequence and transcriptome through protein structure, cellular phenotype and chemical matter to clinical validation, and asks at each tier how far machine learning has penetrated and how strong the supporting evidence is.

2. Machine learning across the evidence hierarchy of transcription factor drugging

2.1. TF screening from sequence and regulatory networks

Three lines of work lead into TF screening. In sequence inference, DeepBind, Basenji and Enformer introduced convolutional neural networks (CNNs) and then attention, extending the receptive field to 100 kb [8-10]. In network inference, GENIE3 ranked candidate links with tree ensembles, GRNBoost2 scaled that ranking to single-cell RNA sequencing (scRNA-seq), SCENIC pruned the result by motif enrichment, SCENIC+ brought the analysis down to the enhancer level, and CellOracle added in silico perturbation of the inferred network [11-15]. Activity inference has moved more slowly, limited throughout by the quality of prior knowledge [16, 17].

Epiregulon, published in 2025, links activity inference to drug response [18]. It weights a regulatory connection by whether TF expression and chromatin accessibility at the TF binding site occur together in the same cell, which covers the case where TF activity is decoupled from mRNA level. It also works from ChIP-seq occupancy data covering 1,377 factors rather than from motifs, so it can score transcriptional coregulators that have no defined motif of their own. In testing, it predicted the effects of both an androgen receptor (AR) antagonist and an AR degrader [18].

AlphaGenome, published this year, loosens the trade-off between input length and output resolution [19]. It takes 1 Mb of sequence and predicts 5,930 human genome tracks across 11 modalities at up to single-base resolution, matching or exceeding the strongest external specialist model in 22 of 24 track tasks and 25 of 26 variant-effect tasks [19]. TF binding is one of its explicit output modalities, so a screen can score binding, accessibility and expression inside one model instead of stitching separate models together.

2.2. Ligandable site identification from protein structure and geometric deep learning

Druggability depends on whether the protein surface carries a site that a small molecule can bind, and the absence of a classical hydrophobic cavity was once treated as the main source of project failure [20]. Pocket detection then developed along two routes. The voxel-based methods came first: DeepSite learned cavity features end to end, P2Rank addressed the balance between speed and accuracy, and DeepPocket ranked candidate sites [21-23]. MaSIF and ScanNet left the Euclidean

DOI: 10.54254/2755-2721/2026.LD39440

grid, the first learning fingerprints on the molecular surface, the second extracting interpretable features end to end [24, 25].

Both routes assume a stably folded protein whose site sits in a cavity visible in the static structure. The intrinsically disordered regions (IDRs) of TFs and their dynamic multivalent interactions break that assumption. PocketMiner works on cryptic pockets instead, using a graph neural network (GNN) to predict from a single structure where a pocket may open during molecular dynamics (MD), with an area under the receiver operating characteristic curve (ROC-AUC) of 0.87 [26]. The prediction target becomes a conformation that the protein populates only transiently.

A 2026 experimental study carried that idea into the disordered region itself. Its subject is the disordered N-terminal activation domain of the androgen receptor, a target in castration-resistant prostate cancer. EPI-001, a small molecule that has entered clinical trials, does not bind the monomer, it selectively recognises an oligomer that this domain forms transiently. The oligomer is more ordered than the monomer and is stabilised by interactions between aromatic residues in partially α-helical regions, with transactivation unit 5 (Tau-5*, residues 330–448) as the main site of interaction. Binding reshapes the conformational ensemble of the target, changes the properties of its phase-separated condensates and reduces recruitment of RNA polymerase II in cells [27]. Selectivity here comes from a conformational state that only some members of the family populate extensively, and owes nothing to pocket geometry. Oligomerisation propensity and the aromatic motifs of partially folded regions would therefore be worth predicting, and a model that could do so would give TFs a chemical entry point.

