Structure learning without context-specific ground truths: a case study in chronic low-dose radiation exposure in human cells

Ashka Shah
Department of Computer Science
University of Chicago
shahashka@uchicago.edu

Rick Stevens
Department of Computer Science
University of Chicago
rstevens@cs.uchicago.edu

Abstract

Methods for gene regulatory network (GRN) inference often leverage structural knowledge from curated databases to constrain the expansive genome-sized graph space or as labels for training, however, this knowledge does not necessarily pertain to the specific context being studied – in this case chronic low-dose radiation exposure. We show how this mismatch between existing knowledge and the context under investigation makes it difficult to tune and evaluate estimated context-specific GRNs. We provide a dataset of RNA-seq gene expression of human cells grown in lab exposed to low-dose ionizing radiation, and compare several algorithms for estimating GRNs. We find that DAG-GNN, an unsupervised causal structure learning model, infers pathways that best align with existing literature. We also find that models that jointly learn from gene expression and radiation level data can directly estimate the genes most impacted by radiation, which greatly enhances downstream pathway analysis.

1 Introduction

Chronic exposure to low-dose radiation has implications for the onset of cancer and cardiovascular disease by activating response pathways in cells (Shimizu et al., 2010). Most research explores the affects of high dose, acute exposure to radiation; this results in high rates of double-stranded DNA breaks triggering programmed cell death or *apoptosis* (Verheij & Bartelink, 2000). However, a complete understanding of the interplay between dose rate, cumulative dose, and cell type in the activation of low-dose radiation pathways remains an open question (Fig. 1), and has implications for people chronically exposed to low-dose radiation (e.g., space exploration, radon exposure). A gene regulatory network (GRN) represents the mechanisms that govern gene expression levels, or the number of copies of RNA transcribed from DNA for a particular portion of DNA (i.e, a gene). Gene expression levels correlate to the number of proteins translated to carry out certain actions in the cell, including response to external stressors such as radiation. We explore the capabilities of causal structure learning and graph learning algorithms to infer GRNs for chronic, low-dose radiation exposure. One challenge is the lack of a context-specific ground truth network to evaluate, tune or constrain these algorithms with. We find that models trained to leverage edges from knowledge bases are not necessarily useful for understanding pathways specific to chronic, low-dose radiation response due to the macroscopic nature of these knowledge databases. Instead, models that do not incorporate prior knowledge are better at capturing dose-dependent mechanisms such as *cell cycle arrest* (a pause in cell division) and *DNA damage repair*. Our contributions and findings are as follows:

1. We provide an RNA-seq gene expression dataset specific to chronic exposure to low-dose radiation in Human Umbilical Vein Endothelial Cells (HUVECs) grown in lab at five different dose rates over 3 weeks.

Accepted to the NeurIPS 2025 Workshop on CauScien: Uncovering Causality in Science.

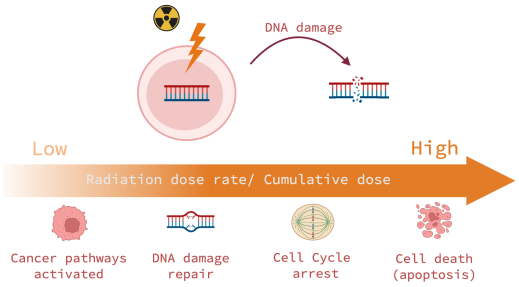

Figure 1: Exposure to ionizing radiation causes DNA breakage in cells. Cells activate pathways in order to repair the damage before the cell divides.

Dataset	Dose Rate (mGy/hr)	# Genes	# samples
A	0.001	918	13
B	0.01	1590	13
C	0.1	1487	13
D	1	1965	13
E	2	801	13

Table 1: A description of the five datasets A through E by dose rates.

2. We provide an evaluation of network recovery and pathway enrichment for six algorithms of interest with varying degrees of prior knowledge incorporation: GENIE3 (Marbach et al., 2012), GENELink (Chen & Liu, 2022), DAG-GNN (Yu et al., 2019), GES (Chickering, 2002), PC (Spirtes et al., 2000), and DirectLiNGAM (Shimizu et al., 2011).
3. We find that GENIE3, DAG-GNN and GES, which incorporate minimal prior knowledge and jointly model the radiation level, identify low-dose radiation pathways. DAG-GNN further identifies dose-dependent pathways related to DNA damage and cell cycle arrest.

2 Related Works

Chronic low-dose radiation exposure datasets High-dose, acute radiation response in mouse and human cell lines has been highly studied and documented in the RadBioDB database (Zanni et al., 2024). To our knowledge, we provide the first RNA-seq expression data for cells exposed to chronic (> 72 hours), low dose (< 100 mGy accumulated dose) radiation. Similar studies to ours include Babini et al. (2022) and Rombouts et al. (2014) who collected microarray expression data of HUVEC cell lines exposed to 1.4 - 4.1 mGy/hr of gamma radiation for up to 10 weeks.

GRN inference Yuan & Duren (2025) show how pretraining on existing RNA-seq expression data from the ENCODE database boosts the recovery of cell-type specific GRNs. Yao et al. (2015); Greenfield et al. (2013) incorporate structural edge priors into GRN inference to improve network recovery for yeast and bacteria GRNs. These works suggests that data and knowledge from other contexts can be used to improve learning of a context-specific GRN.

3 Methods

3.1 Bulk RNA-seq expression data in HUVECs

We collect bulk RNA-seq expression data for HUVECs grown in the lab exposed to low-dose ionizing radiation from a Cesium-137 gamma ray source at five different dose rates measured in miliGray (mGy) per hour [1]. For samples at each dose rate, measurements are made after one, two, and three weeks for both control and exposed samples. A full experimental protocol is described in Appendix B. We subselect genes that are differentially expressed (either over or under expressed) compared to the control samples at each week and filter these to only include genes in the neighborhood of known genes related to radiation exposure, which we describe in the following section. We discuss our reasoning for choosing bulk RNA-seq expression over single-cell RNA-seq, which is popular for causal discovery, in Appendix B.4.

3.2 Curation of a partial ground truth

Following the approach used by Pratapa et al. (2020), we curate a partial ground truth GRN comprised of all currently known transcription factors (genes that code for proteins that control the expression levels of other genes) to target gene regulatory relationships, and protein to protein interaction relationships using four knowledge databases: ENCODE (Consortium et al., 2012), TRUUST (Han et al., 2015), htFTarget (Zhang et al., 2020) and STRING (Von Mering et al., 2005). To tailor this

[1] One Gray is equal to one joule of absorbed radiation energy per kilogram of matter

wide-ranging knowledge graph to our specific context, we manually collect a set of 56 key genes involved in radiation response from a set of publications (these are not necessarily specific to chronic low-dose radiation response). These key genes and associated publications can be found in Appendix Table 3. For each key gene, we take the two-hop neighborhood in the knowledge graph to create our smaller gene set for learning. See Table 1 for the descriptions of each dataset. The partial ground truth graph is the subgraph induced by this subset of genes on the curated knowledge graph.

