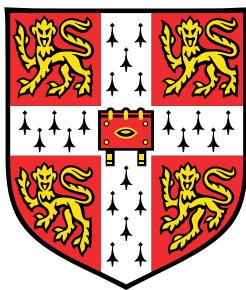


# Statistical framework for the integration of single-cell multi-omics data sets



**Ricard Argelaguet**

European Bioinformatics Institute  
University of Cambridge

This dissertation is submitted for the degree of  
*Doctor of Philosophy*



## Abstract

Single-cell profiling techniques have provided an unprecedented opportunity to study cellular heterogeneity at multiple molecular levels. The maturation of single-cell RNA-sequencing technologies has enabled the identification of transcriptional profiles associated with lineage diversification and cell fate commitment. This represents a remarkable advance over traditional bulk sequencing methods, particularly for the study of complex and heterogeneous biological processes, including the immune system, embryonic development and cancer. However, the accompanying epigenetic changes and the role of other molecular layers in driving cell fate decisions remains poorly understood. Profiling the epigenome at the single-cell level is receiving increasing attention. However, without associated transcriptomic readouts, the conclusions that can be extracted from epigenetic measurements are limited.

More recently, technological advances enabled multiple biological layers to be probed in parallel at the single-cell level, unveiling a powerful approach for investigating regulatory relationships. Such single-cell multi-modal technologies can reveal multiple dimensions of cellular heterogeneity and uncover how this variation is coupled between the different molecular layers, hence enabling a more profound mechanistic insight than can be inferred by analysing a single data modality in separate cells. Yet, multi-modal sequencing protocols face multiple challenges, both from the experimental and the computational front.

In this thesis we propose an experimental methodology and a computational framework for the integrative study of multiple omics in single cells.

The first contribution of this thesis is Nucleosome, Methylome and Transcriptome sequencing (scNMT-seq), a multi-modal single-cell sequencing protocol for profiling RNA expression, DNA methylation and chromatin accessibility in single cells. scNMT-seq provides genome-wide epigenetic readouts at a base-pair resolution, hence expanding our ability to investigate the dynamics of the epigenome across cell fate transitions.

The second contribution of this thesis is Multi-Omics Factor Analysis (MOFA), a statistical framework for the unsupervised integration of large-scale multi-omics data sets. MOFA aims at discovering the principal sources of variation while disentangling the axes of heterogeneity that are shared across multiple modalities from those specific to individual data modalities. This framework enables the unbiased interrogation of large (single-cell) data sets simultaneously across multiple data modalities and across different experiments or conditions.

The third contribution of this thesis is generation of an epigenetic roadmap of mouse gastrulation, resulting from the combined use of scNMT-seq and MOFA. Notably, we show that regulatory elements associated with the formation of the three germ layers are either epigenetically primed or epigenetically remodelled prior to overt cell fate decisions, providing the molecular logic for a hierarchical emergence of the primary germ layers.



## 0.1 Multi-Omics Factor Analysis: a framework for unsupervised integration of multi-omics data sets

The work described in this chapter results from a collaboration with the Multi-omics and statistical computing group lead by Wolfgang Huber at the EMBL (Heidelberg, Germany). It has been peer-reviewed and published in Argelaguet & Velten et al [Argelaguet2018].

The method was conceived by Florian Buettner, Oliver Stegle and me. I performed most of the mathematical derivations and implementation, but with significant contributions from Damien Arnol and Britta Velten. The single-cell application was led by me whereas the CLL data application was led by Britta Velten, but with joint contributions in either cases. Florian Buettner, Wolfgang Huber and Oliver Stegle supervised the project.

The article was jointly written by Britta Velten, Florian Buettner, Wolfgang Huber, Oliver Stegle and me.

### 0.1.1 Model description

MOFA is a multi-view generalisation of traditional Factor Analysis to  $M$  input matrices (or views)  $\mathbf{Y}^m \in \mathbb{R}^{N \times D_m}$  based on the framework of Group Factor Analysis (discussed in Section X).

The input data consists on  $M$  views with non-overlapping features that often represent different assays. However, there is flexibility in the definition of views and they can be tailored to address different hypothesis. Formally, the input data is factorised as:

$$\mathbf{Y}^m = \mathbf{Z}\mathbf{W}^{mT} + \boldsymbol{\varepsilon}^m \quad (1)$$

where  $\mathbf{Z} \in \mathbb{R}^{N \times K}$  is a matrix that contains the factor values and  $\mathbf{W}^m \in \mathbb{R}^{D_m \times K}$  are  $M$  matrices that contain the loadings that relate the high-dimensional space to the low-dimensional latent representation. Finally,  $\boldsymbol{\varepsilon}^m \in \mathbb{R}^{D_m}$  captures the residuals, or the noise, which is assumed to be normally distributed and heteroskedastic:

$$p(\boldsymbol{\varepsilon}_d^m) = \mathcal{N}(\boldsymbol{\varepsilon}_d^m | 0, 1/\tau_d^m) \quad (2)$$

Non-gaussian noise models can also be defined and is discussed in Section XX. Unless otherwise stated, we will always assume Gaussian noise.