2.3. Modelling perturbation response in transcriptomic and imaging read-outs

Three datasets underpin the modelling of perturbation response: the million-scale expression profiles of L1000 [28], the single-cell read-out of Perturb-seq and a genome-scale screen of roughly 2.5 million human cells [29, 30]. On the transcriptomic side, scGen handled transfer to unseen cells, the compositional perturbation autoencoder (CPA) disentangled basal state from perturbation, GEARS inferred the effect of perturbations that were never measured, and CellOT predicted responses without paired samples, the underlying machinery moving from the variational autoencoder (VAE) to optimal transport (OT) [31-34], scGPT trades large-scale pretraining for a general representation [35]. On the imaging side, Cell Painting automated the measurement of morphological features [36], and CPJUMP1, from the JUMP-Cell Painting Consortium (Joint Undertaking in Morphological Profiling), links morphological profiles to compound targets through paired genetic and chemical perturbations [37]. Self-supervised representations outperform hand-crafted features [38], but they also encode technical factors such as batch and well position [39].

The recent State model changes the unit of modelling from the single cell to a set of cells [40]. It has a state embedding model and a state transition model, and trains on groups of cells in a shared embedding space, which separates perturbation effects from the heterogeneity within a population that is easily mistaken for noise. On large datasets, its discriminative performance exceeds the baseline by more than 30%, and because its cell embeddings are trained on observational data from 167 million cells, it can identify strong perturbations in cellular contexts where it saw no perturbation during training [40]. Generalisation across contexts therefore came from matching the unit of modelling to the level at which biological variation occurs, with model capacity a secondary factor.

The perturbation effect of a TF usually appears first as a shift in the expression profile and only later as a change in morphology. A diffusion model from the same period, conditioned on a transcriptomic representation, predicts post-perturbation changes in cell morphology [41]. The two

DOI: 10.54254/2755-2721/2026.LD39440

read-outs are two views of one biological process and each can validate the other, so agreement between them is a usable constraint on the credibility of a screen, weaker than two independent lines of evidence would be.

2.4. Intervention molecule discovery through generative design and complex modelling

Complex structure prediction and target-aware molecular generation were rewritten in parallel. Among complex predictors, EquiBind predicted the binding pose directly, TANKBind encoded physical constraints explicitly, and DiffDock moved docking from regression to generative modelling, reaching a top-1 success rate of 38% on PDBBind against 23% for conventional docking [42-44], DynamicBind went further and returned a receptor conformation specific to the ligand [45]. Among generative methods, Pocket2Mol samples atom by atom conditioned on the pocket, TargetDiff removed the dependence on generation order with an equivariant diffusion model (EDM), and DiffSBDD gave structure-based drug design a unified formulation [46-48].

AlphaFold3 denoises an atom cloud directly with a diffusion module and clearly surpasses the specialist tools of its time at the protein–ligand, protein–nucleic acid and antibody–antigen interfaces [49]. Its use on TFs has a clear limit, because its output in disordered regions carries no experimental constraint and can only be treated as a hypothesis. Structure quality then propagates downstream: when computationally folded structures replace experimental ones, the success rate of the earlier methods falls to at most 10.4%, while generative docking keeps 21.7% [44]. TFs are short of high-resolution experimental structures and highly dynamic at once, so low structural confidence depresses downstream performance across the board, and generative methods lose the least to it.

DiffGui, published in 2025, takes up the problem that generated molecules are often not viable. It places atoms and chemical bonds in a single diffusion process and generates them jointly, and it constrains sampling with binding affinity and drug-likeness through classifier-free guidance [50]. On PDBbind the Jensen–Shannon divergence between its carbon–carbon bond length distribution and the true distribution is 0.1815, below 0.2265 for TargetDiff and 0.2903 for DiffSBDD, and the work includes wet-lab validation [50].

MolGenBench, from the same year, measures generative design against activity data [51]. Built on experimentally validated activity data for 120 protein targets and 600 compound series, it covers both de novo generation and lead optimisation and evaluates 17 mainstream methods under a single protocol. Virtual-screening performance is weak, target awareness is poor, and generalisation to unseen proteins drops, some models improve three-dimensional conformational quality and still fail to produce active molecules reliably [51]. The metrics themselves are part of the problem. Drug-likeness and docking scores correlate strongly with atom count, and an AddCarbon model that merely inserts random carbon atoms into training-set molecules can approach full marks on standard metrics and outperform sophisticated generative models. Such metrics track molecular size, not activity potential, which is a failure of construct validity: the geometric distribution metrics reported by work such as DiffGui are real, but nothing in them assures that active molecules will follow.

