Bioinformatics, 2018, 1–9
doi: 10.1093/bioinformatics/bty847
Advance Access Publication Date: 5 October 2018
Original Paper

Genome analysis

Multi-insight visualization of multi-omics data via ensemble dimension reduction and tensor factorization

Hadi Fanaee-T* and Magne Thoresen

Department of Biostatistics, University of Oslo, Oslo 0317, Norway

*To whom correspondence should be addressed.

Associate Editor: John Hancock

Received on March 21, 2018; revised on August 15, 2018; editorial decision on October 1, 2018; accepted on October 4, 2018

Abstract

Motivation: Visualization of high-dimensional data is an important step in exploratory data analysis and knowledge discovery. However, it is challenging, because the interpretation is highly subjective. If we see dimensionality reduction (DR) techniques as the main tool for data visualization, they are like multiple cameras that look into the data from different perspectives or angles. We can hardly prescribe one single perspective for all datasets and problems. One snapshot of data cannot reveal all the relevant aspects of the data in higher dimensions. The reason is that each of these methods has its own specific strategy, normally based on well-established mathematical theories to obtain a low-dimensional projection of the data, which sometimes is totally different from the others. Therefore, relying only on one single projection can be risky, because it can close our eyes to important parts of the full knowledge space.

Results: We propose the first framework for multi-insight data visualization of multi-omics data. This approach, contrary to single-insight approaches, is able to uncover the majority of data features through multiple insights. The main idea behind the methodology is to combine several DR methods via tensor factorization and group the solutions into an optimal number of clusters (or insights). The experimental evaluation with low-dimensional synthetic data, simulated multi-omics data related to ovarian cancer, as well as real multi-omics data related to breast cancer show the competitive advantage over state-of-the-art methods.

Availability and implementation: https://folk.uio.no/hadift/MIV/ [user/pass via hadift@medisin.uio.no]

Contact: hadift@medisin.uio.no

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Data visualization is one of the most important components of exploratory data analysis. Given a large set of measured variables, but few observations, it is useful to have a meaningful representation of reduced dimensionality. In an ideal scenario the representation should have a dimension that corresponds to the intrinsic dimension of the data.

Visualization of high-dimensional genomic data is a central ingredient of exploratory data analysis in biology and medical research, with the goal of making sense of the data before proceeding with more goal-oriented modeling and analyses. This is usually done via dimensionality reduction (DR) techniques.

Data visualization can be considered as a special case of DR where the dimension is limited to 2 or 3 (the dimension that the

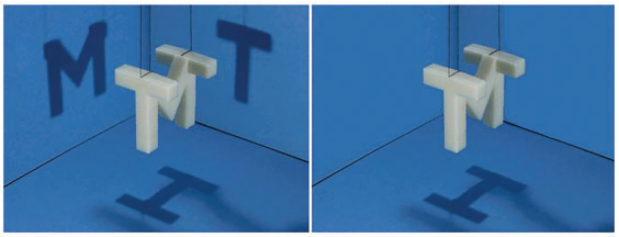

Fig. 1. Multi-insight (left) versus Single-insight visualization (right) (Design: T. Shannon) Copyright 2010 Creative Commons - Attribution - Share Alike license http://www.thingiverse.com/thing:3252

human brain can process). The main objective of visualization is to provide interpretable plots (which we will refer to as insights) to be used by the analyst for exploratory purposes.

Although visualization and DR are related concepts, there are some inherent differences. DR can be formulated as an optimization task where the goal it to optimize the similarity between points in the low- and the high-dimensional space. This optimized solution can be obtained with any price, even with low separation power in the low-dimensional projection. This is the reason why a given DR technique does not necessarily offer the best projection or visualization and sometimes we need to try many such techniques to have a satisfactory and interpretable result. DR methods focus on faithfulness to the high-dimensional space while the focus of visualization techniques is on providing an interpretable picture, understandable for the human brain.

