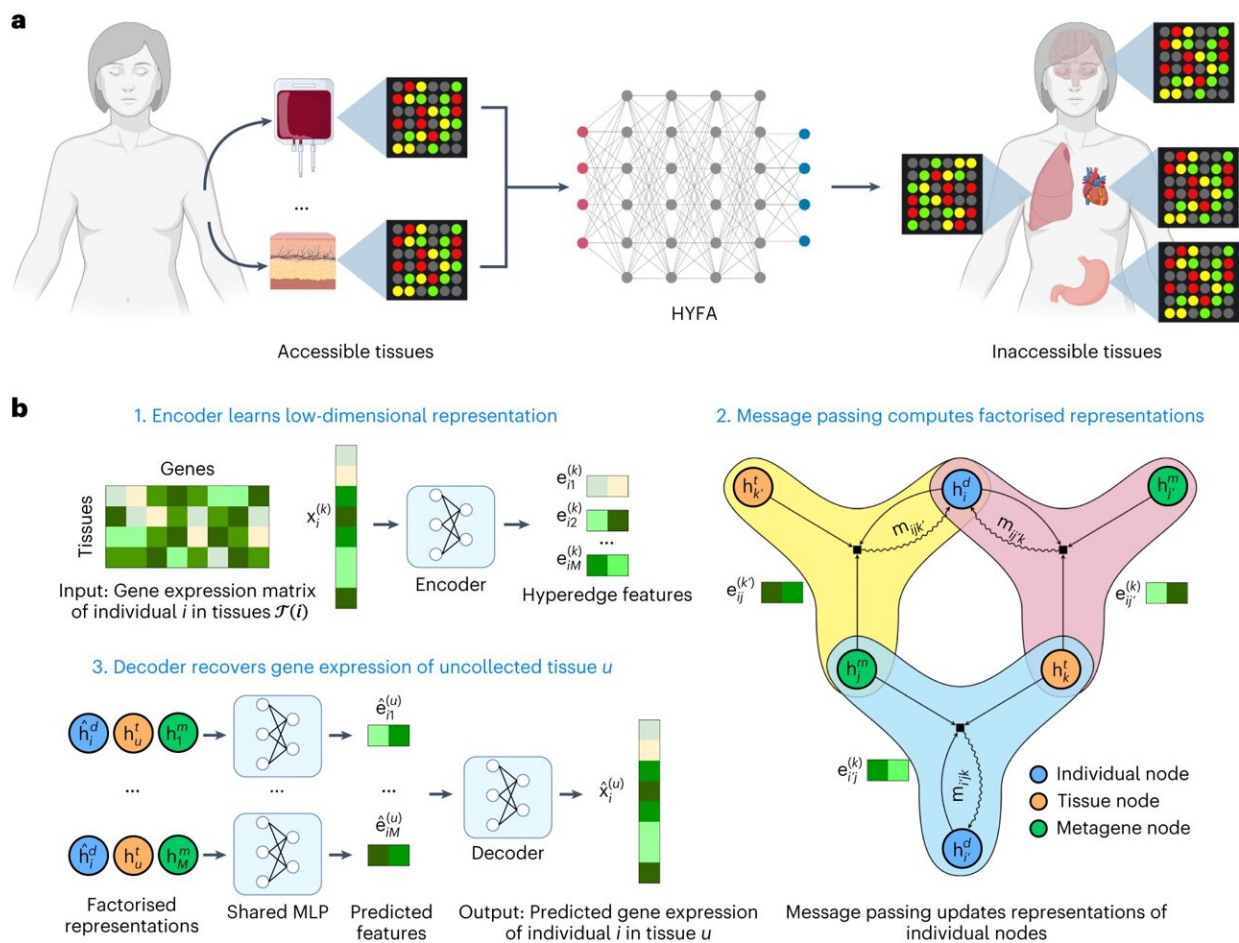# Researchers develop method to predict gene expression across different tissues and cell types

July 28 2023, by Bill Snyder

Overview of HYFA. a, HYFA processes gene expression from a number of collected tissues (for example, accessible tissues) and infers the transcriptomes of uncollected tissues. b, Workflow of HYFA. Credit: *Nature Machine Intelligence* (2023). DOI: 10.1038/s42256-023-00684-8

In recent years, a virtual tidal wave of studies linking the expression of certain genes to complex diseases as varied as cancer and diabetes has raised hopes for major advances in medical treatment and drug discovery.

Yet gene expression datasets from different cell types and tissues are largely disconnected islands of discovery. They do not reveal the rich and often consequential multi-tissue connections that extend, for example, from the brain to the gut.

Now, researchers at Vanderbilt University Medical Center and the University of Cambridge have developed a method of "imputing," or predicting gene expression in hard-to-access tissues like the brain from more accessible tissues, including whole blood.

Their deep-learning/machine-learning approach, called hypergraph factorization, or HYFA, has revealed gene expression patterns shared by the brain and gastrointestinal tract and has validated the expression of certain genes in the blood as a marker for Alzheimer's disease in the brain.

The ability to reconstruct and predict unmeasured gene expression across a broad collection of tissues and cell types "may expand our understanding of the molecular origins of complex traits," the scientists report in a paper featured on the cover of the July issue of the journal *Nature Machine Intelligence*.

"As people generate more molecular data, you need a way of integrating or harmonizing these large datasets," said Eric Gamazon, Ph.D., assistant professor of Medicine in the Division of Genetic Medicine at VUMC, and the paper's co-senior author with Pietro Liò, Ph.D., of the University

of Cambridge.

"Our approach allows one to do that, enabling biomarker discovery and drug repurposing research," Gamazon said.

The study found and validated new genetic variants (changes in the DNA sequence) that can regulate the abundant level of genes in specific tissues and cell types, and which, in turn, may underlie complex diseases and their comorbidities.

Gamazon is a leader in the development and application of gene-expression data, and a contributor to the international Genotype-Tissue Expression (GTEx) project of the National Institutes of Health (NIH) Common Fund.

He and colleagues including Nancy Cox, Ph.D., the Mary Phillips Edmonds Gray Professor of Genetics and director of the Division of Genetic Medicine at VUMC, have previously analyzed transcriptome (gene expression) data from multiple tissues to identify neuroendocrine and gastrointestinal contributors to psychiatric disorders.

The current paper, co-authored by Phillip Lin, a bioinformatics scientist in Gamazon's lab, takes the research to the next level.

The [deep learning approach](#) to gene-expression imputation "promises a systemwide view of human physiology," Gamazon said. "It can also accelerate the integration of these large-scale tissue and cell-type [gene expression](#) biorepositories, as studies, institutions and consortia continue to generate these resources."

More information: Ramon Viñas et al, Hypergraph factorization for multi-tissue gene expression imputation, *Nature Machine Intelligence* (2023). [DOI: 10.1038/s42256-023-00684-8](#)