Altogether, this results in the following likelihood:

$$p(\mathbf{Y}|\mathbf{W}, \mathbf{Z}, \mathbf{T}) = \prod_{m=1}^M \prod_{d=1}^{D_m} \prod_{n=1}^N \mathcal{N}(y_{nd}^m | \mathbf{z}_n^T \mathbf{w}_d^m, 1/\tau_d^m) \quad (3)$$

### Interpretation of the factors

Each factor ordiates cells along a one-dimensional axis centered at zero. Samples with different signs indicate opposite phenotypes, with higher absolute value indicating a stronger effect. For example, if the  $k$ -th factor captures the variability associated with cell cycle, we could expect cells in the Mitosis state to be at one end of the factor (irrespective of the sign, only the relative positioning being of importance). In contrast, cells in G1 phase are expected to be at the other end of the factor. Cells with intermediate phenotype, or with

no clear phenotype (i.e. no cell cycle genes profiled), are expected to be located around zero, as specified by the prior distribution.

### Interpretation of the loadings

The loadings provide a score for each gene on each factor, and are interpreted in a similar way as the factors. Genes with no association with the factor are expected to have values close to zero, as specified by the prior. In contrast, genes with strong association with the factor are expected to have large absolute values. The sign of the loading indicates the direction of the effect: a positive loading indicates that the feature is more active in the cells with positive factor values, and viceversa.

Following the cell cycle example from above, we expect genes that are upregulated in the M phase to have large positive loadings, whereas genes that are downregulated in the M phase (or, equivalently, upregulated in the G1 phase) are expected to have large negative loadings.

### Interpretation of the noise

The use of a probabilistic framework allows the model to explicitly disentangle the signal (i.e. the explained variance) from the noise (i.e. unexplained variance). Large values of  $\tau_d^m$  indicate high certainty on the observations for the feature  $d$  in view  $m$ , as predicted by the latent variables. In contrast, small values of  $\tau_d^m$  are indicative of low predictive power by the latent variables.

### Missing values

The probabilistic formalism naturally accounts for incomplete data matrices, as missing observations do not intervene in the likelihood.

In practice, we implement this using memory-efficient binary masks  $\mathcal{O}^m \in \mathbb{R}^{N \times D_m}$  for each view  $m$ , such that  $\mathcal{O}_{n,d} = 1$  when feature  $d$  is observed for sample  $n$ , 0 otherwise.

### Prior distributions for the factors and the loadings

The key determinant of the model is the regularization used on the prior distributions of the factors and the weights.

For the factors, we follow common practice [XX] and define an isotropic Gaussian prior:

$$p(z_{nk}) = \mathcal{N}(z_{nk} | 0, 1) \quad (4)$$

For the weights we encode two levels of sparsity, a (1) view- and factor-wise sparsity and (2) an individual feature-wise sparsity. The aim of the factor- and view-wise sparsity is to disentangle the activity of factors to the different views, such that the weight vector  $\mathbf{w}_{:,k}^m$  is shrunk to zero if the factor  $k$  does not explain any variation in view  $m$ .

In addition, we place a second layer of sparsity which encourages inactive weights on each individual feature. Mathematically, we express this as a combination of an Automatic Relevance Determination (ARD) prior [Mackay1996] for the view- and factor-wise sparsity and a spike-and-slab prior [25] for the feature-wise sparsity: However, this formulation of the spike-and-slab prior contains a Dirac delta function, which makes

the inference procedure troublesome. To solve this we introduce a re-parametrization of the weights  $w$  as a product of a Gaussian random variable  $\hat{w}$  and a Bernoulli random variable  $s$ , [36] resulting in the following prior: In this formulation  $\alpha_k^m$  controls the activity of factor  $k$  in view  $m$  and  $\theta_k^m$  controls the corresponding fraction of active loadings (i.e. the sparsity levels).

Finally, we define conjugate priors for  $\theta$  and  $\alpha$ :

$$p(\theta_k^m) = \text{Beta}(\theta_k^m | a_0^\theta, b_0^\theta) \quad (5)$$

$$p(\alpha_k^m) = \mathcal{G}(\alpha_k^m | a_0^\alpha, b_0^\alpha), \quad (6)$$

with hyper-parameters  $a_0^\theta, b_0^\theta = 1$  and  $a_0^\alpha, b_0^\alpha = 1e^{-3}$  to get uninformative priors.

