Appendices
Inferring Interaction Networks using the IBP applied to microRNA Target Prediction

A Related work

Zhou et al. [3, 4] present a dependent hierarchical beta process using covariate-dependent features to impose that objects with similar covariates are likely to be clustered. The relationship between objects are summarized by a matrix $\mathbf{A}$ using a kernel $\mathcal{K}$. One way to apply this prior to our biological application would require converting the prior likelihood $\mathbf{C}$ matrix to the summary matrix, by defining a kernel over covariates. In contrast, our model avoids this requirement since all samples are drawn from a single process that encapsulates the dependencies.

B Taking the infinite limit

Lemma B.1. *For any real numbers $a_k (k \geq 1)$, which are constants with respect to n and $1 < T < \infty$,*

$$\lim_{n\to\infty}(1+\sum_{k=1}^{T}\frac{a_k}{n^k})^n = \exp(a_1) \tag{B.1}$$

Proof. The limit is in the indeterminate form 1^∞, we apply a transformation and L'Hôpital's rule:

$$\lim_{n\to\infty}(1+\sum_{k=1}^{T}\frac{a_k}{n^k})^n = \exp\lim_{n\to\infty}\frac{\ln(1+\sum_{k=1}^{T}\frac{a_k}{n^k})}{1/n} \quad \text{(transformation)} \tag{B.2}$$

$$= \exp\lim_{n\to\infty}\frac{-\sum_{k=1}^{T}\frac{ka_k}{n^{k+1}}/(1+\sum_{k=1}^{T}\frac{a_k}{n^k})}{-1/n^2} \quad \text{(L'Hôpital's rule)} \tag{B.3}$$

$$= \exp\lim_{n\to\infty}\frac{\sum_{k=1}^{T}\frac{ka_k}{n^{k-1}}}{1+\sum_{k=1}^{T}\frac{a_k}{n^k}} \tag{B.4}$$

$$= \exp(a_1) \tag{B.5}$$

□

Here we show that:

$$\lim_{K\to\infty}\frac{K!}{\prod_{h=0}^{2^N-1}K_h!}\prod_{k=1}^{K_+}\Phi_{z_k}\frac{B(m_k+\frac{\alpha}{K},N-m_k+1)}{B(\frac{\alpha}{K},N+1)}\prod_{k=1}^{K}\frac{1}{Z'}B(\frac{\alpha}{K},N+1) \tag{B.6}$$

$$= \frac{\alpha^{K+}}{\prod_{h=1}^{2^N-1}K_h!}\prod_{k=1}^{K_+}\Phi_{z_k}\frac{(N-m_k)!(m_k-1)!}{N!}\exp\big(-\alpha\Psi\big) \tag{B.7}$$

We consider each term separately.

$$\frac{K!}{\prod_{h=0}^{2^N-1} K_h!} \prod_{k=1}^{K_+} \Phi_{z_k} \frac{B(m_k + \frac{\alpha}{K}, N - m_k + 1)}{B(\frac{\alpha}{K}, N+1)} \tag{B.8}$$

$$= \frac{K!}{\prod_{h=0}^{2^N-1} K_h!} \prod_{k=1}^{K_+} \Phi_{z_k} \frac{\Gamma(\frac{\alpha}{K} + m_k)\Gamma(N - m_k + 1)}{\Gamma(\frac{\alpha}{K})\Gamma(N+1)} \tag{B.9}$$

$$= \frac{K!}{\prod_{h=0}^{2^N-1} K_h!} \prod_{k=1}^{K_+} \Phi_{z_k} \frac{(N - m_k)! \frac{\alpha}{K} \prod_{j=1}^{m_k-1}(j + \frac{\alpha}{K})}{N!} \tag{B.10}$$

$$= \frac{\alpha^{K+}}{\prod_{h=1}^{2^N-1} K_h!} \frac{K!}{K_0! K^{K+}} \prod_{k=1}^{K_+} \Phi_{z_k} \frac{(N - m_k)! \prod_{j=1}^{m_k-1}(j + \frac{\alpha}{K})}{N!} \tag{B.11}$$

