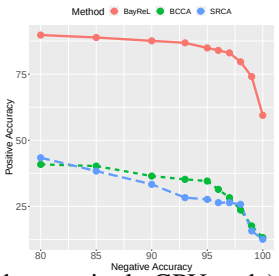
We truly appreciate helpful comments from all the reviewers. We will add suggested references and try our best to improve the presentation. We address the major ones here and would like first to emphasize that the main differences of BayReL from existing omics integrative analysis methods (including MOFA) are: 1) it infers relations in a unified formulation while the existing ones focus on integrating different data types to derive generic latent representations for downstream bioinformatics tasks; 2) it incorporates the graph structure at each view which is crucial for performance (see our response to **R2** below); 3) it learns the relations through non-linear/deep transformations of data as opposed to linear ones in most of the existing methods; 4) unlike co-embedding and matrix completion based methods, it infers relations between different molecular classes, **without any pre-known interactions across classes/views**.

[**R1** & **R4**] α in loss function is known as tempering [Huang, NeurIPS2018] in the context of Bayesian inference, which helps smoothing out the objective function by promoting modes that are close to those of the prior and is designed for better inference. We cannot set α to zero as it would remove the log prior term in our generative model. Appropriately tuned $\alpha > 1$ outperforms the original VAE training ($\alpha = 1$). Empirically on the CF dataset, the positive accuracy (PA, in %) of BayReL at 97% negative accuracy (NA) is 80.3, 82.3, 82.6, 82.7, and 81.2, with $\alpha = 1$, 10, 20, 30, and 50, respectively. We emphasize that, in BayReL, none of the between-view interactions are assumed to be *a priori* known in training. Changing the formulation to avoid using *learned* between-view interactions in prior construction, would change the model and the overall likelihood substantially, which we leave for future studies.

[**R3**] The extension of BCCA (only considering 2 views) to multiple views is known as Bayesian group factor analysis (GFA). In our experiments, we used the GFA implementation released in CCAGFA R package; but since we only had 2-view datasets, we referred to it as BCCA to avoid confusion. We should point out that mathematically, MOFA and GFA are similar, except that MOFA has been extended with Bernoulli and Poisson likelihoods for discrete omics datasets. We evaluate MOFA on the CF dataset, its PA, while NA is set to 97%, is 28.13%, very close to BCCA as we expected. For miRNA-mRNA, the prediction sensitivity of MOFA (in %) is 22.09, 32.95, & 42.17 for avg. deg. 0.2, 0.3, & 0.4, respectively. It slightly outperforms BCCA due to better modeling of RNA-seq count data.

[**R3**] The datasets we used have been extensively studied in the literature (e.g. MOFA studied precision medicine and multi-modal microbiome) and as pointed out in the paper (lines 239-241 & 283), they are of great importance in biology. We tried to pick three very different applications to show applicability of BayReL in biomedicine. Furthermore, these datasets have two types of heterogeneity: both graph and node attributes are different across views. The heterogeneity in the suggested RNA-seq/ChIP-seq would be only because of different types of node attributes (due to different technologies used for characterizing properties of genes/proteins). We also note that miRNA, mRNA, and gene-expression data in our 2nd and 3rd experiments, are indeed RNA-seq data. BayReL will be carefully evaluated (considering practical challenges to validate cross-view relations) with multi-view heterogeneous omics data in future.

[**R2**] We tested two modified versions of BayReL: 1) to show the importance of reconstructing graph structures, we removed the view-specific graph reconstruction (Equation 3) from the model (BayReL-NoRecon); 2) to show the importance of using view-specific graphs, we assume view-specific adjacency matrices are identity (BayReL-NoGraph). Applying to CF, BayReL-NoRecon has shown to be unstable in training with respect to random initialization (at 97% NA, PA can be as low as 10% for some seeds and as high as 79% for others). We argue that the reason for such a behavior is that removing graph reconstruction would cause the embedding to rotate arbitrarily with respect to views, which leads to poor performance. Adding view-specific graph reconstruction ensures that node embeddings are faithful to the view-specific graph structures as well as avoiding arbitrary rotations across views. BayReL-NoGraph also performs worse than BayReL on CF with PA of 44.7% at 97% NA. We note that BayReL-NoGraph outperforms BCCA.



[**R4**] Negative accuracy threshold: While we only discussed the results for one PA threshold (97%) in the paper for brevity, we can see similar improvements in other thresholds too (see the figure on the right for CF data). We note that there is a trade-off between PA and NA, and the optimal point is chosen based on the application. [**R4**] SRCA in Table 2: We did not include it in the original submission for better layout and readability. Prediction sensitivity of SRCA (in %) in TCGA for 25% and 50% of training samples are 25.53 (33.75) & 27.10 (35.79), respectively, for avg. deg. 0.3 (0.4). We will include a complete table in the supplement. Regarding computational complexity [**R2** & **R4**]: We have reported the run time of BayReL in the supplement (Section C). Our current implementation is relatively fast on mid-size graphs (∼15K nodes, <0.5 second per epoch on a single GPU node), indicating its scalability (with respect to the number of nodes, edges, and node attributes) to large datasets. In addition, by deploying sampling based GCNs, e.g. GraphSAGE, FastGCN, and FastGAE, the scalability can be further improved. Regarding accuracy measures [**R4**]: We haven't used F1 score because no true negatives are known in the datasets (only some true positives are known, typical in bioinformatics). Regarding φ^{sim} [**R4**]: We used an inner-product decoder (line 174 in the paper). Regarding hyperparameter selection [**R4**]: For BayReL, as most of the variational unsupervised methods, we chose them based on the training cost. For BCCA, it was described in the supplement (Section C).