



THE 3RD INTERNATIONAL CONFERENCE ON
SYSTEMS MEDICINE, AI & DRUG REPURPOSING

Enhancing the accuracy of Network Medicine through understanding the impact of sample size in gene co-expression networks

Joaquim Aguirre-Plans¹, Bingsheng Chen², Susan Dina Ghiassian³, Alex Jones⁴, Viatcheslav R. Akmaev⁵, Alif Saleh⁶, Deisy Morselli Gysi⁷, Albert-Laszlo Barabasi⁸

¹ [0000-0002-1971-1350](#), Northeastern University, Boston (US)

² [0000-0003-3595-3285](#), Northeastern University, Boston (US)

³ [0000-0003-4745-6331](#), Scipher Medicine Corporation, Waltham (US)

^{4, 5, 6} Scipher Medicine Corporation, Waltham (US)

⁷ [0000-0002-5771-8182](#), Northeastern University, Boston (US)

⁸ [0000-0002-4028-3522](#), Northeastern University, Boston (US)

Abstract

Network Medicine relies on RNA sequencing to infer gene co-expression networks, which are crucial to identify functional gene clusters and gene regulatory interactions, and offer a deeper understanding of disease phenotypes and drug mechanisms¹⁻⁴. Previous studies have used gene co-expression networks to prioritize drug candidates based on their interaction with relevant transcription factors⁵, to identify potential treatments for schizophrenia⁶, or to enrich the disease modules associated to rare diseases⁷. It remains unknown, however, how many samples do we need to make reliable predictions. Here, we propose a power-law model to predict the relationship between the number of inferred significant interactions and sample size, allowing us to quantitatively link sample size to the accuracy of the inferred networks. We apply our model to investigate the effect of sample size on biomarker discovery and differentiation of protein-protein interactions from non-interacting pairs, ultimately unveiling the critical role of data quality in generating meaningful predictions in Network Medicine.

Keywords

sample size, gene co-expression networks, network medicine, rna-seq





THE 3RD INTERNATIONAL CONFERENCE ON
SYSTEMS MEDICINE, AI & DRUG REPURPOSING

References

1. van Dam, S., Vösa, U., van der Graaf, A., Franke, L. & de Magalhães, J. P. Gene co-expression analysis for functional classification and gene-disease predictions. *Brief. Bioinform.* **19**, 575–592 (2018).
2. Gysi, D. M. & Nowick, K. Construction, comparison and evolution of networks in life sciences and other disciplines. *J. R. Soc. Interface* **17**, 20190610 (2020).
3. Chowdhury, H. A., Bhattacharyya, D. K. & Kalita, J. K. (Differential) Co-Expression Analysis of Gene Expression: A Survey of Best Practices. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **17**, 1154–1173 (2020).
4. Paci, P. *et al.* Gene co-expression in the interactome: moving from correlation toward causation via an integrated approach to disease module discovery. *Npj Syst. Biol. Appl.* **7**, 1–11 (2021).
5. De Bastiani, M. *et al.* Master Regulators Connectivity Map: A Transcription Factors-Centered Approach to Drug Repositioning. *Front Pharmacol.* **9**:697 (2018).
6. Truong, T. *et al.* Network-based drug repurposing for schizophrenia. *Neuropsychopharmacology.* **49**(6):983-992 (2024).
7. Buphamalai, P. *et al.* Network analysis reveals rare disease signatures across multiple levels of biological organization. *Nat Commun.* **12**(1):6306 (2021).