Another problem with DR techniques is that they look for a single optimized solution. But in reality, if we look at objects from different angles we may perceive varying impressions. An intuitive example can be Figure 1a where a 3D object (MIT sculpture) is projected into a lower-dimension from different directions. In this case, either of projections *M*, *T* or *I* are correct. A single-insight DR technique (Fig. 1b) in this example needs to make a hard decision and pick one of the projections (e.g. I or M or T) and leave the other two undiscovered. The reason for this problem is that single-insight approaches are restricted to have only one output. It might happen that the letters *T* and *M* get recognized irrelevant by the analyst. But the better strategy is to leave this decision to the analyst and not letting the method decide.

Another widespread problem is that people rely on one given method and come to a consensus on using that particular approach, even if another approach may make more sense for the given problem.

Yet another issue is that we live in the era of *methods overload*. There are tons of methods for DR and data visualization and the number is continuously increasing. There is of course no statistics about the number of developed techniques, but we could list at least 150 of them only for DR. The question is if we need more techniques? Have we explored all the potential of the already developed techniques? Equally important is whether we are able to extract maximum information from our data. Maybe an idea could be to let techniques interact and give us a collaborative solution. Our objective is to answer some of these questions. We do not propose a new DR technique, rather we demonstrate how to combine different DR techniques to give us different insights from the data and how this multi-insight approach outperform some of the state-of-the-art single-insight methods.

To the best of our knowledge this is the first effort towards a multi-insight approach. Also, this is the first time that a systematic framework is proposed for combining different DR techniques.

Table 1. Comparison of methods for high-dimensional data visualization

Dataset	Single-view	Multi-view	Multi-insight
Visualization	✓	✓	✓
Integrating multiple sources	×	✓	✓
No. of patterns	1	2	≥2
Hypothesis discovery	×	×	✓

In Table 1, a comparison is given between single-view, multi-view and the multi-insight techniques. Single-view methods are those traditional DR techniques that can handle single-view data (e.g. only gene expression [GE]), while multi-view methods are developed for the analysis of datasets with multiple representations (e.g. integrating GE and DNA methylation [MET]). The multi-insight approach is beyond the multi-view, in the sense that it cannot only integrate multiple datasets, but also provides multiple insights of the results.

Our multi-insight approach not only provides 2D/3D visualization and multi-source capability, but it can also visualize multiple patterns and discover various cluster structures. Multi-view methods can be considered a special case of multi-insight techniques since they can provide the dominant projection for each view (No. of insights = 1). Besides this, the multi-insight approach enables us to make new hypotheses from the output while the other two do not offer such features.

Note that multiple projections with slight differences are not necessarily equal to multiple insights. For instance, in Figure 1, looking at the 3D object from the bottom viewpoint with -5 or $+5$ degrees variation does not give us any new insight. It just gives us a new view on the same insight, which still does not uncover any information about the other aspects of the object (i.e. M and T). An ideal multi-insight visualization should provide multiple *distinct* snapshots of the data. It should also determine the optimal number of insights. For instance, in Figure 1 the number of optimal insights is three (the whole knowledge space is M, T and I). It means that if we have a lower number of insights we miss some part of total knowledge (e.g. the right picture of Figure 1, which presents only one insight and therefore misses M and T). Likewise, if we end up with a higher number of insights, this is considered redundant information and the chance of introducing more noise to our decision space will be increased.

2 Approach

Our proposed approach consists of several steps. Figure 2 demonstrates the proposed approach on a toy example of a two-view dataset (e.g. GE and DNA MET). Note that we will use the generic term 'view' to refer to different inputs, e.g. data sources.

The central engine of the method is an ensemble of DR methods (Fig. 2-1). $DR_i(i = 1, 2, ..., m)$ can be various DR methods or a single DR method with different combinations of input parameters. In this work we focus on the former. For the latter case suppose that method *M* has *r* input parameters. We can make several DRs by different combinations of *M*'f parameters. We should select our ensemble in such a way that it maximizes the diversity.

