A Bayesian two-step integrative procedure incorporating prior knowledge for the identification of miRNA-mRNAs involved in hepatocellular carcinoma.

Marie Denis, R.S. Varghese, M.E. Barefoot, M.G. Tadesse, H.W. Ressom

EMBC 2022, July 12, 2022

cirad
AGRICULTURAL RESEARCH FOR DEVELOPMENT

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

17 89
GEORGETOWN UNIVERSITY

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 840383.

Outline

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 840383.

Biological context - HCC context

Hepatocellular carcinoma (HCC) is the **most common** type of liver cancer and the **third cause** of cancer deaths worldwide.

An **aggressive** cancer because:

- diagnosed at advanced stages (in many cases, HCC detected in people with liver cirrhosis (CIRR)),
- known diagnostic markers have low sensitivity for early detection.

The identification of **novel** diagnostic biomarkers for early detection of HCC is **crucial** and is still an active research area.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 840383.

Biological context - General context

MicroRNAs (miRNAs) are small single-stranded non-coding RNAs that regulate target gene expression [1]

Role of miRNAs is **crucial** in many biological processes, in particular those underlying diseases.

↬ The study of their effects may be achieved by linking miRNAs to target genes.

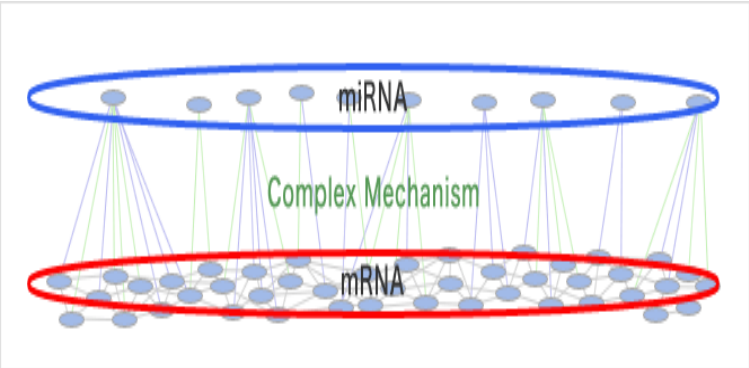


Figure 1: Possible relationships between miRNAs and mRNAs: one miRNA may have multiple mRNA targets and vice-versa.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 840383.

Biological context - Coming back to HCC

Goal

- A better understanding of the biological mechanisms involved in HCC through the identification of miRNA-mRNA disease-associated pairs.

Why ?

- Previous studies: miRNA-mRNA pairs play a **crucial role** in the activation of oncogenic or carcinogenesis pathways in liver diseases or HCC.

But

- The characterization of the relationships between miRNAs and mRNAs is still a **challenge**.

Need to use/develop appropriate statistical methods for a better understanding of links between miRNAs and mRNAs and for identifying relevant pairs.

Statistical context

Objectives

- to improve the understanding of associations / to discover new ones between miRNAs and mRNAs,
- to improve the understanding of associations among mRNAs after considering effects of miRNAs,
- to identify relevant disease-associated miRNA-mRNA pairs.

Many works have shown that:

- the integration of data from various molecular levels leads to better results than analyses considering only one dataset,
- the integration of prior knowledge into statistical models leads to promising results.

We propose a Bayesian two-step **integrative** procedure for analyzing miRNA-seq and mRNA-seq data from patients with HCC or CIRR while integrating **prior knowledge** accumulated from biological experiments or statistical analyses.

Outline

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 840383.

Samples

Human liver tissues from 64 adult patients recruited at MedStar Georgetown University Hospital.

Table 1: Characteristics of patient-derived samples

		HCC (N=39)	CIRR (N=25)	p-value
Age	Mean(SD)	62.02 (11.46)	50.05 (12.1)	**0.0013**
Gender	Male	77%	72%	0.7683
Race	EA	41%	64%	**0.0202**
	AA	33%	32%	
	Asian	26%	0%	
	other	0%	4%	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 840383.

miRNA-seq and mRNA-seq data

Samples

- RNA samples extracted from the 64 liver tissues and analyzed by Illumina Hiseq 4000 [2],
- Gaussianization of data with the R package *huge* [3],
- Selection of 106 mRNAs and 261 miRNAs selected by using Student t-tests and a p-value cut-off of 0.05 after false discovery rate correction.