By the same argument as shown in [2],

$$\lim_{K\to\infty} \frac{K!}{\prod_{h=0}^{2^N-1} K_h!} \prod_{k=1}^{K_+} \Phi_{z_k} \frac{B(m_k + \frac{\alpha}{K}, N - m_k + 1)}{B(\frac{\alpha}{K}, N+1)} \tag{B.12}$$

$$= \frac{\alpha^{K+}}{\prod_{h=1}^{2^N-1} K_h!} \prod_{k=1}^{K_+} \Phi_{z_k} \frac{(N - m_k)!(m_k - 1)!}{N!} \tag{B.13}$$

$$\prod_{k=1}^{K} \frac{Z'}{B(\frac{\alpha}{K}, N+1)} = \prod_{k=1}^{K} \frac{\sum_{h=0}^{2^N-1} \Phi_h B(\frac{\alpha}{K} + m_h, N - m_h + 1)}{B(\frac{\alpha}{K}, N+1)} \tag{B.14}$$

$$= \left(\sum_{h=0}^{2^N-1} \Phi_h \frac{\Gamma(\frac{\alpha}{K} + m_h)\Gamma(N - m_h + 1)}{\Gamma(\frac{\alpha}{K})\Gamma(N+1)} \right)^K \tag{B.15}$$

$$= \left(1 + \frac{\alpha}{K} \sum_{h=1}^{2^N-1} \Phi_h \frac{(N - m_h)! \prod_{j=1}^{m_h-1}(j + \frac{\alpha}{K})}{N!} \right)^K \tag{B.16}$$

$$= \left(1 + \frac{\alpha}{K} \sum_{h=1}^{2^N-1} \Phi_h \frac{(N - m_h)!(m_h - 1)!}{N!} + (\frac{\alpha}{K})^2 \cdots + \ldots \right)^K \tag{B.17}$$

Using Lemma B.1, we get:

$$\lim_{K\to\infty} \prod_{k=1}^{K} \frac{Z'}{B(\frac{\alpha}{K}, N+1)} = \exp\left(\alpha \sum_{h=1}^{2^N-1} \Phi_h \frac{(N - m_h)!(m_h - 1)!}{N!} \right) = \exp\left(\alpha\Psi\right) \tag{B.18}$$

Combining (B.13) and (B.18), we arrive at (B.7).

C The generative process

In Section 2.1.3, we described the generative process using a culinary metaphor. The customers select dishes one after the other as follows. The first customer tries Poisson($\alpha\Psi_1$) dishes. The remaining customers enter one after the others. Customer i selects dishes with a probability that partially depends on the selection of the previous customers. For each dish, the probability that it would be selected is specified by: $\sum_{h:h_i=z_{<ik},h(i)=1} \bar{\Phi}_h / \sum_{h:h_i=z_{<ik}} \bar{\Phi}_h$. He then samples a Poisson($\alpha\Psi_i$) number of new dishes. This process repeats until all customers have made their selections.

We show here that this process simplifies to the Indian Buffet Process when $\Phi_h = 1$ for all h.

Lemma C.1. *If $\Phi_h = 1$ for all h,*

$$\Psi_i = \frac{1}{i} \tag{C.1}$$

Therefore, each customer selects Poisson($\frac{\alpha}{i}$) new dishes as in the IBP.