After selection of the DR ensemble we generate 2D projections of the data via each of the selected DR methods (Fig. 2-2). Two-dimensional projections are the most common type of data visualization, simply because they are easier for human brain to interpret. Note that we can use any arbitrary number of components, but since

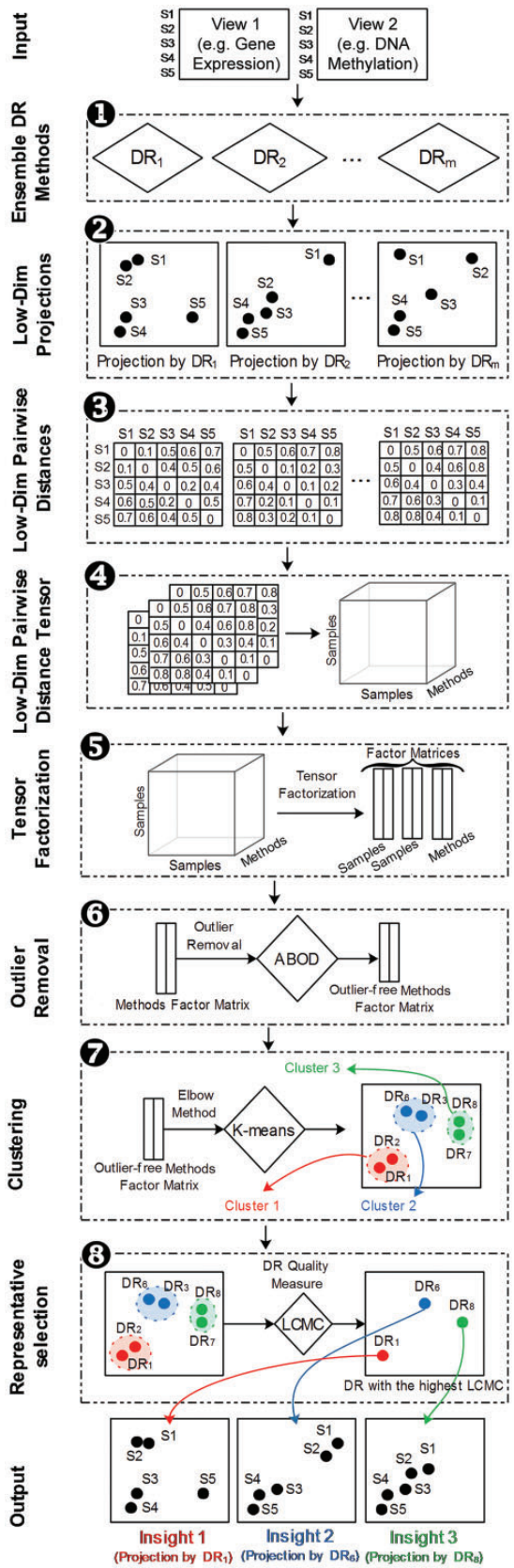

Fig. 2. The proposed apporach. The process is shown for view 1 in the picture but is the same for View 2 as well

our desired output is 2D plots, keeping the number of components equal to two avoids unnecessary re-computations. However, the drawback of 2D projections is that they cannot discriminate between two equally important patterns. Although this barely happens in practice, in such cases the user can go for 3D plots. Subsequently, the user should set the number of components to three and use the appropriate visualization tools to inspect the patterns from different angles.

If the data naturally has only one insight it is very likely that the majority of methods end up with a unified pattern. When the number of insights exceeds one (e.g. left picture of Fig. 1) the divergence between the methods becomes higher and we will have different projections that lead to multiple insights.

In order to obtain the optimal insights, we need to somehow combine the projections obtained by different DR methods. We cannot combine the projection matrices directly, mainly due to the non-linear similarity problem. For instance, suppose that the data has two Clusters A and B. The members of these two clusters can pose in space in an indefinite number of different ways. In order to solve this problem we borrow the idea of non-linear learning where we present the projections in the format of a pairwise distance matrix (Fig. 2-3) instead of raw projections. However, the scale of distances is still different for the different methods. Hence, we perform some scaling after this step to make the distance matrices homogeneous.

Now the problem is reduced to clustering m pairwise distance matrices. One approach to do this is to combine all these matrices via sample-wise concatenation and then apply a clustering algorithm like k-means. But in that case, the simultaneous inter-connection and extra-connections between data-points across different DR solutions are lost. Besides, the dimensionality of the matrix becomes larger, which creates problems for algorithms like k-means.