Prior knowledge between miRNAs and mRNAs

Scores measuring **the belief in the association** between mRNAs and miRNAs (Ingenuity Pathway Analysis (IPA) Target filter analysis tool [4]):

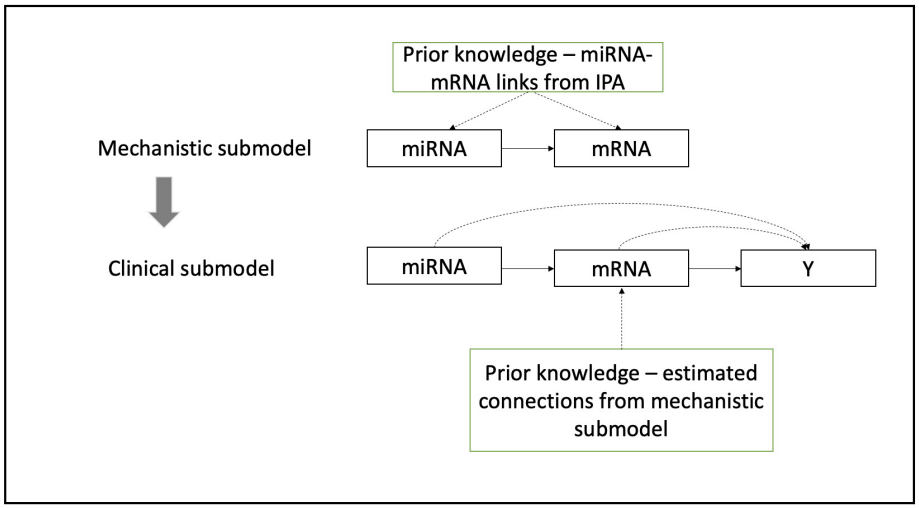
- Experimentally verified associations (score = 1),
- Predicted associations (scores = 0.75 for highly predicted, score = 0.5 for moderate predicted, and score = 0 otherwise).

Statistical model

Bayesian two-step intgrative procedure

Extension of the approach proposed by [5] and consists of two submodels:

- a mechanistic submodel: relating miRNAs and mRNAs,
- a clinical submodel: relating the phenotypic outcome to mRNA and miRNA expression levels



Bayesian two-step integrative procedure

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 840383.

Mechanistic submodel - Variable selection

Bayesian variable selection using spike-and-slab prior integrating prior knowledge from biological experiments

The expression level of gene j ($j = 1, \ldots, q$) is modeled by:

$$G_j = \text{miRNA}_1\beta_1 + \text{miRNA}_2\beta_2 + \cdots + \text{miRNA}_m\beta_m + \cdots + \text{miRNA}_p\beta_p + \varepsilon_j$$

- Prior inclusion probability of miRNA k: $\boldsymbol{p(\gamma_{jk}|\tau)} = \frac{\exp(\eta+\tau s_{jk})}{1+\exp(\eta+\tau s_{jk})}$, s_{jk} score between gene j and miRNA k, $\tau \sim \mathcal{G}(a, b)$ and η fixed

↬ **Selection** of the relevant miRNAs ($\beta_k \neq 0$ for $k = 1, \ldots, p$) while **integrating prior knowledge**

Decomposition of the mRNA expression level into two parts:

$$G_j = \underbrace{G_{miRNA}}_{\text{modulation via miRNAs}} + \underbrace{G_{\overline{miRNA}}}_{\text{modulation via other factors than miRNAs}}$$

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 840383.

Mechanistic submodel - Conditional Gaussian graph

Estimated graph

Gaussian graphical model (GGM) [6] used to estimate a graph structure for gene expressions adjusted for miRNAs ($G_{\overline{miRNA}}$):
$\hookrightarrow$ Covariate-adjusted Gaussian graph or conditional Gaussian graph or **adjusted** graph[7].

Note that we have also estimated an **undajusted** graph based on the raw gene expressions.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 840383.

Clinical submodel

Bayesian variable using spike-and-slab prior integrating prior knowledge from statistical analysis

Probit model where the linear predictor is given by:

$$\text{probit } P(Y=1) \sim G_{miRNA} + \underbrace{G_{\overline{miRNA}}}_{\text{Adjusted graph}} + \overline{\text{mRNA}} + \overline{\text{miRNA}} \quad (1)$$

with

- $\overline{\text{mRNA}}$: the set of mRNAs with no related miRNA in the mechanistic submodel,
- $\overline{\text{miRNA}}$: the set of miRNAs with no association to any of the mRNAs in the mechanistic submodel.

↬ **Selection** of the relevant mRNAs, miRNAs, and miRNA-mRNA pairs associated with HCC status while **integrating dependence structure**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 840383.

Outline

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 840383.