Posterior values of  $\theta_k^m$  close to 0 implies that most of the weights of factor  $k$  in view  $m$  are shrunk to 0 (sparse factor). In contrast, a value of  $\theta_k^m$  close to 1 implies that most of the weights are non-zero (non-sparse factor). A small value of  $\alpha_k^m$  implies that factor  $k$  is active in view  $m$ . In contrast, a large value of  $\alpha_k^m$  implies that factor  $k$  is inactive in view  $m$ .

All together, the joint probability density function of the model is given by

$$\begin{aligned} p(\mathbf{Y}, \hat{\mathbf{W}}, \mathbf{S}, \mathbf{Z}, \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\tau}) = & \prod_{m=1}^M \prod_{n=1}^N \prod_{d=1}^{D_m} \mathcal{N} \left( y_{nd}^m | \sum_{k=1}^K s_{dk}^m \hat{w}_{dk}^m z_{nk}, 1/\tau_d \right) \\ & \prod_{m=1}^M \prod_{d=1}^{D_m} \prod_{k=1}^K \mathcal{N}(\hat{w}_{dk}^m | 0, 1/\alpha_k^m) \text{Ber}(s_{d,k}^m | \theta_k^m) \\ & \prod_{n=1}^N \prod_{k=1}^K \mathcal{N}(z_{nk} | 0, 1) \\ & \prod_{m=1}^M \prod_{k=1}^K \text{Beta}(\theta_k^m | a_0^\theta, b_0^\theta) \\ & \prod_{m=1}^M \prod_{k=1}^K \mathcal{G}(\alpha_k^m | a_0^\alpha, b_0^\alpha) \\ & \prod_{m=1}^M \prod_{d=1}^{D_m} \mathcal{G}(\tau_d^m | a_0^\tau, b_0^\tau). \end{aligned} \quad (7)$$

and the corresponding graphical model is shown in Figure 2. This completes the definition of the MOFA model.

### 0.1.2 Downstream analysis

Once trained, the MOFA model can be queried for a set of downstream analysis:

- **Variance decomposition:** calculate the variance explained ( $R^2$ ) by each factor in each view.
- **Ordination of the samples in the latent space:** scatterplots or beeswarm plots of factors, colored or shaped by sample covariates can reveal the main drivers of sample heterogeneity.

- **Inspection of loadings:** the weights (or loadings) can be interpreted as an activity score for each gene on each factor. Hence, inspecting the top loadings reveals the genes (or other genomic features) that underlie each factor.
- **Imputation:** MOFA generates a condensed and denoised low-dimensional representation of the data without missing values. As discussed in Section X, the data can be reconstructed from the latent space by a simple matrix multiplication:  $\hat{\mathbf{Y}} = \mathbf{Z}\mathbf{W}^T$ .
- **Feature set enrichment analysis:** when a factor is difficult to characterise based only on the inspection of the top loadings, one can compute a statistical test for enrichment of biological pathways using predefined gene-set annotations. The statistical tests that we implemented are outlined in Section X.

The downstream functionalities implemented in MOFA are highlighted in Figure 1.