2.5. Translational progress and clinical evidence

Clinical evidence for TF drugging is concentrated in three routes, none of which acts directly on the DNA-binding domain (DBD). The most complete translational chain belongs to revumenib (SNDX-5613), an inhibitor of the menin–KMT2A interaction. Its phase I trial (NCT04065399) gave a 30% rate of complete remission or complete remission with partial haematological recovery (CR/CRh) in relapsed or refractory acute leukaemia. The differentiation syndrome seen during treatment appeared

DOI: 10.54254/2755-2721/2026.LD39440

together with clearance of residual disease and with markers of differentiation towards normal haematopoietic cells, which is pharmacodynamic confirmation that the differentiation block had been released [52].

The registrational phase II portion of the AUGMENT-101 trial, reported in 2025, found a CR+CRh rate of 22.8% in 57 patients who had failed multiple lines of therapy (P = 0.0036, exceeding the 10% null hypothesis set in the protocol), an overall response rate of 63.2%, and no detectable residual disease in 15 of 22 patients evaluable for measurable residual disease (MRD) [53]. Intervention at a cofactor interface is therefore feasible in patients, which bypasses the absence of a pocket on the TF itself and moves the druggability problem onto a protein whose structure can be resolved.

In the same year, vepdegestrant (ARV-471), a proteolysis-targeting chimera (PROTAC) directed at the oestrogen receptor, completed phase III and was submitted for a new drug application [54], which validated the PROTAC route for TFs, like the menin inhibitor, however, it came out of conventional medicinal chemistry and structural biology. The only known computation-driven case to date is an allosteric covalent inhibitor found by virtual screening at the junction between the coiled-coil and DNA-binding domains of STAT3, after structure-guided optimisation, it reached only low-micromolar potency, an order of magnitude short of the routes already in the clinic, and has not become a candidate [55].

2.6. Evaluation of evidence, main challenges and technical responses

Whether the progress at each tier can be trusted depends first on whether the evaluation is matched to the task. A systematic assessment that trained 62,820 models showed that even in molecular property prediction, the most thoroughly studied area, representation-learning models perform modestly on most datasets, and that differences in how the data are split are enough to make comparisons unfair [56]. Comparability of imaging read-outs depends on batch correction [57]. On real large-scale single-cell perturbation data, network inference that uses interventional information does no better than inference from observational data alone, which reverses the conclusion drawn on synthetic benchmarks [58].

A 2025 evaluation put numbers on this. Trained on CRISPR activation data in K562 cells and tested on 62 double-gene perturbations, five single-cell foundation models together with GEARS and CPA all gave prediction errors well above an additive baseline that simply sums the log fold changes of the two single perturbations, and none beat a no-change baseline that makes no prediction at all [59]. scGPT is among the models that explicitly claim to predict expression changes caused by genetic perturbation, so trading large-scale pretraining for a general representation did not pay off on this task [35]. Seven models failing together implicates the common practice rather than any one architecture: models are compared against each other and almost never against a trivial baseline. Without an additive baseline, there is no way to tell whether a model has learned how two genes interact or has only reproduced the sum of the two single-perturbation effects.

DataSAIL, proposed in the same year, locates a cause further upstream. Information leakage usually takes the form of a test set that resembles the training set too closely and the deployment data too little, rather than of duplicated samples, a model can then score highly by following a similarity shortcut, and the shortcut fails out of distribution [60]. The work formulates leakage-reducing partition as a solvable optimisation problem and argues that the partitioning scheme itself sets how far a metric is inflated, so the partitioning criterion has to be reported alongside the metric [60]. The stratification in MolGenBench by whether a target protein appears in the training set is the same problem in chemical form. Missing baselines and information leakage together make

DOI: 10.54254/2755-2721/2026.LD39440

retrospective evaluation systematically optimistic, which leaves prospective experiments as the only way to settle what machine learning contributes.