Mechanistic submodel

Spike-and-slab variable selection approach:

- 371 miRNA-mRNA pairs identified, 22 of which are experimentally verified.

Estimated graphs:

- before adjusting for miRNAs: 497 edges,
- after adjusting for miRNAS: 101 edges.

↬ Majority of gene-gene interactions not maintained after accounting for the miRNA regulation of these genes.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 840383.

Mechanistic submodel

Focus on 5 genes connected in the unadjusted graph and conditionnaly independent in the adjusted graph.

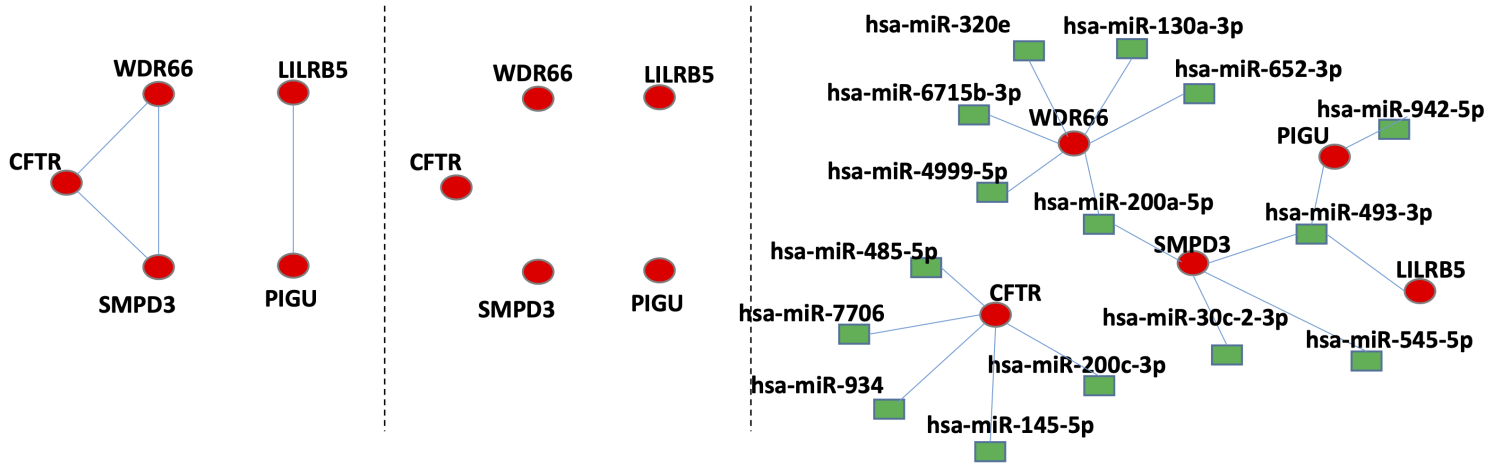


Figure 2: Unadjusted graph (left), adjusted graph (middle), and adjusted graph with associated miRNAs (right).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 840383.

Clinical submodel

Clinical submodel selects: 21 mRNAs, 5 miRNAs, and 66 miRNA-mRNA pairs

- 3 of the 66 pairs are experimentally verified.