3.3 Graph and structure learning models

We chose to evaluate six methods based on their varying ability to incorporate prior knowledge. While this is not a comprehensive list, our objective is to understand if prior knowledge that is not context-specific can improve network recovery and identify low-dose radiation response pathways.

GENELink Chen & Liu (2022) train a supervised graph attention neural network with gene expression data, and positive/negative edge pairs of transcription factors and target genes from a known GRN. Positive pairs correspond to the presence of the edge in the GRN and negative pairs correspond to the absence of an edge. The model infers edge probabilities on a test set of positive/negative pairs.

GENIE3 Marbach et al. (2012) construct an ensemble of Random Forest regression models. Each model predicts the expression value of one transcription factor from all target genes. Edges from transcription factors to target genes are weighted according to the feature importance scores of the fitted models. Known transcription factors are used to initialize the number of models in the ensemble.

DAG-GNN Yu et al. (2019) construct a variational autoencoder parameterized by a novel graph neural network architecture. The model simultaneously learns the structure and weights of a nonlinear structure equation model which results in a learned causal adjacency matrix over the gene set. No known knowledge is needed to train this model.

GES Chickering (2002) design a greedy score-based algorithm that traverses the space of equivalence classes (graphs that imply the same conditional independencies) and outputs a partially directed acyclic graph corresponding to the best scoring graph. Instead of a structural prior, GES assumes a linear Gaussian data distribution and optimizes the Bayesian information criterion.

PC Spirtes et al. (2000) define a constraint-based algorithm that learns edges using conditional independence tests in level sets. Similar to GES, no structural priors are used for the PC algorithm, but the conditional independence test (Fisher z-transformation) assumes a linear Gaussian multivariate model.

DirectLiNGAM Shimizu et al. (2011) assume a linear non-Gaussian acyclic model, for which the causal graph is identifiable without interventions, to learn the causal ordering of variables by successively subtracting the effect of each independent component from the given data in the model. From the causal ordering, the full structure of the causal graph can be learned using linear regression.

4 Experiments

We evaluate GENELink, GENIE3, DAG-GNN, GES, PC, and DirectLiNGAM using our low-dose radiation dataset described in Section 3.1 and partial ground truth network described in Section 3.2. We use a train, test, evaluation split on the partial ground truth edge set following the methods described in Chen & Liu (2022) so that the test set is 1/3 of the total number of edges, and each set is balanced for positive and negative labels. A list of known transcription factors in the train set are provided as input to GENIE3. For a description of training parameters, model architectures and hyperparameters for each method see Appendix C. We bootstrap each model over 10 iterations, and find the optimal threshold for pruning edge weights by optimizing the F1-score of each estimated graph using the training set. Results for the test set on dose rate A are shown Table 2. Further, we estimate the genes directly affected by radiation for GENIE3, DAG-GNN, PC, GES, and DirectLiNGAM. For DAG-GNN, PC, GES and DirectLiNGAM we add the observed cumulative radiation as a random variable in the dataset. For GENIE3, we add radiation as a transcription factor – noting that it should be upstream of all genes. The estimated radiation neighborhood for DAG-GNN is visualized in Appendix Fig. 3. To identify activated pathways we perform pathway enrichment analysis using gProfiler (Raudvere et al., 2019). The enriched pathways and p-values for the genes directly affected by radiation at each dose rate are shown in Fig. 2.

Dose (# edges)	Method	AUC-PR ↑	F1 ↑	TP↑	FP ↓
A (570)	GENELink	**0.287** (0.194)	**0.369** (0.219)	**134.6** (127.7)	967.2 (1633.7)
	GENIE3	0.027 (0.006)	0.048 (0.01)	31.3 (11.7)	714.6 (304.5)
	DAG-GNN	0.020 (0.008)	0.022 (0.0)	2.10 (1.30)	107.0 (154.6)
	GES	0.030 (0.007)	0.033 (0.008)	11.0 (6.0)	382.1 (129.7)
	PC	0.017 (0.011)	0.022 (0.0)	1.4 (1.36)	**65.2** (5.7)
	DirectLiNGAM	0.028 (0.006)	0.031 (0.007)	11.8 (3.1)	488.6 (165.9)

Table 2: Results for recovery of the test set for dose A. TP stands for true positive, FP stands for false positive. Values in parentheses show the standard error across bootstrap runs.

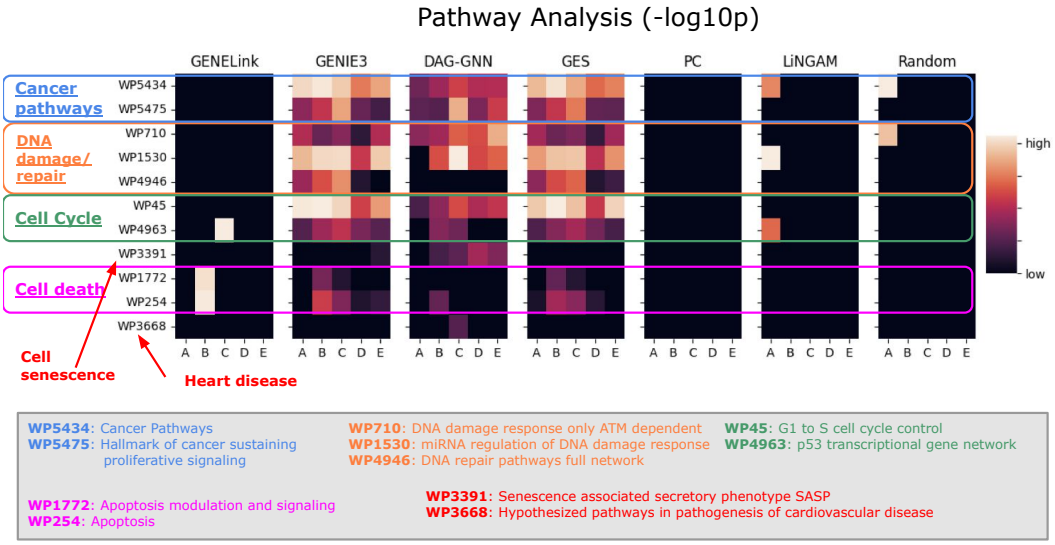

Figure 2: Pathways term names from WikiPathways (Martens et al., 2021) are shown in the text on the left, categorized by the type of radiation response. Brighter values imply more significant enrichment of the corresponding pathway terms for a given model.

