MEDS Clinical Medicine (2026)
Clausius Scientific Press, Canada

DOI: 10.23977/medsc.2026.070315
ISSN 2616-1907 Vol. 7 Num. 3

AI-Assisted Gene Expression Signature Reversal for Candidate Prioritization in Drug Repurposing

Yi Wan

Faculty of Science, University of Technology Sydney, Sydney, 2007, Australia

Keywords: Drug repurposing; Gene expression signature reversal; Connectivity Map; Artificial intelligence; Candidate prioritization

Abstract: Drug repurposing increasingly relies on transcriptomic evidence to identify new therapeutic uses for existing drugs or drug-like compounds. This review examines how AI-assisted gene expression signature reversal contributes to candidate prioritization in drug repurposing. The approach compares disease-associated gene expression signatures with drug-induced transcriptional profiles, using resources such as the Connectivity Map and LINCS as the data foundation. Signature reversal is valuable because it offers a biologically interpretable ranking logic: compounds that oppose disease-associated expression changes may be prioritized for further study. The reviewed literature suggests that this logic is useful, but uneven. Some studies provide evidence that reversal scores can relate to drug efficacy, whereas validation-focused work also shows that reversal signals may be confounded by broad anti-proliferative effects. AI and machine learning extend this framework by supporting high-dimensional expression analysis, transcriptional response prediction, expression ranking, and context-aware prioritization. Recent deep learning approaches highlight the importance of dose and cellular context when predicting drug-induced responses. Overall, AI-assisted signature reversal is most defensible as a candidate triage strategy rather than as evidence of efficacy. Its value depends on combining reversal scores with pathway interpretation, biological context, and targeted follow-up testing.

1. Introduction

Traditional drug discovery is a lengthy, costly, and uncertain process. A new drug usually moves through target identification, laboratory testing, preclinical studies, and several phases of clinical trials before approval. Even after substantial investment, many candidates fail because they do not show sufficient efficacy, raise safety concerns, or perform poorly when translated from experimental systems into clinical use. Drug repurposing responds to this problem by asking a different question: whether existing drugs or drug-like compounds with available biological, pharmacological, or clinical information can be redirected toward new therapeutic indications [1, 2]. Its attraction is not only speed. More importantly, repurposing allows researchers to begin with compounds for which some prior evidence already exists.

Transcriptomic resources have made this strategy more systematic. Many diseases are accompanied by abnormal gene expression patterns, and drug treatment can also produce measurable transcriptional changes. Gene expression signature reversal builds on this relationship. If a compound

induces an expression profile that opposes a disease-associated signature, it may deserve priority as a repurposing candidate [3, 4]. The Connectivity Map and LINCS provide the key reference space for this comparison by linking disease signatures with drug-induced perturbation profiles [5, 6]. L1000CDS ²further translates this idea into a searchable tool by identifying compounds predicted to mimic or reverse a submitted gene expression signature [7].

AI and machine learning add a second layer to this framework. Instead of relying only on direct database matching, these methods can help process high-dimensional expression data, predict drug-induced transcriptional responses, rank expression patterns, and prioritize compounds when the candidate space is large [8-11]. The most useful contribution of AI is therefore not simply automation. It can change how researchers move from noisy transcriptomic data to a smaller, more interpretable candidate list. Yet this advantage introduces a harder interpretive problem: a computationally strong reversal pattern is not necessarily a disease-specific therapeutic signal [12].

This review asks the following research question: How can AI-assisted gene expression signature reversal improve candidate drug prioritization in drug repurposing? To answer this question, the paper first explains the theoretical basis of drug repurposing, signature reversal, and CMap/LINCS-based comparison. It then reviews studies on CMap/LINCS resources, signature reversal methods, and AI-assisted prioritization. The final discussion evaluates where the combined approach is strongest, where it is most vulnerable, and how it should be interpreted before follow-up testing.

2. Theoretical Background

Drug repurposing refers to identifying new therapeutic uses for existing drugs or drug-like compounds. Compared with conventional drug discovery, it often starts from compounds with some known biological, pharmacological, or clinical properties. This can narrow the search space before further testing, rather than beginning entirely from new chemical entities [1, 2]. In this review, drug repurposing is not treated as a broad pharmaceutical business strategy. It is treated more narrowly as the endpoint of a computational process: selecting and ranking existing compounds that may be relevant in a new disease context.