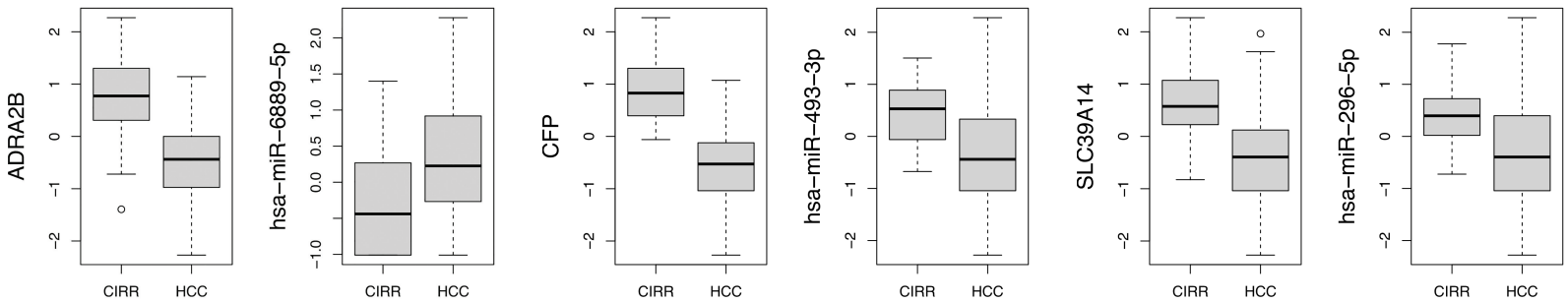
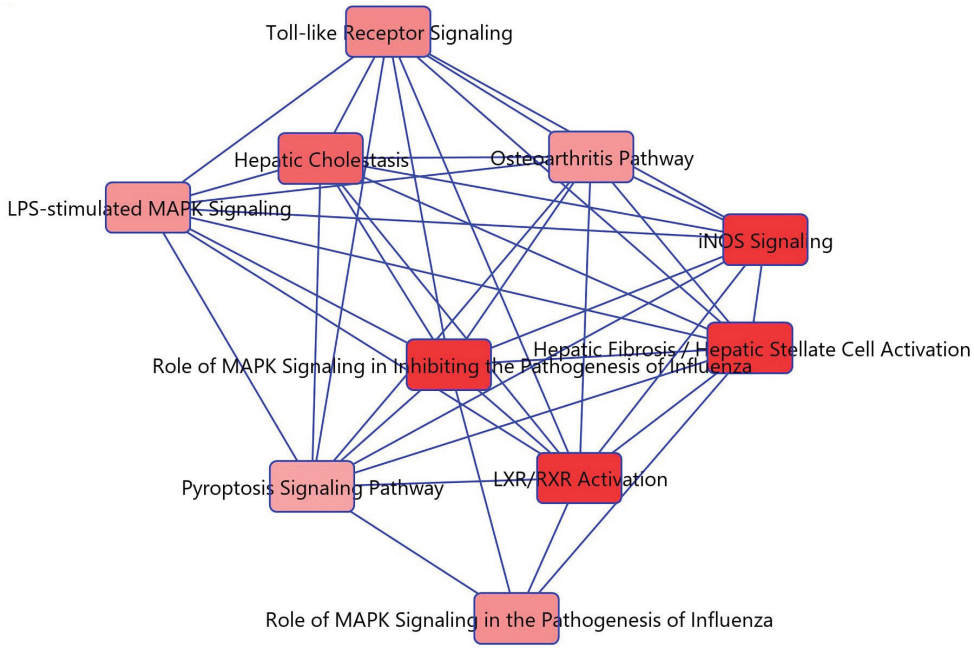


Figure 3: Boxplots of miRNA and mRNA expressions across disease status for three experimentally verified pairs.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 840383.

Clinical submodel: Pathway analysis using the IPA tool

Top 10 pathways represented by the molecules selected from the clinical model. Darker red color for pathways with higher significance.



↬ The three highest significant pathways are known to be significantly enriched in HCC

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 840383.

Outline

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 840383.

Conclusion/Perspectives

New Bayesian integrative approach extending [5] and incorporating knowledge from various sources at the different stages of the modeling for studying diseases.

Conclusion

- Adjusted/unadjusted graphs help improve the understanding of relationships between genes,
- The proposed approach helps narrow down to the most important mRNAs and miRNAs as well as miRNA-mRNA pairs,
- Key pathways identified.

↬ Biological relevance of studying molecular interactions and of integrating prior knowledge when analyzing mRNA-seq and miRNA-seq.

Perspectives

- Computational improvement for analyzing high-dimensional dataset,
- Findings need to be experimentally validated on larger dataset to confirm their potential as diagnostic or prognostic biomarkers,.

Thanks for your attention!

marie.denis@cirad.fr

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 840383.

Lin, Shuibin and Gregory, Richard I, MicroRNA biogenesis pathways in cancer, Nature reviews cancer, vol: 5(6), pp321333, 2015

Y. Chen, M. E. Barefoot, R. Varghese, K. Wang, C. Di Poto, H. W. Ressom, Integrative Analysis to Identify Race-Associated Metabolite Biomarkers for Hepatocellular Carcinoma, 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp 53005303, 2020

Zhaof, T., Liu, H., Roeder, K., Lafferty, J., and Wasserman, L., The huge package for high-dimensional undirected graph estimation in R, The Journal of Machine Learning Research, vol: 13(1), pp: 10591062, 2012

Causal analysis approaches in Ingenuity Pathway Analysis, Bioinformatics, vol: 30(4), pp: 523530, 2014

Denis, M. and Tadesse, M. G., Evaluation of hierarchical models for integrative genomic analyses, Bioinformatics, vol: 32(5), pp738746, 2016

A.P., Dempster, Covariance selection, Biometrics, pages 157175, 1972

Friedman, J., Hastie, T., and Tibshirani, R., Sparse inverse covariance estimation with the graphical lasso, Biostatistics, vol: 9(3), pp: 432441, 2008

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 840383.