5 Discussion

Recovery of edges from knowledge databases Table 2 shows that GENELink is the best at recovering the partial ground truth graph for dose rate A, achieving the best AUC-PR and F1 scores – this holds across dose rates (Appendix Table 4). The remaining algorithms have roughly the same, and relatively poor, performance. For reference, a random estimator would have an AUC-PR value of 0.011 (# positive edges in test set/ total number of edges in test set). Given that GENELink is trained using a subset of partial ground truth edges, it's comparative performance is expected. We note that all methods have very large standard errors across bootstrap runs, likely due to our low sample size.

Pathway enrichment analysis Fig. 2 shows pathway enrichment using gProfiler. Pathways that are highly enriched by a gene set have high -log10 p-values; implying higher confidence that the gene set overlaps with known genes in that pathway. For GENIE3, DAG-GNN, GES, PC and DirectLiNGAM we perform enrichment analysis on genes predicted to be directly affected by radiation. For GENELink we take the 100 genes with the highest out-degrees. We include enrichment of a set of 100 random genes in each dataset for reference. We expect to see enrichment of pathways shown in Fig. 1 with a dose dependent pattern; certain pathways may be highly enriched in a specific dose range. DAG-GNN, GENIE3 and GES enrich more pathways implicated in low-dose radiation response compared to the other methods and the random baseline. Cancer pathways are enriched across doses – the highest being at 0.1 mGy/hr (dose rate C). DNA damage response pathways are active across all doses, but from DAG-GNN graphs we see a dose-dependent activation of specific damage pathways. According to DAG-GNN graphs, the *mirRNA regulation of DNA damage response*

pathway (which depends on a combination of ATM and ATR genes) is enriched at higher dose rates compared to the *DNA damage response only ATM dependent* pathway. DAG-GNN, GENIE3 and GES corroborate evidence that miRNA plays a role in DNA damage response (Wouters et al., 2011). The enrichment of the *G1 to S cell cycle control* implies the presence of cell cycle arrest across doses. While cell cycle arrest is a temporary pause in cell division, *cell senescence* is a permanent stop in cell division due to irreparable DNA damage. There is evidence that cellular senescence in HUVECs is active at dose rate 1.4mGy/hr (Rombouts et al., 2014); DAG-GNN graphs show that senescence pathways are active at dose rates as low as 0.01 mGy/hr (dose rate B). Apoptosis has limited enrichment across doses, as expected given the low dose rate. DAG-GNN graphs show some low enrichment of pathways related to heart disease at 0.1 mGy/hr (dose rate C). GES and GENIE3 have almost identical enrichment of pathways; this is largely because both learned networks are dense and the radiation neighborhood is close to the size of the full gene set. In contrast, DAG-GNN has a smaller radiation neighborhood (see Fig. 3), but the enrichment of important pathways shows that these genes are specific to our context. In general, GENIE3, DAG-GNN and GES enrich low-dose radiation pathways of interest, however, DAG-GNN provides a more nuanced dose-dependent description of these pathways.

Limitations GENIE3, DAG-GNN and GES do not effectively recover the exact edges involved in these pathways. For this task, collecting single-cell, time series, and targeted gene perturbation data will likely yield more accurate results because this data is closer to meeting causal assumptions – including having access to larger sample sizes. Instead, we show that these methods can be effective data analysis tools to understand dose-dependent behavior despite limitations in the data.

6 Conclusion

Analysis of chronic exposure to low-dose radiation is an under explored area, but it has potential to reveal mechanisms crucial to understanding long term prognosis for people exposed to low levels of radiation over time. We provide a dataset of RNA-seq gene expression for HUVEC cells chronically exposed to low-dose radiation, which is outside the scope of existing studies. We perform GRN inference and find that structural knowledge from databases is generally not transferrable to understanding pathways activated by low-dose radiation. Moreover, models that jointly learn a graph over gene sets and radiation levels allow us to find the genes most affected by radiation for downstream pathway analysis; of these, DAG-GNN has the best dose-dependent pathway enrichment. We posit that methods like GENIE3, DAG-GNN and GES can be powerful tools for similar under-explored scientific areas, especially in cases with small sample sizes, and with observed environmental perturbations.

Acknowledgments and Disclosure of Funding

We thank Becca Weinburg for helpful discussions and explanations of experimental protocols. AS is supported by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration. This research used resources of the Argonne Leadership Computing Facility. Argonne National Laboratory's work on the Exploration of the Potential for Artificial Intelligence and Machine Learning to Advance Low-Dose Radiation Biology Research (RadBio-AI) project was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environment Research, under contract DE-AC02-06CH11357.

References

Akuwudike, P., López-Riego, M., Marczyk, M., Kocibalova, Z., Brückner, F., Polańska, J., Wojcik, A., and Lundholm, L. Short-and long-term effects of radiation exposure at low dose and low dose rate in normal human vh10 fibroblasts. *Frontiers in Public Health*, 11:1297942, 2023.

Babini, G., Baiocco, G., Barbieri, S., Morini, J., Sangsuwan, T., Haghdoost, S., Yentrapalli, R., Azimzadeh, O., Rombouts, C., Aerts, A., et al. A systems radiation biology approach to unravel the role of chronic low-dose-rate gamma-irradiation in inducing premature senescence in endothelial cells. *Plos one*, 17(3):e0265281, 2022.

Bao, L., Ma, J., Chen, G., Hou, J., Hei, T. K., Yu, K., and Han, W. Role of heme oxygenase-1 in low dose radioadaptive response. *Redox Biology*, 8:333–340, 2016.

Brouillard, P., Lachapelle, S., Lacoste, A., Lacoste-Julien, S., and Drouin, A. Differentiable causal discovery from interventional data. *Advances in Neural Information Processing Systems*, 33: 21865–21877, 2020.

Chen, G. and Liu, Z.-P. Graph attention network for link prediction of gene regulations from single-cell rna-sequencing data. *Bioinformatics*, 38(19):4522–4529, 2022.

Chevalley, M., Roohani, Y. H., Mehrjou, A., Leskovec, J., and Schwab, P. A large-scale benchmark for network inference from single-cell perturbation data. *Communications Biology*, 8(1):412, 2025.

Chickering, D. M. Learning equivalence classes of bayesian-network structures. *Journal of machine learning research*, 2(Feb):445–498, 2002.

Consortium, E. P. et al. An integrated encyclopedia of dna elements in the human genome. *Nature*, 489(7414):57, 2012.