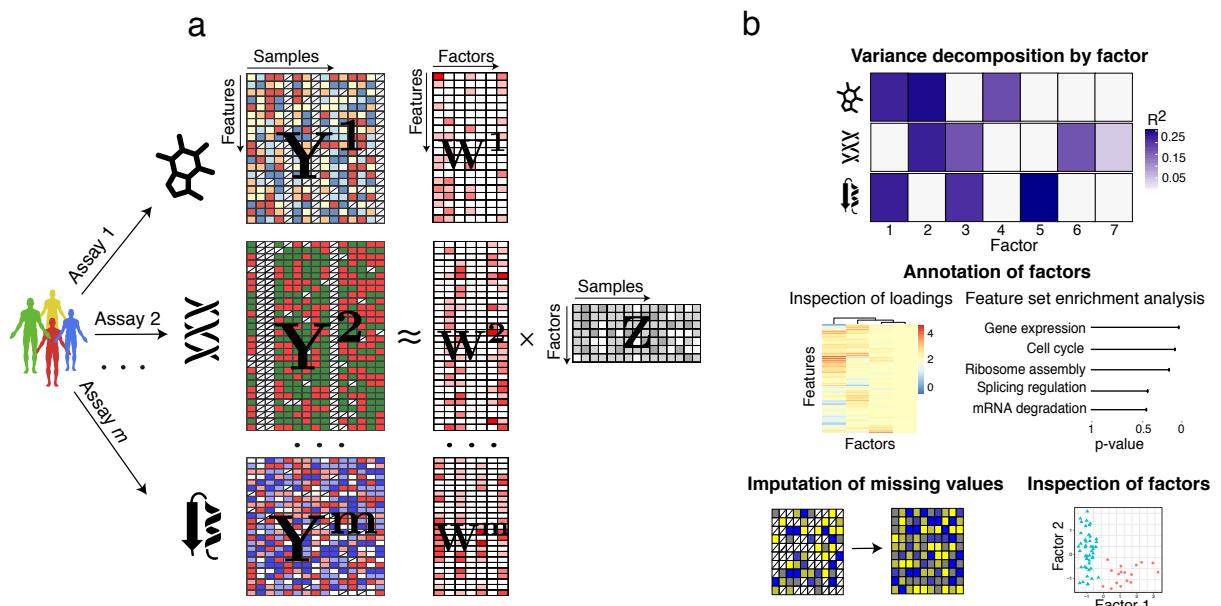


Fig. 1 MOFA overview. The model takes  $M$  data matrices as input ( $\mathbf{Y}^1, \dots, \mathbf{Y}^M$ ), one or more from each data modality, with co-occurring samples but features that are not necessarily related and can differ in numbers. MOFA decomposes these matrices into a matrix of factors ( $\mathbf{Z}$ ) and  $M$  weight matrices, one for each data modality ( $\mathbf{W}^1, \dots, \mathbf{W}^M$ ). White cells in the weight matrices correspond to zeros, i.e. inactive features, whereas the cross symbol in the data matrices denotes missing values. The fitted MOFA model can be queried for different downstream analyses, including a variance decomposition to assess the proportion of variance explained by each factor in each data modality.

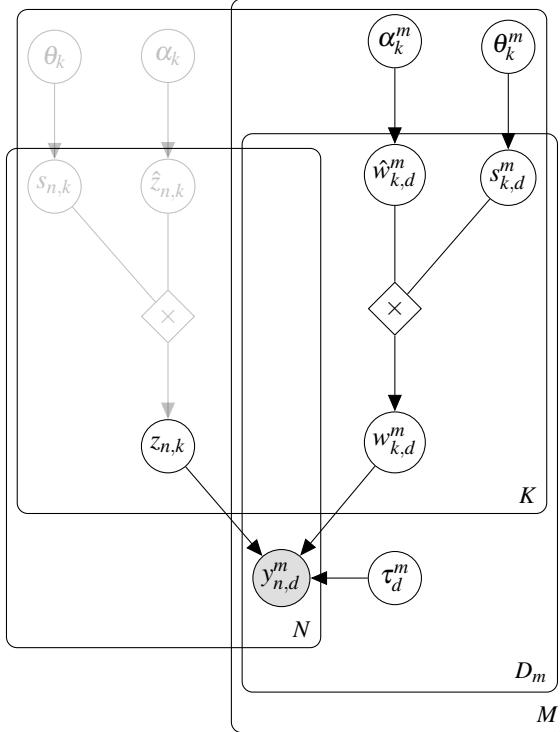


Fig. 2 Graphical model for MOFA. The white circles represent hidden variables that are inferred by the model, whereas the grey circles represent the observed variables. There are a total of four plates, each one representing a dimension of the model:  $M$  for the number of views,  $N$  for the number of samples,  $K$  for the number of factors and  $D_m$  for the number of features in view  $m$

## Inference

To make the model scalable to large data sets we adopt a Variational inference framework with a structured mean field approximation. A detailed overview is given in section XX, and details on the variational updates for the MOFA model are given in Appendix XX.

To enable efficient inference for non-Gaussian likelihoods we employ local bounds [XX]. This is described in detail in Section X

### 0.1.3 Monitoring convergence

An attractive property of Variational inference is that the objective function, the Evidence Lower Bound (ELBO), is required to monotonically increase at every iteration. This provides a simple way of monitoring convergence Figure 3. This is indeed one of the reasons why we selected this inference framework over Expectation Propagation or sampling-based approaches.

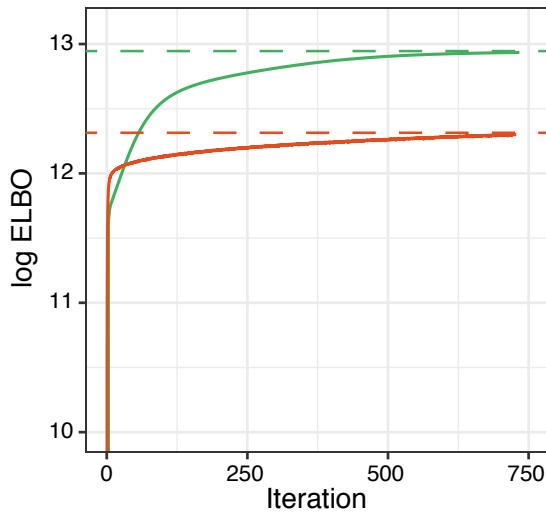


Fig. 3 XX

#### 0.1.4 Model selection and consistency across random initializations

The optimisation problem in MOFA is not convex and the algorithm is not guaranteed to find the optimal solution [XX]. Therefore, posterior distributions will vary depending on the initialisation and it becomes mandatory to perform model selection and assess the consistency of the factors across different trials.