Gene expression signature reversal provides the theoretical logic for that process. A gene expression signature is a pattern of genes that are upregulated or downregulated under a particular biological condition. In disease research, a disease-associated signature is usually derived by comparing disease samples with normal or control samples. Such a signature does not represent the whole disease, but it can capture part of the molecular state associated with that condition. Drug treatment, in turn, can produce measurable transcriptional changes known as drug-induced expression profiles [5, 6].

The reversal idea is straightforward but powerful. If a disease increases the expression of one group of genes and decreases another, then a candidate drug may be interesting if it pushes those changes in the opposite direction. This inverse relationship gives signature reversal its appeal as a prioritization method [3, 4]. It is also where the method becomes vulnerable. A strong inverse score can indicate an expression-level opposition, but it does not by itself explain why a compound should work therapeutically. The score needs to be read as a ranking signal, not as a mechanistic conclusion [12].

Connectivity Map and LINCS make this type of reasoning operational. They contain large collections of perturbation-induced expression profiles generated under different experimental conditions. By comparing a disease signature with these profiles, researchers can estimate whether a compound is positively or negatively connected with a disease state [5, 6]. For drug repurposing, negative connectivity is usually more informative because it suggests that the compound may oppose disease-related expression changes. L1000CDS ²provides a more direct query-based implementation

of this logic by identifying compounds predicted to mimic or reverse submitted signatures [7].

Within this framework, AI and machine learning should be viewed as computational amplifiers rather than replacements for biological reasoning. They can help identify informative expression patterns, reduce the burden of high-dimensional data, predict drug-induced transcriptional responses, and rank candidate compounds [8, 9, 10, 13]. Newer models also address dose-dependent and context-specific responses, which matters because drug effects are not fixed properties of compounds alone; they vary with cell type, tissue state, exposure level, and phenotype readout [11]. The theoretical contribution of AI-assisted signature reversal is therefore sharper than a general claim about "AI in drug discovery": it is a way to improve candidate ranking when transcriptomic evidence is available but difficult to interpret.

3. Literature Review

3.1. CMap / LINCS-Based Drug Repurposing Studies

CMap and LINCS-based approaches have become central to gene expression-based drug repurposing because they provide a shared computational space for comparing disease-associated signatures with drug-induced profiles [5, 6, 14]. This is a different logic from classical target-based screening. Rather than asking whether a compound binds to one predefined target, the method asks whether a compound shifts a broader transcriptional state in a direction that is relevant to disease. That distinction matters most for complex diseases, where pathological states often involve coordinated changes across genes, pathways, and regulatory networks rather than a single dominant molecular lesion [2, 14].

The original Connectivity Map introduced the idea that genes, small molecules, and disease states could be connected through expression signatures [5]. Its most important contribution was conceptual: it made transcriptional profiles usable as query objects. The next-generation CMap and L1000 platform expanded that idea in scale, allowing many more compounds and experimental conditions to be examined computationally [6]. As a result, gene expression-based repurposing became less dependent on isolated case-by-case comparisons and more suited to systematic candidate screening.

In most CMap/LINCS-based studies, the workflow begins with a disease signature. Researchers identify genes that are increased or decreased in disease samples relative to normal or control samples, then compare that signature with drug-induced profiles. A positive connection may indicate that a drug produces a disease-like transcriptional state, whereas a negative connection suggests an opposing state [14, 15]. For repurposing, the second relationship is usually more useful because it offers a route for identifying compounds that may counteract disease-associated molecular changes.

The literature shows that CMap/LINCS-based strategies can support candidate discovery across disease areas, including oncology and viral infection [1, 2, 15]. These studies should not be read as evidence that database matching alone can establish efficacy. Their stronger contribution is more practical: they show how transcriptomic comparison can reduce a broad and disordered compound space into a smaller, biologically interpretable set of candidates. For this review, CMap and LINCS therefore function mainly as infrastructure. They supply perturbation profiles; the critical scientific question is how those profiles are scored, ranked, and interpreted.

3.2. Signature Reversal Methods in Drug Repurposing

Signature reversal is the most direct bridge between disease biology and candidate prioritization in this review. It assumes that a disease state can be partly represented by an abnormal expression signature and that a useful compound may produce an opposite transcriptional effect. In practical terms, genes increased in the disease may be reduced after drug treatment, while genes reduced in the

disease may be increased. This inverse relationship gives the method a clear biological interpretation and explains why it has become attractive for repurposing studies [3, 4].

A typical reversal workflow has three linked steps. First, a disease-associated signature is generated from transcriptomic data. Second, this signature is compared with drug-induced profiles from databases such as CMap or LINCS. Third, compounds are ranked according to the strength of the inverse relationship [7, 15]. Negative connectivity scores or related reversal scores are often used to express this relationship. A stronger negative score suggests that the drug-induced profile is more opposed to the disease-associated profile and may therefore deserve higher priority for follow-up analysis [4, 7, 15].