Dai, H., Ng, I., Luo, G., Spirtes, P., Stojanov, P., and Zhang, K. Gene regulatory network inference in the presence of dropouts: a causal view. *arXiv preprint arXiv:2403.15500*, 2024.

Fang, F., Yu, X., Wang, X., Zhu, X., Liu, L., Rong, L., Niu, D., and Li, J. Transcriptomic profiling reveals gene expression in human peripheral blood after exposure to low-dose ionizing radiation. *Journal of radiation research*, 63(1):8–18, 2022.

Ghandhi, S. A., Yaghoubian, B., and Amundson, S. A. Global gene expression analyses of bystander and alpha particle irradiated normal human lung fibroblasts: synchronous and differential responses. *BMC medical genomics*, 1(1):63, 2008.

Ghandhi, S. A., Ming, L., Ivanov, V. N., Hei, T. K., and Amundson, S. A. Regulation of early signaling and gene expression in the α-particle and bystander response of imr-90 human fibroblasts. *BMC medical genomics*, 3(1):31, 2010.

Ghandhi, S. A., Sinha, A., Markatou, M., and Amundson, S. A. Time-series clustering of gene expression in irradiated and bystander fibroblasts: an application of fbpa clustering. *Bmc Genomics*, 12(1):2, 2011.

Greenfield, A., Hafemeister, C., and Bonneau, R. Robust data-driven incorporation of prior knowledge into the inference of dynamic regulatory networks. *Bioinformatics*, 29(8):1060–1067, 2013.

Han, H., Shim, H., Shin, D., Shim, J. E., Ko, Y., Shin, J., Kim, H., Cho, A., Kim, E., Lee, T., et al. Trrust: a reference database of human transcriptional regulatory interactions. *Scientific reports*, 5 (1):11432, 2015.

Katsura, M., Urade, Y., Nansai, H., Kobayashi, M., Taguchi, A., Ishikawa, Y., Ito, T., Fukunaga, H., Tozawa, H., Chikaoka, Y., et al. Low-dose radiation induces unstable gene expression in developing human ipsc-derived retinal ganglion organoids. *Scientific Reports*, 13(1):12888, 2023.

Keshavarzi, O., Haddadi, G., Fardid, R., Haghani, M., Kalantari, T., and Namdari, A. Investigating the expression levels of bax and bcl-2 genes in peripheral blood lymphocytes of industrial radiation workers in the asaluyeh region. *Journal of Biomedical Physics & Engineering*, 14(3):275, 2024.

Kim, R., Inoue, H., and Toge, T. Bax is an important determinant for radiation sensitivity in esophageal carcinoma cells. *International journal of molecular medicine*, 14(4):697–1403, 2004.

Lee, E.-S., Won, Y. J., Kim, B.-C., Park, D., Bae, J.-H., Park, S.-J., Noh, S. J., Kang, Y.-R., Choi, S. H., Yoon, J.-H., et al. Low-dose irradiation promotes rad51 expression by down-regulating mir-193b-3p in hepatocytes. *Scientific reports*, 6(1):25723, 2016.

Liu, X., He, Y., Li, F., Huang, Q., Kato, T. A., Hall, R. P., and Li, C.-Y. Caspase-3 promotes genetic instability and carcinogenesis. *Molecular cell*, 58(2):284–296, 2015.

Lopez, R., Hütter, J.-C., Pritchard, J., and Regev, A. Large-scale differentiable causal discovery of factor graphs. *Advances in Neural Information Processing Systems*, 35:19290–19303, 2022.

Marbach, D., Costello, J. C., Küffner, R., Vega, N. M., Prill, R. J., Camacho, D. M., Allison, K. R., Kellis, M., Collins, J. J., et al. Wisdom of crowds for robust gene network inference. *Nature methods*, 9(8):796–804, 2012.

Martens, M., Ammar, A., Riutta, A., Waagmeester, A., Slenter, D. N., Hanspers, K., A. Miller, R., Digles, D., Lopes, E. N., Ehrhart, F., et al. Wikipathways: connecting communities. *Nucleic acids research*, 49(D1):D613–D621, 2021.

Okazaki, R. Role of p53 in regulating radiation responses. *Life*, 12(7):1099, 2022.

Pratapa, A., Jalihal, A. P., Law, J. N., Bharadwaj, A., and Murali, T. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nature methods*, 17(2): 147–154, 2020.

Ramsey, J. and Andrews, B. Py-tetrad and rpy-tetrad: A new python interface with r support for tetrad causal search. In *Causal Analysis Workshop Series*, pp. 40–51. PMLR, 2023.

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., and Vilo, J. g: Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic acids research*, 47(W1):W191–W198, 2019.

Rombouts, C., Aerts, A., Quintens, R., Baselet, B., El-Saghire, H., Harms-Ringdahl, M., Haghdoost, S., Janssen, A., Michaux, A., Yentrapalli, R., et al. Transcriptomic profiling suggests a role for igfbp5 in premature senescence of endothelial cells after chronic low dose rate irradiation. *International journal of radiation biology*, 2014.

Shah, A., DePavia, A. F., Hudson, N. C., Foster, I., and Stevens, R. Causal discovery over high-dimensional structured hypothesis spaces with causal graph partitioning. *Transactions on Machine Learning Research*, 2025. ISSN 2835-8856. URL https://openreview.net/forum?id=FecsgPCOHk.

Shimizu, S., Inazumi, T., Sogawa, Y., Hyvarinen, A., Kawahara, Y., Washio, T., Hoyer, P. O., Bollen, K., and Hoyer, P. Directlingam: A direct method for learning a linear non-gaussian structural equation model. *Journal of Machine Learning Research-JMLR*, 12(Apr):1225–1248, 2011.

Shimizu, Y., Kodama, K., Nishi, N., Kasagi, F., Suyama, A., Soda, M., Grant, E. J., Sugiyama, H., Sakata, R., Moriwaki, H., et al. Radiation exposure and circulatory disease risk: Hiroshima and nagasaki atomic bomb survivor data, 1950-2003. *Bmj*, 340, 2010.

Spirtes, P., Glymour, C. N., and Scheines, R. *Causation, prediction, and search*. MIT press, 2000.

Verheij, M. and Bartelink, H. Radiation-induced apoptosis. *Cell and tissue research*, 301(1):133–142, 2000.

Von Mering, C., Jensen, L. J., Snel, B., Hooper, S. D., Krupp, M., Foglierini, M., Jouffre, N., Huynen, M. A., and Bork, P. String: known and predicted protein–protein associations, integrated and transferred across organisms. *Nucleic acids research*, 33(suppl_1):D433–D437, 2005.