Among current technical responses, only protein conformational ensembles reach the bottleneck of chemical tractability. Once prediction of a single conformation was largely solved, attention moved to generating conformational distributions through multiple sequence alignment (MSA) perturbation, score-based generative models or protein language models [61, 62], which answer both the need for cryptic pocket prediction and the need for selective targeting of disordered regions [26, 27]. Multi-omic integration can bring evidence from several tiers together in a specific cellular context [63, 64], but what it yields is still target validity. Transfer learning, large-scale pretraining and knowledge graphs remain only as reliable as the coverage of the available data allows [65-68], as the baseline comparison above shows. Adding model capacity on the target-identification side will not narrow the gap at the clinical end.

3. Discussion

Confusion between types of evidence runs through all tiers. Regulatory network inference yields target validity, while whether a target can be modulated by a small molecule with sufficient potency and selectivity has to be answered separately by structure, phenotype and chemical perturbation. Molecular generation splits the same way: geometric plausibility keeps improving, yet a unified benchmark shows that such metrics do not track the ability to produce active molecules. As Section 2.6 shows, missing trivial baselines and leakage between training and test sets make retrospective evaluation systematically optimistic, so the real contribution of a model can only be settled by prospective experiment.

Model capability is probably not what produces the gap along the axis, the likelier cause is the decreasing supply of verifiable labels at successive tiers. At the front of the axis there are accepted standards that support large-scale supervision, and sequence, accessibility and expression can be measured at genome scale and compared across laboratories. By the middle, labels already carry batch and covariate effects, which weakens the advantage of complex models: a representation that scores well on one screen may be fitting well position as much as biology. At the chemical and clinical end, labels are scarce and hard to reproduce or retrace, since activity data come from heterogeneous assays and a clinical outcome cannot be repeated at will. The density of machine learning involvement therefore tracks what each tier can be made to measure, and which testable labels a tier can provide deserves consideration before any further iteration of models and algorithms.

4. Conclusion

Machine learning involvement decreases tier by tier along the evidence axis and approaches zero at clinical validation. The upstream gains are real. Binding, accessibility and expression can now be scored within one model, ligandable sites have widened to include transient conformational states, and perturbation response is modelled at the level of cell populations. Downstream the evidence thins, and no clinical route to date can be traced to substantive machine learning input. This pattern follows mainly from how many verifiable labels each tier can supply. Two practices follow from it: ask what testable label a tier can supply before choosing a model, and report trivial baselines and the data-partitioning criterion alongside every metric. For TFs, the most practical near-term task may be to ask whether a given factor carries a druggable bypass dependency, which is a more efficient route to intervention than predicting binding sites in its disordered region.

DOI: 10.54254/2755-2721/2026.LD39440

References

[1] Lambert, S. A., Jolma, A., Campitelli, L. F., et al. (2018). The human transcription factors. *Cell, 172*(4), 650–665.

[2] Henley, M. J., & Koehler, A. N. (2021). Advances in targeting "undruggable" transcription factors with small molecules. *Nature Reviews Drug Discovery, 20*(9), 669–688.

[3] Lee, T. I., & Young, R. A. (2013). Transcriptional regulation and its misregulation in disease. *Cell, 152*(6), 1237–1251.

[4] Boija, A., Klein, I. A., Sabari, B. R., et al. (2018). Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell, 175*(7), 1842–1855.

[5] Bushweller, J. H. (2019). Targeting transcription factors in cancer — from undruggable to reality. *Nature Reviews Cancer, 19*(11), 611–624.

[6] Chen, H., Fu, R., & Wu, Z. Q. (2026). Transcription factor targeting strategies in cancer: Mechanisms, challenges and cutting-edge progress. *Acta Pharmacologica Sinica*. DOI: 10.1038/s41401-026-01877-8.

[7] Huang, K., Fu, T., Gao, W., et al. (2022). Artificial intelligence foundation for therapeutic science. *Nature Chemical Biology, 18*(10), 1033–1036.