The method is valuable partly because it is not restricted to drugs with known mechanisms in the target disease. By comparing transcriptional patterns, signature reversal can identify unexpected relationships between existing compounds and disease states. This feature is especially relevant to repurposing, where the strongest candidate may not be the compound with the most obvious original indication [2, 3]. L1000CDS ² makes this logic practical by allowing a user-submitted expression signature to be queried for small molecules predicted to mimic or reverse it [7].

Evidence for the method, however, is not all of the same strength. Studies linking reversal of cancer gene expression to drug efficacy provide important support because they connect the score to biological activity rather than leaving it as a purely computational output [4]. At the same time, validation-focused work on transcriptome signature reversion in oncology gives one of the clearest cautions in the literature: reversal scores may partly capture general anti-proliferative effects rather than disease-specific therapeutic activity [12]. This is a more serious limitation than a generic need for “more validation.” It means that a compound can look promising because it broadly suppresses proliferating cells, not because it targets the disease mechanism under study.

For that reason, reversal-based ranking should be interpreted as a disciplined starting point, not a finished answer. The quality of the disease signature, the scoring design, the cell line used to generate the drug profile, and the dose and time of treatment can all shape the final ranking [6, 12, 14]. The method remains useful, but its usefulness depends on interpretation. A high-ranked candidate is more convincing when the reversal score is supported by pathway-level relevance, disease biology, and a plausible drug-response pattern.

3.3. AI and Machine Learning in Candidate Prioritization

AI and machine learning enter this field because transcriptomic repurposing creates more information than researchers can interpret manually. Gene expression datasets may contain thousands of measured or inferred features, while sample sizes are often much smaller. Meaningful signals can be mixed with technical noise, redundant genes, and context-specific variation. Machine learning methods can help identify informative patterns, reduce dimensional complexity, and support more systematic comparison between disease states and drug-induced profiles [8, 13].

The most immediate use of AI is candidate ranking. Computational screens rarely return one obvious compound; they produce lists that need to be ordered. Machine learning-based approaches have been used to generate ranked lists of possible repurposing candidates from disease-related expression information [8]. Other approaches use disease clustering before reversal-based comparison, which shows that AI can contribute before the final ranking step by organizing disease states into more meaningful groups [13]. This is an important distinction: AI does not only score drugs; it can also structure the input space in which drug-disease comparisons are made.

Deep learning extends the framework further by predicting drug-induced transcriptional responses rather than relying only on profiles already measured in a database. DLEPS uses chemically induced transcriptional profiles from L1000 to predict drug efficacy from disease-related gene signatures and

selected compound libraries [9]. CIGER focuses on predicting chemical-induced gene expressions for de novo chemicals and applies that prediction to pancreatic cancer drug repurposing [10]. These studies are particularly relevant because they move prioritization beyond simple lookup. They suggest that AI can estimate how a compound may perturb expression even when direct experimental data are incomplete.

The most methodologically important development is context-aware prediction. MultiDCP predicts dose-dependent and context-specific drug-induced gene expression and cell viability, addressing the weakness of models that use a small number of fixed cell lines as substitutes for disease-relevant tissues or patient contexts [11]. This is not just technical refinement. For repurposing, dose and cellular context can change whether a candidate appears promising, ineffective, or unsafe. AI therefore contributes most when it helps prioritize compounds under biologically relevant conditions, rather than merely producing a longer ranked list.

Taken together, the literature supports a pipeline in which CMap/LINCS resources provide perturbation data, signature reversal supplies the inverse-matching logic, and AI/ML methods improve prediction, pattern recognition, and candidate ranking. The combined approach matters because it makes prioritization more explicit. It does not simply say that a compound is "related" to a disease; it asks how that relationship is represented in expression space, how strongly it opposes the disease state, and whether the ranking remains plausible under biological constraints.

4. Discussion & Synthesis

The reviewed literature supports a qualified but useful answer to the research question. AI-assisted gene expression signature reversal can improve candidate prioritization by linking three elements that are often treated separately: disease-associated transcriptional change, perturbation databases, and computational ranking. The strongest contribution is triage. Instead of beginning with a large and weakly organized compound pool, researchers can begin with candidates whose transcriptional effects have at least some relationships to the disease signature [3, 4].