Wouters, M. D., van Gent, D. C., Hoeijmakers, J. H., and Pothof, J. Micrornas, the dna damage response and cancer. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 717(1-2):54–66, 2011.

Wyrobek, A., Manohar, C., Krishnan, V., Nelson, D., Furtado, M., Bhattacharya, M., Marchetti, F., and Coleman, M. Low dose radiation response curves, networks and pathways in human lymphoblastoid cells exposed from 1 to 10 cgy of acute gamma radiation. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 722(2):119–130, 2011.

Yao, S., Yoo, S., and Yu, D. Prior knowledge driven granger causality analysis on gene regulatory network discovery. *BMC bioinformatics*, 16(1):273, 2015.

Yu, Y., Chen, J., Gao, T., and Yu, M. Dag-gnn: Dag structure learning with graph neural networks. In *International conference on machine learning*, pp. 7154–7163. PMLR, 2019.

Yuan, Q. and Duren, Z. Inferring gene regulatory networks from single-cell multiome data using atlas-scale external data. *Nature Biotechnology*, 43(2):247–257, 2025.

Zanni, V., Papakonstantinou, D., Kalospyros, S. A., Karaoulanis, D., Biz, G. M., Manti, L., Adamopoulos, A., Pavlopoulou, A., and Georgakilas, A. G. Radphysbio: A radiobiological database for the prediction of cell survival upon exposure to ionizing radiation. *International Journal of Molecular Sciences*, 25(9):4729, 2024.

Zhang, Q., Liu, W., Zhang, H.-M., Xie, G.-Y., Miao, Y.-R., Xia, M., and Guo, A.-Y. htftarget: a comprehensive database for regulations of human transcription factors and their targets. *Genomics, proteomics & bioinformatics*, 18(2):120–128, 2020.

A Radiation Specific Genes

List of genes involved in radiation response and corresponding citations which were manually curated. Genes are included in this list if they were shown to be differentially expressed and listed in a table, or in the case of theoretical papers, if they were mentioned in a discussion of radiation response. This list is dose agnostic.

Table 3: Radiation specific genes with citations

Gene name	Citations
TP53	Ghandhi et al. (2011); Akuwudike et al. (2023); Wyrobek et al. (2011); Katsura et al. (2023); Okazaki (2022)
MYC	Ghandhi et al. (2011); Akuwudike et al. (2023); Wyrobek et al. (2011); Katsura et al. (2023); Okazaki (2022); Bao et al. (2016); Lee et al. (2016)
FOS	Ghandhi et al. (2011); Wyrobek et al. (2011); Fang et al. (2022); Ghandhi et al. (2010, 2008)
BCL2	Wyrobek et al. (2011); Keshavarzi et al. (2024); Okazaki (2022); Ghandhi et al. (2008, 2010); Liu et al. (2015)
FAS	Ghandhi et al. (2008, 2010, 2011); Akuwudike et al. (2023); Wyrobek et al. (2011); Fang et al. (2022); Katsura et al. (2023); Okazaki (2022); Bao et al. (2016); Liu et al. (2015)
NFYB	Wyrobek et al. (2011)
E2F4	Wyrobek et al. (2011)
B2M	Wyrobek et al. (2011)
EGR2	Wyrobek et al. (2011)
CDKN1A	Ghandhi et al. (2010, 2011); Akuwudike et al. (2023); Wyrobek et al. (2011)
GADD45A	Ghandhi et al. (2011); Akuwudike et al. (2023); Wyrobek et al. (2011)
ATM	Okazaki (2022)
ATR	Okazaki (2022)
SOD2	Ghandhi et al. (2008)
GPX1	Okazaki (2022)
HMOX1	Bao et al. (2016)
BAX	Okazaki (2022); Keshavarzi et al. (2024); Kim et al. (2004)
CASP3	Liu et al. (2015)
ILB	Ghandhi et al. (2008, 2010, 2011)
IL6	Ghandhi et al. (2008, 2010, 2011)
IL8	Ghandhi et al. (2008, 2010, 2011)
IL33	Ghandhi et al. (2008, 2010, 2011)
TNF	Ghandhi et al. (2008, 2010, 2011); Wyrobek et al. (2011); Fang et al. (2022); Okazaki (2022)
TNFAIP3	Ghandhi et al. (2008, 2010, 2011); Wyrobek et al. (2011); Fang et al. (2022); Okazaki (2022)
TNF-alpha	[Ghandhi et al. (2008, 2010, 2011); Wyrobek et al. (2011); Fang et al. (2022); Okazaki (2022)
NFKB1	Ghandhi et al. (2011)
EGR1	Wyrobek et al. (2011)
RAD51	Lee et al. (2016)
MDM2	Akuwudike et al. (2023); Okazaki (2022); Wyrobek et al. (2011); Ghandhi et al. (2008, 2011)

XPC	Akuwudike et al. (2023); Okazaki (2022)
DDB2	Okazaki (2022); Wyrobek et al. (2011); Ghandhi et al. (2008, 2010)
TGF-beta-1	Okazaki (2022)
CXCL2	Ghandhi et al. (2008, 2010, 2011)
CXCL3	Ghandhi et al. (2008, 2010, 2011)
CXCL4	Ghandhi et al. (2008, 2010, 2011)
GDF15	Ghandhi et al. (2008, 2011)
FDXR	Ghandhi et al. (2008, 2010)
PTGS2	Ghandhi et al. (2008, 2010)
FGF2	Ghandhi et al. (2008); Katsura et al. (2023)
POU5F1	Ghandhi et al. (2008); Katsura et al. (2023)
MMP1	Ghandhi et al. (2008, 2010)
MMP3	Ghandhi et al. (2008, 2010)
DKK1	Ghandhi et al. (2008, 2010)
SERPINB2	Ghandhi et al. (2008)
IL1A	Ghandhi et al. (2008)
IL1B	Ghandhi et al. (2008)
LIF	Ghandhi et al. (2008)
MMP10	Ghandhi et al. (2008)
ATF3	Ghandhi et al. (2008)
BCL2A1	Ghandhi et al. (2008, 2010)
MT1E	Ghandhi et al. (2011)
KDM5B	Ghandhi et al. (2011)
BMP2	Ghandhi et al. (2008)
KYNU	Ghandhi et al. (2008)
LAMB3	Ghandhi et al. (2008)
ETS1	Wyrobek et al. (2011)

B Experimental Protocols

B.1 Radiation source plates

To produce radioactive source plates optimized for long term low-dose rate exposure of cells, gamma-emitting nuclides (137 Cs) were dissolved in carrier solvent and deposited into select wells of a 96-well plate. Specific "hot" well locations were selected to maximize separation between different doses on the plates. After drying, a radiation-resistant epoxy resin was added to create a sealed source. This source plate was inverted to allow close proximity to cell culturing plates placed on top. Collimators were fabricated using commercially available high-Z shielding material (tungsten bismuth polymer), and these collimators placed between the source plate and the culturing plate. Our custom radiation source plates allowed investigation of impacts of a broad range of extremely low dose rate exposures: 0.001 mGy/hr, 0.01 mGy/hr, 0.1 mGy/hr, 1 mGy/hr and 2 mGy/hr.