Proof.

$$\Psi_i = \sum_{h:h_i=0,h(i)=1} \bar{\Phi}_h \tag{C.2}$$

$$= \sum_{h:h_i=0,h(i)=1} \frac{(N-m_h)!(m_h-1)!}{N!} \tag{C.3}$$

$$= \sum_{t=0}^{N-i} \binom{N-i}{t} \frac{(N-t-1)!t!}{N!} \tag{C.4}$$

$$= \frac{(i-1)!(N-i)!}{N!} \sum_{t=0}^{N-i} \binom{N-t-1}{i-1} \tag{C.5}$$

$$= \frac{(i-1)!(N-i)!}{N!} \frac{N\binom{N-1}{i-1}}{i} \tag{C.6}$$

$$= \frac{1}{i} \tag{C.7}$$

□

Lemma C.2. *If $\Phi_h = 1$ for all h,*

$$\frac{\sum_{h:h_i=z_{<ik},h(i)=1} \bar{\Phi}_h}{\sum_{h:h_i=z_{<ik}} \bar{\Phi}_h} = \frac{m_k}{i} \tag{C.8}$$

Therefore, each customer selects an old dish with probability $\frac{m_k}{i}$ as in the IBP.

Proof.

$$\sum_{h:h_i=z_{<ik},h(i)=1} \bar{\Phi}_h = \sum_{h:z_{<ik},h(i)=1} \frac{(N-m_h)!(m_h-1)!}{N!} \tag{C.9}$$

$$= \sum_{t=0}^{N-i} \binom{N-i}{t} \frac{(N-t-m_k-1)!(t+m_k)!}{N!} \tag{C.10}$$

$$= \frac{1}{(m_k+1)\binom{i}{m_k+1}} \tag{C.11}$$

$$\sum_{h:h_i=z_{<ik}} \bar{\Phi}_h = \sum_{h:z_{<ik}} \frac{(N-m_h)!(m_h-1)!}{N!} \tag{C.12}$$

$$= \sum_{t=0}^{N-i+1} \binom{N-i+1}{t} \frac{(N-t-m_k)!(t+m_k-1)!}{N!} \tag{C.13}$$

$$= \sum_{t=0}^{N-i+1} \binom{N-i+1}{t} \frac{(N-t-m_k)!(t+m_k-1)!}{N!} \tag{C.14}$$

$$= \frac{1}{m_k\binom{i-1}{m_k}} \tag{C.15}$$

Together,

$$\frac{\sum_{h:h_i=z_{<ik},h(i)=1} \bar{\Phi}_h}{\sum_{h:h_i=z_{<ik}} \bar{\Phi}_h} = \frac{m_k\binom{i-1}{m_k}}{(m_k+1)\binom{i}{m_k+1}} \tag{C.16}$$

$$= \frac{m_k}{i} \tag{C.17}$$

□

Furthermore, an equivalence class $[\mathbf{Z}]$ can be represented by a frequency vector $\mathbf{K} = (K_1, \ldots, K_{2^N-1})$. We can define a distribution on $\mathbf{K}$ by assuming that each K_h is generated independently by a Poisson distribution with parameters $\alpha\bar{\Phi}_h$. The probability is given by:

$$P(\mathbf{K}) = \prod_{h=1}^{2^N-1} \frac{(\alpha\bar{\Phi}_h)^{K_h}}{K_h!} \exp(-\alpha\bar{\Phi}_h) \tag{C.18}$$

This could be easily seen to be the same as Equation (8).

D GO results for clusters in Figure 4

Table D.1,D.2 and D.3 show the GO enrichment results for cluster (b), (c) and (f) in Figure 4 by GOstat[1]. We only show terms with corrected P-value less than 0.01. Cluster (a), (d) and (f) have no significant terms.

Term ID	Description	P value
GO:22402	cell cycle process	7.32E-05
GO:7049	cell cycle	1.77E-04
GO:22403	cell cycle phase	3.17E-04
GO:278	mitotic cell cycle	2.94E-03
GO:279	M phase	2.94E-03