The strategy we follow here is to train several MOFA models (e.g. 10 trials) under different parameter initialisations, and subsequently select the model with the highest ELBO for downstream analysis. In addition, we evaluate the robustness of the factors by plotting the Pearson correlations between factors across all trials. Figure 4.

A similar strategy has also been proposed in [Hore2016].

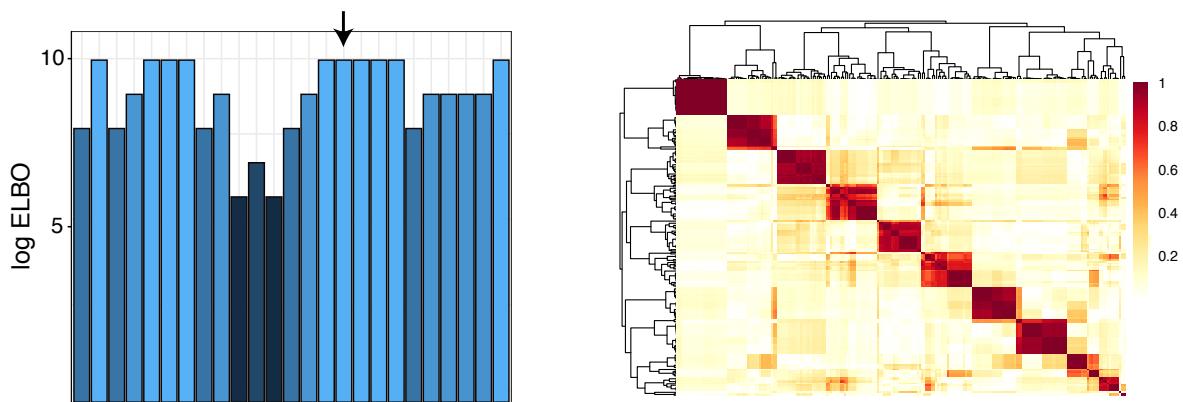


Fig. 4 XX... A block-diagonal matrix ensures good consistency across random initialisations

### 0.1.5 Learning the number of factors

As described in section X, the use of an ARD prior allows factors to be actively pruned by the model if they explain negligible variation. In the implementation this is controlled by a hyperparameter that defines a threshold on the minimum fraction of variance explained by a factor (across all views).

Additionally, because of the non-convexity of the variational inference algorithm, different model instances can potentially yield solutions with different number of active factors (??). The optimal number of factors need to be selected by the model selection strategy outlined in Section 0.1.4.

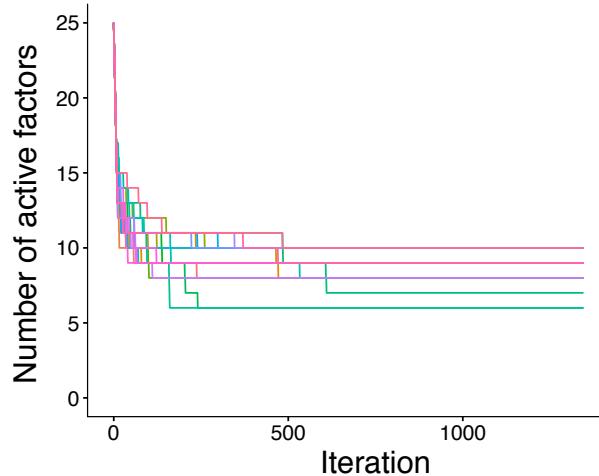


Fig. 5 XX

### 0.1.6 Application to single-cell multi-omics

The emergence of single-cell multi-modal techniques has created open opportunities the development of novel computational techniques that integrate data sets across multiple modalities [XX]. Here, we investigated the potential of MOFA to unravel the heterogeneity in one of the earliest single-cell multi-omics data sets [**Angermueller2016**].

The data set consists on 87 embryonic stem cells (ESCs) where RNA expression and DNA methylation were simultaneously profiled using single-cell Methylation and Transcriptome sequencing (scM&T-seq). Two populations of ESCs were profiled: the first one contains 16 cells grown in 2i media, which induces a native pluripotency state with genome-wide DNA hypomethylation [XX]. the second population contains 71 cells grown in serum media, which triggers a primed pluripotency state poised for differentiation [XX].