B.2 Cell culture

Single-donor human vascular endothelial cells (HUVECs) (Cat. No. C-12200, PromoCell) were cultured in Human Large Vessel Endothelial Cell Basal Medium (Gibco) supplemented with 1X Large Vesselplement (Gibco). HUVEC trypsinization was performed using phosphate-buffered saline (PBS, Cat. No. SH30256.01 Cytiva) for rinsing, 0.25% Trypsin (Cat. No. SH40003.01, Cytiva) for detachment, and Trypsin Neutralizing Solution (Cat. No. CC-5002, Lonza). HUVECs were harvested until cell replication was insufficient to continue, at 3 weeks.

Cells were split twice per week to prevent overgrowth between weekly harvest days. The split ratio was determined based on a visual assessment of confluency and adjusted as growth rates diminished progressively throughout the experiment, ranging between 1:4 and 1:12. Cells were pooled by dose rate, diluted with media, and reseeded on fresh 96-well plates. Culturing was carried out in Nunc Edge 2.0 96-well Plates (Thermo Fisher Scientific) for continued cell culture and nucleic acid extractions, and in PhenoPlate 96-well microplates (Revvity) for cell painting experiments. All empty wells were filled with sterile water to mitigate edge effects. The moats of the Nunc Edge 2.0 96-well plates were filled with 3 mL sterile water per moat to reduce edge effects further.

B.3 RNA-seq

Raw, paired-end sequencing FASTQ files and were assessed for quality using FASTQC to identify potential technical issues. Adapter trimming and removal of low-quality bases were carried out using TrimGalore (v0.6.10) to ensure high-quality input for alignment. The cleaned reads were then aligned to the GRCh38.p14 human reference genome using the STAR aligner (v2.7.11b). Gene-level counts were generated with featureCounts (v2.0.6), and transcript abundance was estimated with TPMCalculator (v0.0.3). Each irradiated condition included two biological replicates. Differential expression analysis was performed using DESeq2 (pydeseq2 v0.4.9), by contrasting weekly irradiated samples against untreated controls. Genes were considered significantly differentially expressed if they met the criteria of an adjusted p-value < 0.05 and an absolute log2 fold-change ≥ 1.0.

B.4 A note about single-cell versus bulk RNA-seq data

A natural question is what the reasoning is for our choice of measuring bulk rather than single-cell gene expression – especially since single-cell datasets have been increasingly used for causal discovery because it better matches the assumptions needed for identifiability of the causal structure and generates tens of thousands of samples (Brouillard et al., 2020; Lopez et al., 2022; Chevalley et al., 2025). Our reasoning is two-fold: First, single-cell RNA-seq costs around 10x more than bulk RNA-seq (this ends up being $100k for the duration of our experiments); second, single-cell RNA-seq data is most valuable when analyzing the expression profile of heterogeneous cells. For example, in tissue samples we see heterogeneity between cells because a tissue sample contains cells of varying cell types – different cell types have different expression profiles. In our case, the cells come from a *cell-line*, meaning they are all the same cell type. If we were to measure single-cell expression, we would likely observe cells in different stages of the cell-cycle. While this is valuable for understanding gene to gene causal relationships relevant to the cell-cycle, since we are more interested in the cells' response to radiation our focus is to collect data where this signal would be strongest. In bulk expression data, cells are pooled and expression values are averaged. This improves the statistical power of the data, and strengthens radiation signals which we expect to affect all cells. In contrast, single-cell RNA-seq is highly susceptible to noise and dropouts, which has been well-documented (Dai et al., 2024). There is some heterogeneity specific to radiation response; some cells will die earlier than others, and therefore it is of interest to measure the expression of each cell to understand why a cell survived. Further, there is a coupling between cell-cycle and radiation exposure. we are able to measure cell-cycle arrest and cell senescence signals in bulk-expression data based on Fig. 2, however single-cell may be able to provide more information to the coupling between cell-cycle and radiation response. Given the cost of single-cell, we see more value in measuring bulk-expression of cells under radiation, however single-cell could be valuable for understanding cell-cycle related radiation response in the future.

C Model Training and Hyperparameters

We use the default settings for training/running all methods. Here we describe these settings in detail.

- GENELink embeds the prior GRN and gene expression data with two GAT layers (each with three attention heads), with `hidden_dim` = 128 and `output_dim` = 64. The outputs of the GAT layers are connected to two separate two-layer MLP channels both with `input_dim`=64, `hidden_dim`=32, `output_dim`=16. Each channel corresponds to transcription factors (TF) and target gene embeddings. For a given (TF, target) train or test pair, edge probabilities are computed as the dot product of the transcription factor and target gene embeddings. GENELink is trained for 5 epochs using the Adam optimizer with batch size 256, learning rate of 3e-3 (`gamma`=0.99), and binary cross entropy loss.
- For GENIE3, each RandomForestRegressor model has `n_estimators` = 1000 trees and `max_features` = $\sqrt{\#genes}$. GENIE3 ensemble models are fit in parallel with 16 threads.
- DAG-GNN has a trainable adjacency matrix $A \in \mathbb{R}^{\#genes \times \#genes}$. The encoder is $z = (I - A^T)x$ for input $x \in \mathbb{R}^{\#genes \times \#samples}$, where I is the identity matrix. The decoder is $(I - A^T)^{-1}z$. DAG-GNN is trained by maximizing the evidence lower bound (ELBO) and until convergence of the acyclicity constraint described in Yu et al. (2019). DAG-GNN is trained for 300 epochs using Adam with a learning rate of 3e-3 (`gamma`=1.0, `lr_decay`= 200). The threshold for the adjacency matrix A is set to 0. We tune A separately according to the optimal F1-score with respect to the train set of GENELink.

- For GES, we use the Tetrad implementation (Ramsey & Andrews, 2023) and use the Bayesian information criterion (BIC) score function.
- For the PC algorithm we use the Tetrad implementation (Ramsey & Andrews, 2023) and the Fisher z-score conditional independence test. We set the significance threshold for pruning edges $\alpha = 0.05$, which is the default value.
- For DirectLiNGAM we use the Tetrad implementation (Ramsey & Andrews, 2023).