[8] Alipanahi, B., Delong, A., Weirauch, M. T., et al. (2015). Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nature Biotechnology, 33*(8), 831–838.

[9] Kelley, D. R., Reshef, Y. A., Belanger, D., et al. (2018). Sequential regulatory activity prediction across chromosomes with convolutional neural networks. *Genome Research, 28*(5), 739–750.

[10] Avsec, Ž., Agarwal, V., Visentin, D., et al. (2021). Effective gene expression prediction from sequence by integrating long-range interactions. *Nature Methods, 18*(10), 1196–1203.

[11] Huynh-Thu, V. A., Irrthum, A., Wehenkel, L., et al. (2010). Inferring regulatory networks from expression data using tree-based methods. *PLoS ONE, 5*(9), e12776.

[12] Moerman, T., Aibar Santos, S., Bravo González-Blas, C., et al. (2019). GRNBoost2 and Arboreto: Efficient and scalable inference of gene regulatory networks. *Bioinformatics, 35*(12), 2159–2161.

[13] Aibar, S., González-Blas, C. B., Moerman, T., et al. (2017). SCENIC: Single-cell regulatory network inference and clustering. *Nature Methods, 14*(11), 1083–1086.

[14] Bravo González-Blas, C., De Winter, S., Hulselmans, G., et al. (2023). SCENIC+: Single-cell multiomic inference of enhancers and gene regulatory networks. *Nature Methods, 20*(9), 1355–1367.

[15] Kamimoto, K., Stringa, B., Hoffmann, C. M., et al. (2023). Dissecting cell identity via network inference and in silico gene perturbation. *Nature, 614*(7949), 742–751.

[16] Garcia-Alonso, L., Holland, C. H., Ibrahim, M. M., et al. (2019). Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Research, 29*(8), 1363–1375.

[17] Badia-i-Mompel, P., Vélez Santiago, J., Braunger, J., et al. (2022). decoupleR: Ensemble of computational methods to infer biological activities from omics data. *Bioinformatics Advances, 2*(1), vbac016.

[18] Włodarczyk, T., Lun, A., Wu, D., et al. (2025). Epiregulon: Single-cell transcription factor activity inference to predict drug response and drivers of cell states. *Nature Communications, 16*, 7118.

[19] Avsec, Ž., Latysheva, N., Cheng, J., et al. (2026). Advancing regulatory variant effect prediction with AlphaGenome. *Nature, 649*(8099), 1206–1218.

[20] Schmidtke, P., & Barril, X. (2010). Understanding and predicting druggability: A high-throughput method for detection of drug binding sites. *Journal of Medicinal Chemistry, 53*(15), 5858–5867.

[21] Krivák, R., & Hoksza, D. (2018). P2Rank: Machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. *Journal of Cheminformatics, 10*, 39.

[22] Jiménez, J., Doerr, S., Martínez-Rosell, G., et al. (2017). DeepSite: Protein-binding site predictor using 3D-convolutional neural networks. *Bioinformatics, 33*(19), 3036–3042.

[23] Aggarwal, R., Gupta, A., Chelur, V., et al. (2022). DeepPocket: Ligand binding site detection and segmentation using 3D convolutional neural networks. *Journal of Chemical Information and Modeling, 62*(21), 5069–5079.

[24] Gainza, P., Sverrisson, F., Monti, F., et al. (2020). Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nature Methods, 17*(2), 184–192.

[25] Tubiana, J., Schneidman-Duhovny, D., & Wolfson, H. J. (2022). ScanNet: An interpretable geometric deep learning model for structure-based protein binding site prediction. *Nature Methods, 19*(6), 730–739.

[26] Meller, A., Ward, M., Borowsky, J., et al. (2023). Predicting locations of cryptic pockets from single protein structures using the PocketMiner graph neural network. *Nature Communications, 14*, 1177.

[27] Bielskutė-García, S., Mateos, B., Awawdy, M., et al. (2026). Oligomerization enables the selective targeting of an intrinsically disordered region by a small molecule. *Science Advances, 12*, eadz7400.