Table D.1: GO enrichment analysis of cluster (b) in Figure 4

Term ID	Description	P value
GO:724	double-strand break repair via homologous recombination	3.29E-03
GO:725	recombinational repair	3.29E-03
GO:6281	DNA repair	3.29E-03
GO:6974	response to DNA damage stimulus	3.93E-03
GO:6310	DNA recombination	3.93E-03
GO:9314	response to radiation	3.93E-03
GO:9719	response to endogenous stimulus	3.93E-03
GO:51053	negative regulation of DNA metabolic process	3.93E-03
GO:8630	DNA damage response, signal transduction resulting in induction of apoptosis	4.13E-03
GO:6302	double-strand break repair	4.78E-03
GO:51052	regulation of DNA metabolic process	4.78E-03
GO:10212	response to ionizing radiation	4.78E-03
GO:9411	response to UV	6.28E-03
GO:7568	aging	7.34E-03
GO:8629	induction of apoptosis by intracellular signals	7.87E-03
GO:6996	organelle organization	9.95E-03
GO:9628	response to abiotic stimulus	9.95E-03
GO:42770	DNA damage response, signal transduction	9.95E-03

Table D.2: GO enrichment analysis of cluster (c) in Figure 4

Term ID	Description	P value
GO:45859	regulation of protein kinase activity	8.83E-03
GO:51338	regulation of transferase activity	8.83E-03
GO:165	MAPKKK cascade	8.83E-03

Table D.3: GO enrichment analysis of cluster (f) in Figure 4

E GenMiR++

Figure E.1 shows the network inferred by GenMiR++ with threshold of 0.9. We did not find any significant enrichment with corrected P-value less than 0.01.

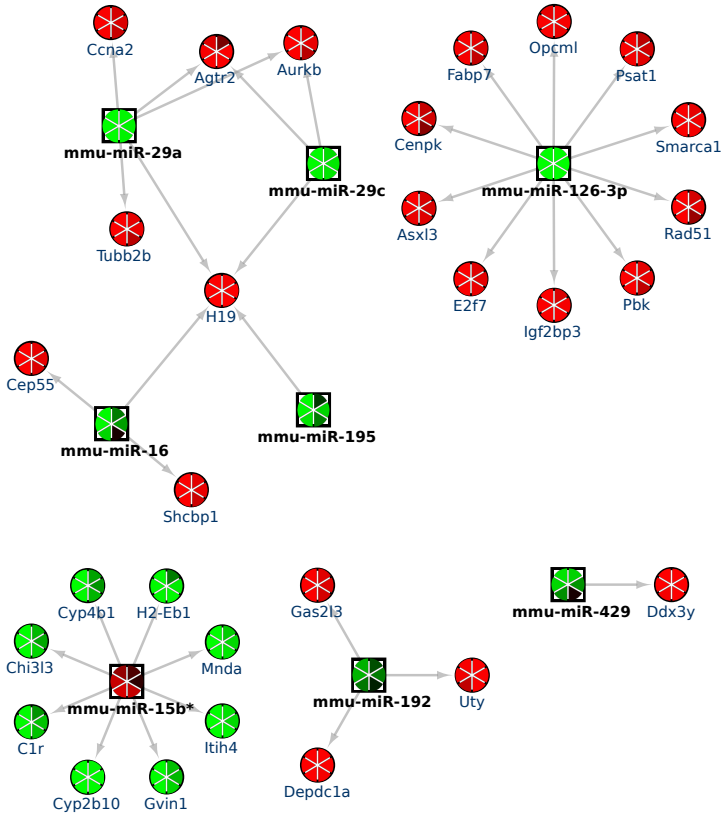

Figure E.1: GenMiR++ with threshold = 0.9

F Comparison with K-means

We ran K-means on the same set of mRNAs in Figure 4 using $k = 6$ as inferred by GroupMiR. We did not find any significant enrichment indicating that only by integrating sets of miRNAs with the mRNAs for this data we can find functional biological groupings.

G Comparison to IBP

We also tested with the original IBP ($\mathbf{W} = 0$). Not surprisingly, the results for both the synthetic and real data were weak (the IBP is of course not intended for our data since it cannot use the prior interaction information). Specifically, for the synthetic data the average F1 when using a noise level of 0.4 (a high but reasonable level) is 0.8418 for our method and only 0.5163 for the original IBP. For the real data, the IBP failed to recover any significant groupings. Without the priors the ability to identify significant interactions is greatly weakened.