All models were trained on a NVIDIA Tesla V100-SXM2-32GB GPU. Due to memory constraints with the DAG-GNN model architecture, we partitioned the gene set according to a causal partition (Shah et al., 2025) with respect to the partial ground truth network. Training times depend on the gene set size, which vary by dose rate. For dose rate A (918 genes) the average training times are as follows: GENELink (1 min), GENIE3 (10.7 min), DAG-GNN (23.3 min), PC (8.9 min), GES (57.7 min), and DirectLiNGAM (1.5 min).

D DAG-GNN Radiation Neighborhoods

Fig. 3 shows the two-hop neighborhood for “radiation" after including the cumulative radiation dose as a random variable in DAG-GNN graph estimation. The size of the node corresponds to the out-degree of the node and the width of the edge corresponds to the edge weight magnitude. These graphs were generated using the consensus over 10 bootstrap runs so that edges that occurred in $\geq 50\%$ of runs were retained and visualized – edge weights were averaged. The node set is also the gene set used for pathway enrichment in Fig. 2. GENIE3 and GES graphs are not shown here because the neighborhoods are very large and difficult to visualize.

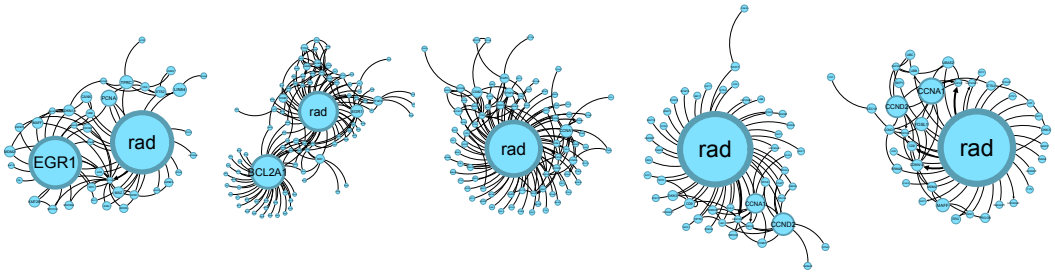

Figure 3: The radiation (`rad`) neighborhoods estimated by DAG-GNN for dose rates A (left most) through E (right most). Thicker edges indicate larger weights between nodes.

E Pathway Enrichment

We use gProfiler (Raudvere et al., 2019) for pathway enrichment analysis of genes in the neighborhood of radiation for GENIE3, DAG-GNN, PC and GES, the top 100 out-degree nodes for GENELink, and a set of 100 random genes in the dataset as comparison. gprofiler performs functional profiling of gene lists using various kinds of biological evidence. The tool performs statistical enrichment analysis to find over-representation of information in the gene set from Gene Ontology terms, biological pathways, regulatory DNA elements, human disease gene annotations, and protein-protein interaction networks. For our analysis, we focus on pathways defined in WikiPathways (Martens et al., 2021). gProfiler uses Fisher’s one-tailed test, also known as cumulative hypergeometric probability, as the p-value measuring the randomness of the intersection between the query gene set and the WikiPathway term term. The p-value represents the probability of the observed intersection plus probabilities of all larger, more extreme intersections. The higher the -log10 p-value the stronger the enrichment of that term is in the gene set.

F Doses A-E Results

Dose (# positive edges in test set)	Method	AUC-PR ↑	F1 ↑	TP↑	FP ↓
A (570)	GENELink	**0.287** (0.194)	**0.369** (0.219)	**134.6** (127.7)	967.2 (1633.7)
	GENIE3	0.027 (0.006)	0.048 (0.01)	31.3 (11.7)	714.6 (304.5)
	DAG-GNN	0.020 (0.008)	0.022 (0.0)	2.10 (1.30)	107.0 (154.6)
	GES	0.030 (0.007)	0.033 (0.008)	11.0 (6.0)	382.1 (129.7)
	PC	0.017 (0.011)	0.022 (0.0)	1.4 (1.4)	**65.2** (5.7)
	DirectLiNGAM	0.028 (0.006)	0.031 (0.007)	11.8 (3.1)	488.6 (165.9)
B (1241)	GENELink	**0.268** (0.173)	**0.398** (0.203)	**481.7** (327.3)	534.1 (234.8)
	GENIE3	0.021 (0.001)	0.044 (0.006)	111.5 (75.2)	3522.2 (2869.9)
	DAG-GNN	0.022 (0.006)	0.021 (0.008)	16.70 (29.7)	615.4 (1286.3)
	GES	0.019 (0.002)	0.020 (0.002)	15.6 (3.7)	890.7 (243.3)
	PC	0.012 (0.006)	0.017 (0.0)	1.5 (1.1)	**107.7** (12.7)
	DirectLiNGAM	0.019 (0.004)	0.021 (0.003)	16.2 (2.6)	965.7 (375.9)
C (1118)	GENELink	**0.236** (0.211)	**0.312** (0.231)	**428.9** (366.2)	877.3 (577.8)
	GENIE3	0.015 (6e-4)	0.033 (0.002)	138.6 (33.1)	7130.9 (1730.1)
	DAG-GNN	0.019 (0.001)	0.017 (3e-4)	4.6 (7.4)	252.5 (540.6)
	GES	0.018 (0.003)	0.020 (0.004)	15.6 (6.0)	1032.1 (504.9)
	PC	0.015 (0.007)	0.017 (0.0)	2.2 (1.5)	**111.9** (10.2)
	DirectLiNGAM	0.018 (0.003)	0.019 (0.002)	14.7 (6.1)	1054.0 (532.1)
D (928)	GENELink	**0.074** (0.075)	**0.135** (0.151)	**248.4** (211.1)	8224.9 (11982.6)
	GENIE3	0.015 (0.003)	0.025 (0.005)	23.3 (9.7)	1019.7 (721.1)
	DAG-GNN	0.042 (0.029)	0.017 (0.005)	9.7 (10.0)	478.6 (1099.5)
	GES	0.017 (0.004)	0.019 (0.006)	12.2 (2.6)	782.7 (194.8)
	PC	0.010 (0.006)	0.019 (0.0)	1.6 (1.2)	**118.8** (10.1)
	DirectLiNGAM	0.022 (0.004)	0.025 (0.006)	16.1 (2.9)	719.7 (54.5)
E (480)	GENELink	**0.404** (0.245)	**0.377** (0.030)	**221.0** (176.2)	984.2 (2501.2)
	GENIE3	0.040 (0.007)	0.062 (0.013)	28.4 (10.2)	416.6 (154.3)
	DAG-GNN	0.065 (0.024)	0.028 (0.006)	5.8 (2.6)	**44.8** (6.1)
	GES	0.032 (0.006)	0.035 (0.007)	12.5 (2.9)	446.0 (99.0)
	PC	0.025 (0.006)	0.024 (0.0)	2.1 (0.7)	60.9 (5.5)
	DirectLiNGAM	0.033 (0.004)	0.035 (0.005)	11.4 (4.5)	384.6 (130.1)

Table 4: Metrics for recovery of the test split on the partial ground truth network for all doses. TP stands for true positive, FP stands for false positive. Values in parentheses are standard error.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We claim to provide RNA-seq expression data, a thorough evaluation of three algorithms, and a demonstration of how GENIE3 and DAG-GNN enrich known pathways for radiation response. We support all these claims in our experiments and discussions section.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.