A better solution for preserving the inter/extra relationships in the pairwise distances is to model the data with tensors. This is Step 4 of our method and is depicted in Figure 2-4. Tensors are mathematical data models that allow us to represent the data in its natural structure without any information loss. This kind of data modeling has a long history rooted in psychometrics, chemometrics and signal processing. For a detailed survey about tensors refer to (Papalexakis *et al.*, 2016).

To analyze tensor data we use a group of techniques called tensor factorizations (*TF*s). They are similar to matrix factorization with some differences in the properties and also in the limitation in the number of data dimensions. When data has a tensor structure in the sense that more than two phenomena are simultaneously influencing the data items, TF will be a more accurate data model compared with matrix factorization (Fanaee-T and Gama, 2016).

*TF*s can model all intra/interactions between multiple dimensions of undergoing processes at the same time. The most common application of TFs is in the analysis of time-evolving networks (i.e. tensor of $node \times node \times time$) which is a natural form of tensor-structured data. Time and interactions between nodes are two sources of variation that can simultaneously influence each other. TFs are perfect for modeling such complex interactions.

Our pairwise distance tensor ($sample \times sample \times DR_{m}ethod$) is very similar to the time-evolving graph tensors. The similarity between samples is equivalent to nodes and the DR methods' equivalent is time. After modeling the data in the tensor format we decompose the pairwise distance tensor via CP TF (Carroll and Chang, 1970; Harshman,R.A. unpublished work). The reason why we prefer CP over Tucker or other models is that CP contrary to Tucker provides a unique solution under mild conditions

(Papalexakis *et al.*, 2016). Uniqueness is an essential property for our multi-insight methodology. Besides, we want to operate on the methods, factor matrix for clustering purpose. This is much easier in CP than in Tucker since in CP we do not have the core tensor and factor matrices may directly be used as input to a clustering algorithm.

The CP model (Fig. 2-5) decomposes the tensor $sample \times sample \times DR_method$) to three factor matrices of sizes $n \times p$, $n \times p$ and $m \times p$ such that n represents the number of samples and p is the number of components in the CP model and m is the number of DR methods in the ensemble. The first two matrices represent the decomposition of the sample mode and the third matrix represents the DR methods mode.

Some of DR methods might provide meaningless projections due to some reason, for instance false parameter tuning or unpredictable bugs in implementations. Therefore, we need to remove the outlier projections from the factor matrix before we proceed. In an ideal setting when the DR methods work perfectly on all configurations and datasets we may skip the outlier removal step. However, in practice, implementation of DR techniques sometimes exhibit odd results on certain datasets and configurations. These odd projections, if included into the model, might ruin the result of clustering algorithms like k-means that are sensitive to outliers. Of course, some of these odd projections can be novelties, but we believe that when we have a sufficient number of diverse DR methods, that kind of novelties should naturally appear as new insights in the clustering output, even with a low support. A naive solution to this problem might be to retrieve the outliers as an output of the method for further inspection. However, the problem is that if a given pattern is supported by for instance only one method it would be difficult to discriminate between this and an unwanted outlier.

Our choice for outlier removal is an algorithm called angle-based outlier detection or ABOD (Kriegel *et al.*, 2008). ABOD assesses the variance in the angles between the difference vectors of a point to the other points. Contrary to major outlier detection algorithms, ABOD is parameter-free and consequently this keeps our method parameter-free as well. We generate an outlier-free factor matrix by excluding the top 5% of DR methods identified by the ABOD algorithm (Fig. 2-6). This outlier-free matrix will be the input to the clustering algorithm.

If we want to find k distinct insights we need to find k distinct clusters of DR methods. The DR factor matrix contains a summary of intra/extra (non-linear) relationships between samples in low-dimensional space. Therefore, applying a clustering algorithm to the DR factor matrix (outlier-free one) gives our desired answer. There are many approaches to perform clustering, but we prefer a parameter-free approach to keep our method parameter-free too. We propose to use k-means clustering along with an automated version of the Elbow method to tune the number of clusters (the k parameter in k-means). In order to compute the best k we compute the distortion under different numbers of clusters counting from 1 to 10. The parameter k is the cluster number corresponding to 90% of variance explained. Given k we apply k-means clustering to obtain the distinct groups of DR methods. Each of the obtained clusters in Figure 2-7 correspond to one insight. For instance, DR_1 and DR_2 correspond to Insight 1 and DR_3 and DR_6 correspond to Insight 2.