The RNA expression data was processed using standard pipelines [**Lun2016**] to obtain log normalised counts, followed by a selection of the top 5,000 most overdispersed genes [XX].

The DNA methylation data was processed as described in section XXX. Briefly, for each CpG site, we calculated a binary methylation rate from the ratio of methylated read counts to total read counts. Next, CpG sites were classified by overlapping with genomic contexts, namely promoters, CpG islands and enhancers (defined by the presence of distal H3K27ac marks). Finally, for each annotation we selected the top 5,000 most variable CpG sites with a minimum coverage of 10% across cells.

Each of the resulting matrices was input as a separate view to MOFA.

In this data set, MOFA learnt 3 factors (minimum explained variance of 2%). Factor 1 captured the transition from naive to primed pluripotent states, which MOFA links to widespread coordinated changes between DNA methylation and RNA expression. Inspection of the gene loadings for Factor 1 pinpoints important pluripotency markers including *Rex1/Zpf42* or *Essrb* [XX]. As previously described both in vitro (Angermueller et al, 2016) and in vivo (Auclair et al, 2014), the dynamics of DNA methylation are driven by a genome-wide increase in DNA methylation levels [XX]

Factor 2 captured a second axis of differentiation from a primed pluripotency state to a differentiated state (Figure 6b,c). e to a differentiated state with highest RNA loadings for known differentiation markers such as keratins and annexins