Dose	Method	AUC-PR ↑	F1 ↑	TP↑	FP ↓
A	GENELink	0.113 (0.141)	0.125 (0.0)	**11339.7** (13569.9)	163367.7 (162630.6)
	GENIE3	**0.172** (0.008)	**0.127** (0.006)	3226.3 (1388.9)	9579.8 (5354.5)
	DAG-GNN	0.130 (0.002)	0.125 (0.0)	239.8 (159.4)	**1451.6** (2187.7)
	GES	0.105 (0.003)	0.126 (0.0)	1743.7 (0.0)	11907.0 (2435.3)
	PC	0.116 (0.007)	0.125 (0.0)	176.8 (14.7)	910.8 (0.1)
	DirectLiNGAM	0.100 (0.003)	0.125 (0.0)	1784.8 (85.5)	13305.8 (1290.5)
B	GENELink	0.108 (0.140)	0.104 (0.0)	**15258.1** (22406.8)	265531.6 (403917.5)
	GENIE3	**0.135** (0.005)	**0.126** (0.029)	11553.2 (8115.0)	50422.6 (40052.1)
	DAG-GNN	0.085 (0.016)	0.104 (0.0)	1198.6 (2295.8)	**14439** (29548.0)
	GES	0.084 (0.001)	0.100 (0.0)	2849.7 (188.1)	24595.5 (1683.1)
	PC	0.100 (0.005)	0.104 (0.0)	275.3 (219.2)	1666.7 (80.9)
	DirectLiNGAM	0.083 (0.002)	0.104 (0.0)	2759.7 (251.5)	24463.1 (2970.8)
C	GENELink	0.102 (0.142)	0.100 (0.0)	11600.5 (17924.9)	216024.3 (337188.2)
	GENIE3	**0.122** (0.002)	**0.178** (0.023)	**25509.0** (8810.6)	135448.8 (51799.5)
	DAG-GNN	0.094 (0.018)	0.100 (0.0)	360.3 (591.4)	**4189.2** (9018.0)
	GES	0.081 (0.002)	0.100 (0.0)	2631.8 (169.7)	24024.4 (1668.4)
	PC	0.094 (0.005)	0.100 (0.0)	249.7 (0.0)	1618.7 (77.1)
	DirectLiNGAM	0.079 (0.002)	0.100 (0.0)	2591.0 (121.6)	23997.5 (1721.7)
D	GENELink	0.033 (0.008)	0.063 (0.001)	**21071.5** (230207.6)	614387.1 (680192.4)
	GENIE3	**0.108** (0.005)	**0.066** (0.011)	3941.3 (2081.7)	19179.1 (12234.4)
	DAG-GNN	0.072 (0.029)	0.062 (0.0)	545.4 (982.5)	**14647** (32770.0)
	GES	0.054 (0.002)	0.062 (0.0)	1987.4 (482.6)	27406.0 (6597.0)
	PC	0.067 (0.004)	0.062 (0.0)	260.5 (0.0)	2363.1 (154.2)
	DirectLiNGAM	0.052 (0.002)	0.062 (0.0)	2114.0 (92.7)	31581.4 (2624.4)
E	GENELink	0.112 (0.142)	**0.131** (0.0)	**10000.4** (12228.4)	135080.7 (165169.6)
	GENIE3	**0.183** (0.008)	**0.131** (0.0)	1716.8 (563.6)	4349.5 (1622.1)
	DAG-GNN	0.124 (0.008)	**0.131** (0.0)	135.6 (25.6)	**621.6** (88.7)
	GES	0.109 (0.003)	**0.131** (0.0)	1654.6 (117.0)	11008.4 (978.3)
	PC	0.130 (0.005)	**0.131** (0.0)	180.1 (12.1)	784.9 (33.9)
	DirectLiNGAM	0.104 (0.002)	**0.131** (0.0)	1628.1 (106.8)	11856.9 (697.5)

Table 5: Metrics for recovery of the entire partial ground truth network for all doses. TP stands for true positive, FP stands for false positive. Values in parentheses are standard error.

- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We have a limitations paragraph in the discussion section, mainly relating to the fact that we have a small sample size and that the actual pathways are not necessarily being recovered in the learned graphs.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.

- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

   Justification: This is not a theoretical paper.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We provide a description of how the data was collected from cell cultures in Appendix B and Model training in Appendix C. Due to space limitations we could not fit them in paper.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same

dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.

- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: We will provide the datasets described in Table 1 and the code for model training and evaluation if the paper is accepted and we can deanonymize our work.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: We provide a description of model training in Appendix C. paper.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: For all metrics in Tables 2 we report the standard error over 10 bootstrap runs. For the pathway analysis we do not provide error bars, as this is not standard practice in pathway enrichment however we include the -logp values for the pathways of interest.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
   - If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We report the compute specs in Appendix C and training times for each model. Results are over 10 bootstrapped runs, so the runtimes to replicated the evalutions are ten-fold of what is report in this section.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

Answer: [Yes]

Justification: We have read and followed the guidelines. Regarding human cells, the cells are grown in lab and there are no patient identifiers in the dataset.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: The main societal impact for this paper relates to public health, as the research conducted in this paper is meant to understand the onset of cancer from radiation exposure. We mention this in the introduction and conclusion of the paper. We do not believe there are potential negative impacts.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: We don't believe this paper poses this risk.

Guidelines:

- The answer NA means that the paper poses no such risks.

- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: We credit the authors of GENIE3, GENELink and DAG-GNN. Each paper has an associated code base.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: No new assets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

Justification: The Central Department of Energy Institutional Review Board has determined that this project is Exempt human subjects research.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: The Central Department of Energy Institutional Review Board has determined that this project is Exempt human subjects research.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: LLMs are not used for the research in this paper.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.