

---

# DeCovarT: Network-Driven Deconvolution of Transcriptomics data to reveal organoid Cellular Heterogeneity

Bastien Chassagnol<sup>\*1</sup>, Etienne Becht<sup>2</sup>, Grégory Nuel, and Anaïs Baudot<sup>†3</sup>

<sup>1</sup>Marseille medical genetics - Centre de génétique médicale de Marseille – Aix Marseille Université, Institut National de la Santé et de la Recherche Médicale – France

<sup>2</sup>Institut de Recherches SERVIER – Institut de Recherche Servier – France

<sup>3</sup>Aix Marseille Univ, INSERM, MMG, 13385, Marseille, France – Institut National de la Santé et de la Recherche Médicale - INSERM, Aix-Marseille Université - AMU, Aix Marseille Univ, INSERM, MMG, Marseille Medical Genetics, Marseille, France – France

## Abstract

Introduction: Bulk transcriptomic datasets have contributed to the characterization of biological systems. However, their ability to reveal biological mechanisms is limited by the aggregation of heterogeneous signals coming from the aggregated cells. Indeed, variations of gene expression induced by cell-type composition or micro-environmental fluctuations are confounded into a single measurement. Single-cell technologies overcome part of this limitation, but they remain costly and are affected by technical artefacts, such as uneven capture efficiency across cell types.

To address these issues, deconvolution methods have been proposed to separate bulk expression profiles into their cell population components. Deconvolution works by identifying the cell type composition of bulk expression profiles, using reference profiles obtained from bulk or single-cell RNA-seq data. However, current deconvolution approaches struggle to discriminate closely related cell types that share similar expression patterns. We hypothesize that deconvolution performances could be improved by considering gene interactions and regulatory networks.

Methods: We designed DeCovarT, a generative model that describes the distribution of the observed bulk expression profile conditioned to latent cell-type specific expression profiles, described by gene regulatory networks (GRNs). As a first step, we integrated a large set of gene expression profiles for different cell type populations. We inferred gene–gene interactions for each cell type, assuming they follow multivariate Gaussian models. Specifically, we used a weighted version of the graphical Lasso framework.

We assumed that the total bulk expression arises from the sum of cell-type contributions, weighted by their relative abundances. However, instead of assigning each cell type to a static transcriptomic profile, we assumed that the expression of a given cell type was sampled from a multivariate Gaussian distribution. This Gaussian distribution is parameterised by the GRN inferred in the previous step. By assuming independence between cell populations, the

---

\*Speaker

†Corresponding author: [anais.baudot@univ-amu.fr](mailto:anais.baudot@univ-amu.fr)

total bulk expression can be described as a weighted convolution of their cell-specific expression profiles, also identifiable to a multivariate Gaussian distribution. This explicit formulation enables closed-form derivations for the log-likelihood of the model, its gradient, and its Hessian. Cell-type proportions and latent expression profiles were then estimated through maximum likelihood using a Levenberg–Marquardt optimization scheme. A SoftMax-based reparameterization ensured that the inferred cell-type proportions satisfy the unit-simplex constraints.

**Results:** The generative formulation of DeCovarT enables computation of asymptotic confidence intervals for all estimated parameters, providing a principled quantification of uncertainty. Preliminary numerical simulations on synthetic datasets indicate that increasing gene expression correlation strongly influences the performance of tested deconvolution algorithms. Specifically, increasing overlap between cell type expression profiles reduces performance. However, DeCovarT consistently outperforms existing deconvolution algorithms. **Use-case:** Organoids are three-dimensional cellular structures derived from stem cells, increasingly used to model organ and tissue development and function. In this context, bulk transcriptomes are frequently profiled to study organoids. However, because bulk transcriptomic data collapse heterogeneous cell types into a single signal, they obscure cell-type specific differentiation trajectories. We applied a preliminary implementation of DeCovarT to characterize the cellular composition of organoids, providing a more resolved view of cellular differentiation.

**Keywords:** deconvolution, bulk transcriptomics, organoids, graphical Lasso, generative model, optimisation