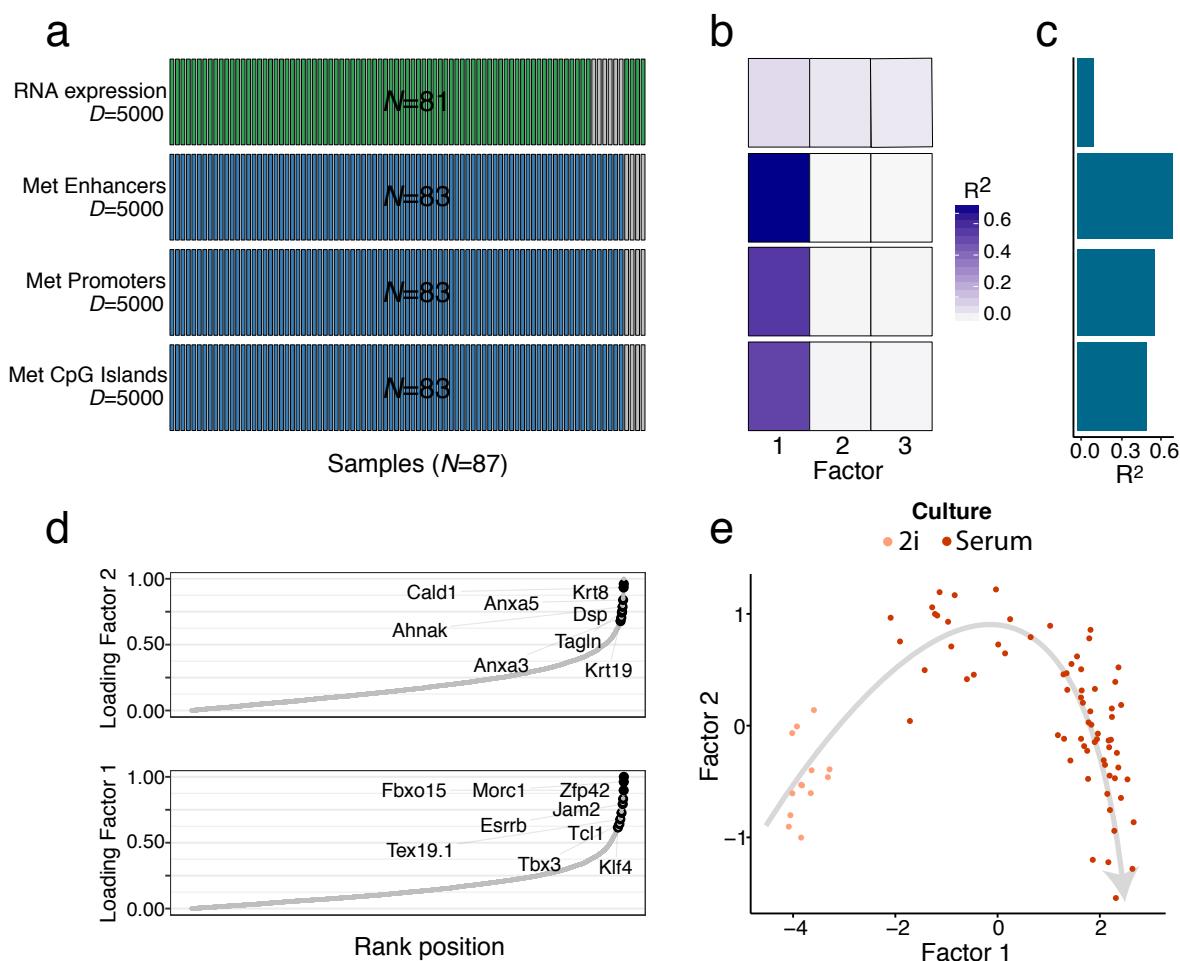


Fig. 6 XX

### 0.1.7 Open perspectives

# Bibliography

- [1] F. R. Bach and M. I. Jordan. *A probabilistic interpretation of canonical correlation analysis*. Tech. rep. 2005.
- [2] C. Bishop. “Variational Principal Components”. In: *Proceedings Ninth International Conference on Artificial Neural Networks, ICANN'99*. Vol. 1. IEE, Jan. 1999, pp. 509–514.
- [3] C. M. Bishop. “Bayesian PCA”. In: *Proceedings of the 1998 Conference on Advances in Neural Information Processing Systems II*. Cambridge, MA, USA: MIT Press, 1999, pp. 382–388. ISBN: 0-262-11245-0.
- [4] C. M. Bishop. “Pattern recognition”. In: *Machine Learning* 128 (2006), pp. 1–58.
- [5] K. Bunte et al. “Sparse group factor analysis for biclustering of multiple data sources”. In: *Bioinformatics* 32.16 (2016), pp. 2457–2463.
- [6] L. Dietz. *Directed Factor Graph Notation for Generative Models*. 2010.
- [7] C. Gao, C. D. Brown, and B. E. Engelhardt. “A latent factor model with a mixture of sparse and dense factors to model gene expression data with confounding effects”. In: *arXiv e-prints*, arXiv:1310.4792 (2013), arXiv:1310.4792. arXiv: 1310.4792 [stat.AP].
- [8] Y. Guo et al. “Sufficient Canonical Correlation Analysis”. In: *Trans. Img. Proc.* 25.6 (June 2016), pp. 2610–2619. ISSN: 1057-7149. DOI: 10.1109/TIP.2016.2551374.
- [9] W. Hardle and L. Simar. *Applied Multivariate Statistical Analysis*. Springer, 2007, pp. 321–30. ISBN: 978-3-540-72243-4.
- [10] Y. Hasin, M. Seldin, and A. Lusis. “Multi-omics approaches to disease”. In: *Genome Biology* 18.1 (2017), p. 83. DOI: 10.1186/s13059-017-1215-1.
- [11] H. Hotelling. “Analysis of a complex of statistical variables into principal components”. In: *Journal of Educational Psychology* 24.6 (1933), pp. 417–441.
- [12] H. Hotelling. “Relations between two sets of variates”. In: *Biometrika* 28.3-4 (Dec. 1936), pp. 321–377. ISSN: 0006-3444. DOI: 10.1093/biomet/28.3-4.321. eprint: <http://oup.prod.sis.lan/biomet/article-pdf/28/3-4/321/586830/28-3-4-321.pdf>.
- [13] S. Huang, K. Chaudhary, and L. X. Garmire. “More Is Better: Recent Progress in Multi-Omics Data Integration Methods”. In: *Frontiers in Genetics* 8 (2017), p. 84. ISSN: 1664-8021. DOI: 10.3389/fgene.2017.00084.
- [14] A. Ilin and T. Raiko. “Practical Approaches to Principal Component Analysis in the Presence of Missing Values”. In: *J. Mach. Learn. Res.* 11 (Aug. 2010), pp. 1957–2000. ISSN: 1532-4435.