Note that due to freedom of k we might have too many insights which may overwhelm the user. In order to alleviate this problem the user can simply set k manually in the k-means algorithm or set a lower maximum ceiling in the automated k-means to control the number of insights.

The next step is how to pick the best representative visualization of the insight group. For instance, in the above example, among DR_1 and DR_2 which one should represent the insight? We propose to use the local continuity meta-criterion (LCMC) from Chen and Buja (2009) (implementation in coranking R package) to pick the best representative projection (Fig. 2-8). LCMC is a parameter-free quality measure for DR. It can be defined as the average number of overlaps between R-nearest neighborhoods in the high-dimensional space and the low-dimensional projection. If we plot LCMC versus varying R (neighborhood size) we can define two quality criteria (Lee and Verleysen, 2010) called QL and QG which are defined as the mean respectively, over the LCMC values left and right to the maximum of the plot (R_{max}). QL represents the quality of DR with respect to local structure preservation while QG accounts for global structure. QL is considered to be more important than QG because preserving the local structure is of higher importance and is also more difficult. Therefore, we here consider QL as our quality criterion.

We compute the QL for each projection in each cluster and pick the one with the highest QL. The visualization obtained from this projection is the output of our method (Fig. 2 Output).

Alternatively, if we pick the best overall projection among all DR methods based on the QL, the solution simply converts to a single-insight method. An intuitive example of this can be picking only one of the projections M, I or T in Figure 1. In summary, what our solution does is that it first estimates the optimal number of insights (i.e. 3 in Fig. 1) and then choose the best visualization for each insight (M, I, and T). So our approach provides more information compared with a single-insight approach.

3 Experimental evaluation

In order to demonstrate the effectiveness of the proposed approach we design three experiments. The first experiment (Hollow Sphere Dataset) aims at showing the effectiveness of the proposed approach in comparison to existing approaches on visualization of low-dimensional data. Since the original data have dimension three and the low-dimensional projection has dimension two we can judge the effectiveness of the approaches in a more tangible manner. The second experiment (Simulated Ovarian Cancer Dataset) is a simulation performed to generate multi-omics data related to ovarian cancer with five hypothetical cancer sub-types. The goal of this experiment is to see how the proposed approach performs compared with state-of-the-art methods in reflecting the true structure of data in higher dimensions. Finally, the third experiment (Breast Cancer Dataset) is designed to assess the performance of our proposed approach in a real scenario, again compared with existing techniques when the goal is sub-type discovery based on high-dimensional multi-omics data. Note that although our case studies are based on mRNA expression/DNA MET, the framework is general and can operate with any multi-view datasets.

3.1 Datasets

3.1.1 Hollow sphere dataset

Yet Another Fishbowl Dataset (Silva and Tenenbaum, 2003) is usually used for evaluation of nonlinear DR techniques. It consists of a hollow sphere with a hole on top. In order to generate these data, we first sample θ from a uniform distribution with min $= 0$ and max $= 2\pi$. Then z is sampled from a uniform distribution with min $= -1$ and max $= 0.8$ (0.2 less for the hole). Finally, x and y are computed as $x = \sqrt{(1-z^2)}\cos\theta$ and $y = \sqrt{(1-z^2)}\sin\theta$. We set

$N = 1000$ as the number of samples and generate (x, y, z). However, since our focus in this work is on multi-view datasets we generate a multi-view version of the above dataset following a trick proposed by Stražar *et al.* (2016). All positive values are stored in the positive matrix (first view), while absolute values of negative values are stored in the negative matrix (second view). We call the final multi-view dataset '*Hollow Sphere*' for further reference.