DOI: 10.54254/2755-2721/2026.LD39440

[28] Subramanian, A., Narayan, R., Corsello, S. M., et al. (2017). A next generation connectivity map: L1000 platform and the first 1, 000, 000 profiles. *Cell, 171*(6), 1437–1452.
[29] Dixit, A., Parnas, O., Li, B., et al. (2016). Perturb-seq: Dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell, 167*(7), 1853–1866.
[30] Replogle, J. M., Saunders, R. A., Pogson, A. N., et al. (2022). Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell, 185*(14), 2559–2575.
[31] Lotfollahi, M., Wolf, F. A., & Theis, F. J. (2019). scGen predicts single-cell perturbation responses. *Nature Methods, 16*(8), 715–721.
[32] Lotfollahi, M., Klimovskaia Susmelj, A., De Donno, C., et al. (2023). Predicting cellular responses to complex perturbations in high-throughput screens. *Molecular Systems Biology, 19*, e11517.
[33] Roohani, Y., Huang, K., & Leskovec, J. (2024). Predicting transcriptional outcomes of novel multigene perturbations with GEARS. *Nature Biotechnology, 42*(6), 927–935.
[34] Bunne, C., Stark, S. G., Gut, G., et al. (2023). Learning single-cell perturbation responses using neural optimal transport. *Nature Methods, 20*(11), 1759–1768.
[35] Cui, H., Wang, C., Maan, H., et al. (2024). scGPT: Toward building a foundation model for single-cell multi-omics using generative AI. *Nature Methods, 21*(8), 1470–1480.
[36] Bray, M. A., Singh, S., Han, H., et al. (2016). Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nature Protocols, 11*(9), 1757–1774.
[37] Chandrasekaran, S. N., Cimini, B. A., Goodale, A., et al. (2024). Three million images and morphological profiles of cells treated with matched chemical and genetic perturbations. *Nature Methods, 21*(6), 1114–1121.
[38] Kim, V., Adaloglou, N., Osterland, M., et al. (2023). Self-supervision advances morphological profiling by unlocking powerful image representations. *Scientific Reports, 13*, 21921.
[39] Moshkov, N., Bornholdt, M., Benoit, S., et al. (2024). Learning representations for image-based profiling of perturbations. *Nature Communications, 15*, 1594.
[40] Adduri, A. K., Gautam, D., Bevilacqua, B., et al. (2026). Predicting cellular responses to perturbation across diverse contexts with State. *Cell, 189*, 5914–5931.
[41] Wang, X., Fan, Y., Guo, Y., et al. (2025). Prediction of cellular morphology changes under perturbations with a transcriptome-guided diffusion model. *Nature Communications, 16*, 8172.
[42] Stärk, H., Ganea, O. E., Pattanaik, L., et al. (2022). EquiBind: Geometric deep learning for drug binding structure prediction. In *Proceedings of the 39th International Conference on Machine Learning* (pp. 20503–20521). PMLR.
[43] Lu, W., Wu, Q., Zhang, J., et al. (2022). TANKBind: Trigonometry-aware neural networks for drug-protein binding structure prediction. In *Advances in Neural Information Processing Systems 35* (pp. 7236–7249).
[44] Corso, G., Stärk, H., Jing, B., et al. (2023). DiffDock: Diffusion steps, twists, and turns for molecular docking. In *The Eleventh International Conference on Learning Representations*.
[45] Lu, W., Zhang, J., Huang, W., et al. (2024). DynamicBind: Predicting ligand-specific protein-ligand complex structure with a deep equivariant generative model. *Nature Communications, 15*, 1071.
[46] Peng, X., Luo, S., Guan, J., et al. (2022). Pocket2Mol: Efficient molecular sampling based on 3D protein pockets. In *Proceedings of the 39th International Conference on Machine Learning* (pp. 17644–17655). PMLR.
[47] Guan, J., Qian, W. W., Peng, X., et al. (2023). 3D equivariant diffusion for target-aware molecule generation and affinity prediction. In *The Eleventh International Conference on Learning Representations*.
[48] Schneuing, A., Harris, C., Du, Y., et al. (2024). Structure-based drug design with equivariant diffusion models. *Nature Computational Science, 4*(12), 899–909.
[49] Abramson, J., Adler, J., Dunger, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature, 630*(8016), 493–500.
[50] Hu, Q., Sun, C., He, H., et al. (2025). Target-aware 3D molecular generation based on guided equivariant diffusion. *Nature Communications, 16*, 7928.
[51] Cao, D., Fan, Z., Yu, J., et al. (2026). Evaluating generative models for structure-based drug discovery from de novo design to lead optimization. *Nature Communications*. DOI: 10.1038/s41467-026-78105-8.
[52] Issa, G. C., Aldoss, I., DiPersio, J., et al. (2023). The menin inhibitor revumenib in KMT2A-rearranged or NPM1-mutant leukaemia. *Nature, 615*(7954), 920–924.
[53] Issa, G. C., Aldoss, I., Thirman, M. J., et al. (2025). Menin inhibition with revumenib for KMT2A-rearranged relapsed or refractory acute leukemia (AUGMENT-101). *Journal of Clinical Oncology, 43*(1), 75–84.
[54] Ma, Z., & Zhou, J. (2025). NDA submission of vepdegestrant (ARV-471) to U.S. FDA: The beginning of a new era of PROTAC degraders. *Journal of Medicinal Chemistry, 68*(14), 14129–14136.
[55] Szalai, T. V., Di Lorenzo, V., Péczka, N., et al. (2025). Allosteric covalent inhibitors of the STAT3 transcription factor from virtual screening. *ACS Medicinal Chemistry Letters, 16*(6), 991-997.