- [15] S. A. Khan et al. “Identification of structural features in chemicals associated with cancer drug response: a systematic data-driven analysis”. In: *Bioinformatics* 30.17 (2014), pp. i497–i504.
- [16] A. Klami and S. Kaski. “Probabilistic approach to detecting dependencies between data sets”. In: *Neurocomputing* 72.1 (2008), pp. 39–46.
- [17] A. Klami et al. “Group factor analysis”. In: *IEEE transactions on neural networks and learning systems* 26.9 (2015), pp. 2136–2147.
- [18] S. Komili and P. A. Silver. “Coupling and coordination in gene expression processes: a systems biology view”. In: *Nat. Rev. Genet.* 9 (Jan. 2008), p. 38.
- [19] N. D. Lawrence et al. “Efficient inference for sparse latent variable models of transcriptional regulation”. In: *Bioinformatics* 33.23 (Aug. 2017), pp. 3776–3783. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btx508. eprint: <http://oup.prod.sis.lan/bioinformatics/article-pdf/33/23/3776/25168082/btx508.pdf>.
- [20] J. T. Leek and J. D. Storey. “Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable Analysis”. In: *PLoS Genet.* 3.9 (Sept. 2007), e161.
- [21] X. Li, B. Xiao, and X.-S. Chen. “DNA Methylation: a New Player in Multiple Sclerosis”. In: *Molecular Neurobiology* 54.6 (Aug. 2017), pp. 4049–4059. ISSN: 1559-1182. DOI: 10.1007/s12035-016-9966-3.
- [22] Y. Li, F.-X. Wu, and A. Ngom. “A review on machine learning principles for multi-view biological data integration”. In: *Briefings in Bioinformatics* 19.2 (Dec. 2016), pp. 325–340. ISSN: 1477-4054. DOI: 10.1093/bib/bbw113. eprint: <http://oup.prod.sis.lan/bib/article-pdf/19/2/325/25524236/bbw113.pdf>.
- [23] S. D. McCabe, D.-Y. Lin, and M. I. Love. “MOVIE: Multi-Omics VIualization of Estimated contributions”. In: *bioRxiv* (2018). DOI: 10.1101/379115. eprint: <https://www.biorxiv.org/content/early/2018/07/29/379115.full.pdf>.
- [24] C. Meng et al. “Dimension reduction techniques for the integrative analysis of multi-omics data”. In: *Brief. Bioinform.* 17.4 (July 2016), pp. 628–641.
- [25] T. J. Mitchell and J. J. Beauchamp. “Bayesian variable selection in linear regression”. In: *Journal of the American Statistical Association* 83.404 (1988), pp. 1023–1032.
- [26] R. M. Neal. *Bayesian learning for neural networks*. 1995.
- [27] M. Pilling. “Handbook of Applied Modelling: Non-Gaussian and Correlated Data”. In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 181.4 (2018), pp. 1264–1265. DOI: 10.1111/rssa.12402. eprint: <https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/rssa.12402>.
- [28] I. Pournara and L. Wernisch. “Factor analysis for gene regulatory networks and transcription factor activity profiles”. In: *BMC Bioinformatics* 8.1 (2007), p. 61.
- [29] M. Rattray et al. “Inference algorithms and learning theory for Bayesian sparse factor analysis”. In: *Journal of Physics: Conference Series* 197 (Dec. 2009), p. 012002. DOI: 10.1088/1742-6596/197/1/012002.
- [30] M. Ringnér. “What is principal component analysis?” In: *Nat. Biotechnol.* 26 (Mar. 2008), p. 303.
- [31] M. D. Ritchie et al. “Methods of integrating data to uncover genotype–phenotype interactions”. In: *Nature Reviews Genetics* 16 (Jan. 2015),

- [32] D. B. Rubin and D. T. Thayer. “EM algorithms for ML factor analysis”. In: *Psychometrika* 47.1 (1982), pp. 69–76.
- [33] O. Stegle et al. “Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses”. en. In: *Nat. Protoc.* 7.3 (Feb. 2012), pp. 500–507.
- [34] G. L. Stein-O’Brien et al. “Enter the Matrix: Factorization Uncovers Knowledge from Omics”. In: *Trends in Genetics* 34.10 (2018), pp. 790–805.
- [35] M. Tipping and C. Bishop. “Probabilistic Principal Component Analysis”. In: *Journal of the Royal Statistical Society* 61(3) (1999), pp. 611–22.
- [36] M. K. Titsias and M. Lázaro-Gredilla. “Spike and slab variational inference for multi-task and multiple kernel learning”. In: *Advances in neural information processing systems*. 2011, pp. 2339–2347.
- [37] S. Virtanen et al. “Bayesian group factor analysis”. In: *Artificial Intelligence and Statistics*. 2012, pp. 1269–1277.
- [38] C. Xu, D. Tao, and C. Xu. “A Survey on Multi-view Learning”. In: *arXiv e-prints*, arXiv:1304.5634 (Apr. 2013), arXiv:1304.5634. arXiv: 1304.5634 [cs.LG].
- [39] I. S. L. Zeng and T. Lumley. “Review of Statistical Learning Methods in Integrated Omics Studies (An Integrated Information Science)”. In: *Bioinformatics and Biology Insights* 12 (2019/03/16 2018), p. 1177932218759292.
- [40] Z. Zhang et al. “Opening the black box of neural networks: methods for interpreting neural network models in clinical applications.” In: *Annals of translational medicine* 6 11 (2018), p. 216.
- [41] S. Zhao et al. “Bayesian group factor analysis with structured sparsity”. In: *Journal of Machine Learning Research* 17.196 (2016), pp. 1–47.