Among the components of this framework, the data foundation is relatively mature. CMap and LINCS have established a practical reference space for perturbation-based comparison [5, 6], and L1000CDS [2] provides an accessible way to query mimic or reversal relationships [7]. The more difficult problem is not whether profiles can be compared, but what the comparison means. A negative connection is attractive because it is interpretable, but interpretability can also be misleading if the biological basis of the reversal is not examined.

The most serious methodological risk is the specificity of the reversal signal. Validation-focused work on transcriptome signature reversion suggests that some reversal scores in oncology may reflect broad anti-proliferative effects rather than selective targeting of disease-specific pathways [12]. This point changes how the method should be used. The task is not simply to find the most negative score. A stronger workflow would ask whether the reversed genes are connected to disease-relevant pathways, whether the drug response is plausible at a usable dose, and whether the signal persists in an appropriate cellular context.

AI can improve this workflow, but only when its role is specified carefully. The strongest AI contributions in the reviewed literature are not vague claims about faster drug discovery. They are more concrete: ranking candidates from disease-related expression information [8], organizing disease groups before reversal analysis [13], predicting transcriptional responses from chemical information [9, 10], and modeling dose- and context-dependent responses [11]. These functions address real bottlenecks in transcriptomic repurposing. They also make the pipeline more complex, because predicted profiles and rankings must be interpreted rather than accepted automatically.

Dose and cellular context deserve particular attention. A compound that reverses a signature in

one cell type, at one concentration, and under one treatment duration may not behave the same way in the tissue or patient group of interest. MultiDCP is important in this discussion because it treats drug response as context-dependent rather than assuming that a small set of cell lines can stand in for all disease settings [11]. For candidate prioritization, this may be more consequential than small improvements in ranking accuracy. A biologically relevant prediction is more useful than a mathematically strong ranking produced under inappropriate conditions.

The overall implication is that AI-assisted signature reversal should be developed as an evidence-weighting workflow. A candidate should become more convincing when several layers point in the same direction: inverse expression matching, pathway-level relevance, appropriate cellular context, plausible dose-response behavior, and interpretable model output. If these layers disagree, the ranking should be treated cautiously even when the reversal score is strong. This layered interpretation is where the approach has the greatest value for drug repurposing. It does not remove uncertainty, but it makes the uncertainty more visible and easier to manage.

5. Conclusion

AI-assisted gene expression signature reversal offers a focused strategy for prioritizing drug repurposing candidates. Its value comes from combining transcriptomic databases, inverse expression matching, and computational ranking. CMap and LINCS provide perturbation profiles, L1000CDS [2] helps operationalize mimic and reversal searches, and AI/ML methods extend the framework through prediction, ranking, and context-aware analysis.

The review supports a cautious conclusion. Signature reversal is useful when it is treated as a ranking logic anchored in disease biology, not as proof that a compound will be effective. Evidence linking reversal to drug efficacy strengthens the method, while validation-focused work shows that reversal scores can also be influenced by non-specific anti-proliferative effects. The most defensible use of the approach is therefore not to identify a final therapeutic answer, but to select candidates that deserve closer biological and experimental examination.

Future work should focus less on producing longer candidate lists and more on improving the quality of prioritization. The most promising direction is an integrated workflow that combines reversal scores with pathway interpretation, disease-relevant cell or tissue context, dose-response information, and transparent model reasoning. Used in this way, AI-assisted signature reversal can make early-stage drug repurposing more selective, interpretable, and biologically grounded.

References

[1] Zhao, Y., Chen, X., Chen, J., & Qi, X. (2023). Decoding Connectivity Map-based drug repurposing for oncotherapy. Briefings in Bioinformatics, 24(3), bbad142. https://doi.org/10.1093/bib/bbad142

[2] Kwon, O. S., Kim, W., Cha, H. J., & Lee, H. (2019). In silico drug repositioning: From large-scale transcriptome data to therapeutics. Archives of Pharmacal Research, 42(10), 879–889. https://doi.org/10.1007/s12272-019-01176-3

[3] Fisher, J. L., Wilk, E. J., Oza, V. H., Gary, S. E., Howton, T. C., Flanary, V. L., Clark, A. D., Hjelmeland, A. B., & Lasseigne, B. N. (2024). Signature reversion of three disease-associated gene signatures prioritizes cancer drug repurposing candidates. FEBS Open Bio, 14(5), 803–830. https://doi.org/10.1002/2211-5463.13796

[4] Chen, B., Ma, L., Paik, H., Sirota, M., Wei, W., Chua, M. S., So, S., & Butte, A. J. (2017). Reversal of cancer gene expression correlates with drug efficacy and reveals therapeutic targets. Nature Communications, 8, 16022. https://doi.org/10.1038/ncomms16022