DOI: 10.54254/2755-2721/2026.LD39440

[56] Deng, J., Yang, Z., Wang, H., et al. (2023). A systematic study of key elements underlying molecular property prediction. *Nature Communications, 14*, 6395.
[57] Arevalo, J., Su, E., Ewald, J. D., et al. (2024). Evaluating batch correction methods for image-based cell profiling. *Nature Communications, 15*, 6516.
[58] Chevalley, M., Roohani, Y. H., Mehrjou, A., et al. (2025). A large-scale benchmark for network inference from single-cell perturbation data. *Communications Biology, 8*, 412.
[59] Ahlmann-Eltze, C., Huber, W., & Anders, S. (2025). Deep-learning-based gene perturbation effect prediction does not yet outperform simple linear baselines. *Nature Methods, 22*(8), 1657-1661.
[60] Joeres, R., Blumenthal, D. B., & Kalinina, O. V. (2025). Data splitting to avoid information leakage with DataSAIL. *Nature Communications, 16*, 3337.
[61] Bowman, G. R. (2024). AlphaFold and protein folding: Not dead yet! The frontier is conformational ensembles. *Annual Review of Biomedical Data Science, 7*, 51–57.
[62] Sledzieski, S., & Hanson, S. M. (2025). The landscape of machine learning approaches for modeling protein conformational ensembles. *Current Opinion in Structural Biology*.
[63] Argelaguet, R., Arnol, D., Bredikhin, D., et al. (2020). MOFA+: A statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biology, 21*, 111.
[64] Cao, Z. J., & Gao, G. (2022). Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nature Biotechnology, 40*(10), 1458–1466.
[65] Theodoris, C. V., Xiao, L., Chopra, A., et al. (2023). Transfer learning enables predictions in network biology. *Nature, 618*, 616–624.
[66] Hao, M., Gong, J., Zeng, X., et al. (2024). Large-scale foundation model on single-cell transcriptomics. *Nature Methods, 21*(8), 1481–1491.
[67] Yu, H., Qian, W., Song, Y., et al. (2025). PerturbNet predicts single-cell responses to unseen chemical and genetic perturbations. *Molecular Systems Biology, 21*(8), 960–982.
[68] Wenkel, F., Tu, W., Masschelein, C., et al. (2026). TxPert: Using multiple knowledge graphs for prediction of transcriptomic perturbation effects. *Nature Biotechnology*. DOI: 10.1038/s41587-026-03113-4.