H Networks at 60% posterior probability.

We also report networks constructed with 60% posterior probability by GroupMiR in Figure H.1 and 0.6 threshold by GenMiR++ in Figure H.2.

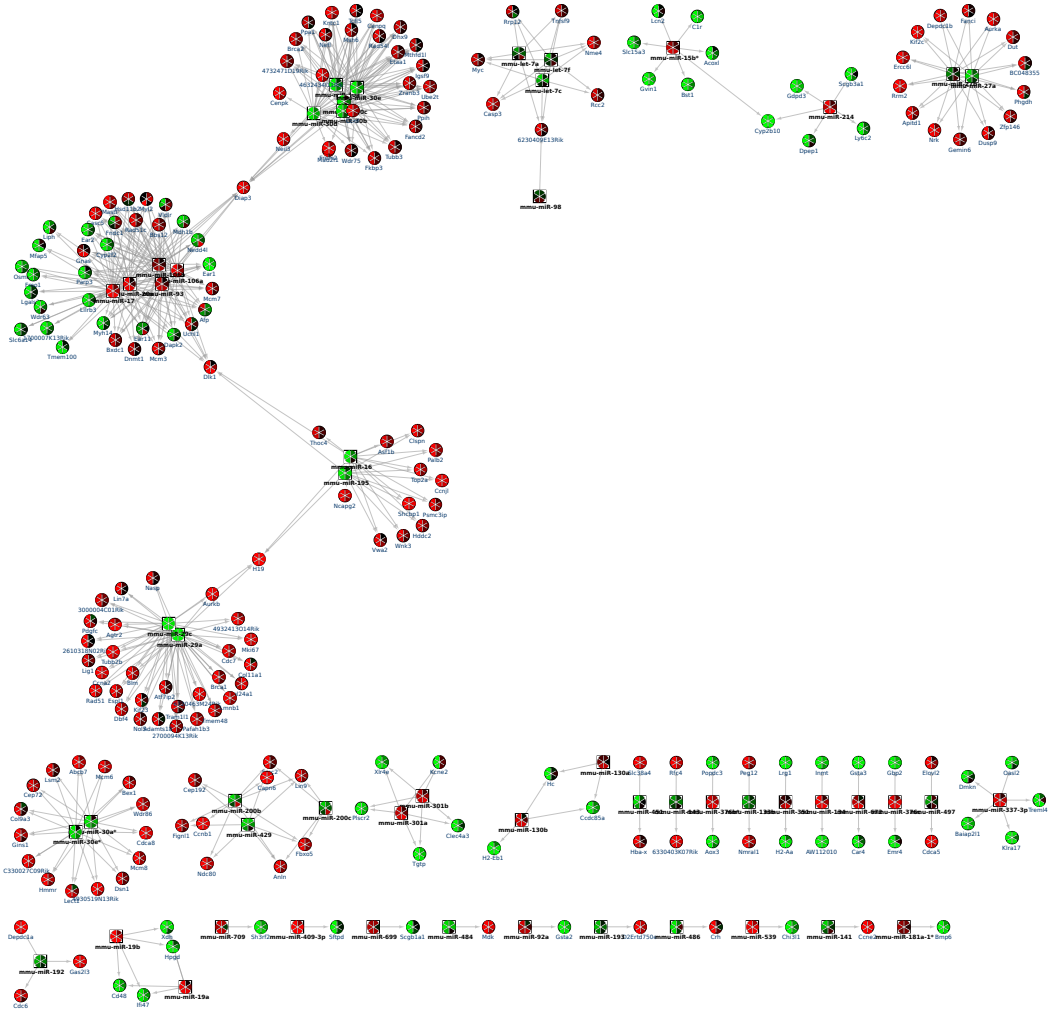

Figure H.1: Network inferred by GroupMiR with 60% posterior probability

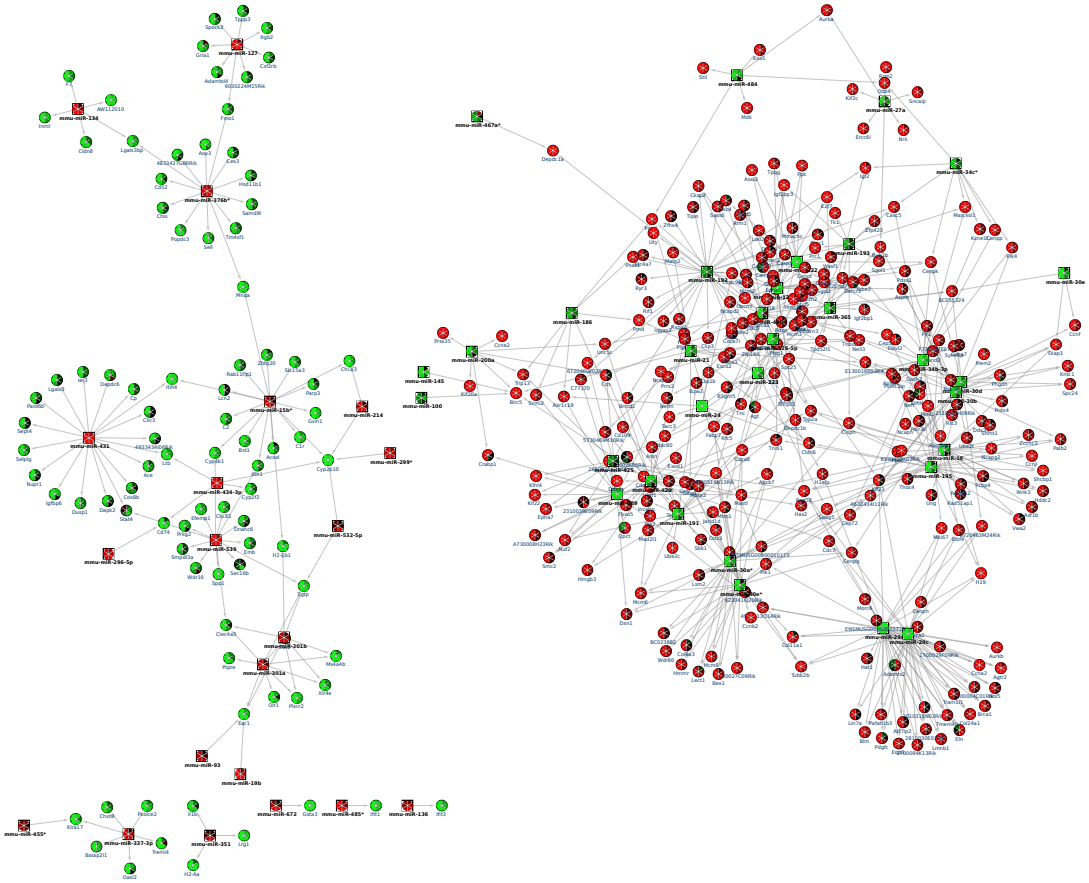

Figure H.2: Network inferred by GenMiR++ with threshold of 0.6

References

[1] T. Beißbarth and T.P. Speed. GOstat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics*, 20(9):1464, 2004.

[2] T. Griffiths and Z. Ghahramani. Infinite latent feature models and the Indian buffet process. *In Advances in Neural Information Processing Systems*, 18:475, 2006.

[3] M. Zhou, H. Yang, G. Sapiro, D. Dunson, and L. Carin. Dependent hierarchical beta process for image interpolation and denoising. *Proc. Artificial Intelligence and Statistics (AISTATS)*, 2011.

[4] Mingyuan Zhou, Hongxia Yang, Guillermo Sapiro, David B. Dunson, and Lawrence Carin. Covariate-dependent dictionary learning and sparse coding. In *ICASSP*, pages 5824–5827. IEEE, 2011.