[5] Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., Lerner, J., Brunet, J. P., Subramanian, A., Ross, K. N., Reich, M., Hieronymus, H., Wei, G., Armstrong, S. A., Haggarty, S. J., Clemons, P. A., Wei, R., Carr, S. A., Lander, E. S., & Golub, T. R. (2006). The Connectivity Map: Using gene-expression signatures to connect small molecules, genes, and disease. Science, 313(5795), 1929–1935. https://doi.org/10.1126/science.1132939

[6] Subramanian, A., Narayan, R., Corsello, S. M., Peck, D. D., Natoli, T. E., Lu, X., Gould, J., Davis, J. F., Tubelli, A. A., Asiedu, J. K., Lahr, D. L., Hirschman, J. E., Liu, Z., Donahue, M., Julian, B., Khan, M., Wadden, D., Smith, I. C., Lam, D., ... Golub, T. R. (2017). A next generation Connectivity Map: L1000 platform and the first 1,000,000 profiles. Cell,

171(6), 1437–1452.e17. https://doi.org/10.1016/j.cell.2017.10.049
[7] Duan, Q., Reid, S. P., Clark, N. R., Wang, Z., Fernandez, N. F., Rouillard, A. D., Readhead, B., Tritsch, S. R., Hodos, R., Hafner, M., Niepel, M., Sorger, P. K., Dudley, J. T., Bavari, S., Panchal, R. G., & Ma'ayan, A. (2016). L1000CDS²: LINCS L1000 characteristic direction signatures search engine. npj Systems Biology and Applications, 2, 16015. https://doi.org/10.1038/npjsba.2016.15
[8] Rodriguez, S., Hug, C., Todorov, P., Moret, N., Boswell, S. A., Evans, K., Zhou, G., Johnson, N. T., Hyman, B. T., Sorger, P. K., Albers, M. W., & Sokolov, A. (2021). Machine learning identifies candidates for drug repurposing in Alzheimer's disease. Nature Communications, 12(1), 1033. https://doi.org/10.1038/s41467-021-21330-0
[9] Zhu, J., Wang, J., Wang, X., Gao, M., Guo, B., Gao, M., Liu, J., Yu, Y., Wang, L., Kong, W., An, Y., Liu, Z., Sun, X., Huang, Z., Zhou, H., Zhang, N., Zheng, R., & Xie, Z. (2021). Prediction of drug efficacy from transcriptional profiles with deep learning. Nature Biotechnology, 39(11), 1444–1452. https://doi.org/10.1038/s41587-021-00946-z
[10] Pham, T. H., Qiu, Y., Liu, J., Zimmer, S., O'Neill, E., Xie, L., & Zhang, P. (2022). Chemical-induced gene expression ranking and its application to pancreatic cancer drug repurposing. Patterns, 3(4), 100441. https://doi.org/10.1016/j.patter.2022.100441
[11] Wu, Y., Liu, Q., Qiu, Y., & Xie, L. (2022). Deep learning prediction of chemical-induced dose-dependent and context-specific multiplex phenotype responses and its application to personalized Alzheimer's disease drug repurposing. PLOS Computational Biology, 18(8), e1010367. https://doi.org/10.1371/journal.pcbi.1010367
[12] Koudijs, K. K. M., Böhringer, S., & Guchelaar, H. J. (2023). Validation of transcriptome signature reversion for drug repurposing in oncology. Briefings in Bioinformatics, 24(1), bbac490. https://doi.org/10.1093/bib/bbac490
[13] Cong, Y., Shintani, M., Imanari, F., Osada, N., & Endo, T. (2022). A new approach to drug repurposing with two-stage prediction, machine learning, and unsupervised clustering of gene expression. OMICS: A Journal of Integrative Biology, 26(6), 339–347. https://doi.org/10.1089/omi.2022.0026
[14] Musa, A., Ghoraie, L. S., Zhang, S. D., Glazko, G., Yli-Harja, O., Dehmer, M., Haibe-Kains, B., & Emmert-Streib, F. (2018). A review of Connectivity Map and computational approaches in pharmacogenomics. Briefings in Bioinformatics, 19(3), 506–523. https://doi.org/10.1093/bib/bbw112
[15] Mousavi, S. Z., Rahmanian, M., & Sami, A. (2020). A connectivity map-based drug repurposing study and integrative analysis of transcriptomic profiling of SARS-CoV-2 infection. Infection, Genetics and Evolution, 86, 104610. https://doi.org/10.1016/j.meegid.2020.104610