3.1.2 Simulated ovarian cancer dataset

InterSIM is an R package developed by Chalise *et al.* (2016). It is a simulation tool for integrative multi-omics datasets. As it is described, it simulates three types of interrelated omics data (DNA MET, mRNA GE, and protein expression) with realistic intra- and inter-relationships between features of the same type extracted from The Cancer Genome Atlas (TCGA) ovarian cancer study. Additionally, InterSIM can simulate any arbitrary number of clusters which makes it an ideal tool for evaluation of integrative clustering or DR algorithms. We simulate a dataset with pairs of mRNA GE and DNA MET. The configurations we use for the simulation are as follows (refer to Chalise *et al.*, 2016 for detailed descriptions). (i) Number of samples: 100; (ii) Number of true clusters: 5 with the proportion of samples in the clusters: (0.18, 0.19, 0.20, 0.21 and 0.22); (iii) Cluster mean shift for GE and MET: 2.5; (iv) proportion of Differentially Expressed CpG sites: 0.2. The rest of the parameters are set as NULL according to the packageac default. We finally add noise of SNR = 1 (signal-to-noise ratio) to the output matrices to make it a more difficult problem for the DR methods.

3.1.3 Breast cancer dataset

This dataset is extracted from the R package Bayesian Consensus Clustering (https://github.com/ttriche/bayesCC) packaged up from the work of Lock and Dunson (2013). It is a small subset of the full dataset that is publicly available data from TCGA. Four views (sources) of data are available for the same set of 348 breast cancer samples. The available views are mRNA (645 genes), DNA MET (574 CpG sites), miRNA (423 miRNAs) and proteomics (171 proteins). We pick two views from this set: mRNA GE and DNA MET.

3.2 Data pre-processing

Some of the DR methods are from the family of non-negative matrix factorization (NMF) and accepts only positive values as input. Thus, the following pre-processing step is performed on the data to make the elements positive. First, the absolute value of the minimum of the matrix is added to all elements and then all elements are divided by the maximum value.

3.3 Compared methods

We compare our proposed method with six established methods from the literature: integrative orthogonal nonnegative matrix factorization (iONMF) (Stražar *et al.*, 2016), joint non-negative matrix factorization (JNMF) (Zhang *et al.*, 2012), Multiple Co-Inertia Analysis (MCOA) (Chessel and Hanafi, 1996), probabilistic canonical correlation analysis (PCCA) (Klami *et al.*, 2013), NMF (Berry *et al.*, 2007), and principal component analysis (PCA). The former four approaches are multi-view methods and the latter two are single-view methods. Note that for PCA and NMF we apply them on each data source separately without concatenation. We denote these projections by S1 and S2 not to confuse it with V1 and V2 of multi-view methods.

3.4 DR methods in the ensemble

Our ensemble includes 99 DR methods from different categories: DR methods, NMF, joint matrix factorization, JNMF, multi-block data methods, Bayesian Multi-Block models and Joint/Separated Matrix Factorization. The full list of the selected methods is available in Supplementary Appendix S1. Also the family and type (multi-view or single-view) of method, the link to download the code and corresponding references are listed in Supplementary Appendix. The reader can also refer to Li *et al.* (2018), Meng *et al.* (2016) (multi-view) and Mokbel *et al.* (2013) (single-view) for further overview of the majority of these methods. The input parameters of the DR methods are chosen based on the default parameters suggested by the authors in their implementations (see Supplementary Appendix S2). Note that though each individual method might be sensitive to the choice of input parameters, we assume that the ensemble should be robust to this sensitivity because of the outlier removal phase described earlier.

For regular DR techniques without multi-view support we concatenate the matrices for each datasource/view in a sample-wise manner and create a larger single matrix. After computation we again separate the projections corresponding to each view, so that for two views we would have two projection matrices. For methods like JIVE, COBE or O2PLS etc. that provide individual and joint components we make projections corresponding to both joint and individual components and pick the projection that presents the higher QL.

Note that the output of our method is presented as individual projections for each view. This is because a large number of methods in our ensemble do not provide a joint representation (i.e. methods from the family of DR and NMF), and this is the only solution if we want to include all of them together. However, this does not necessarily mean that those methods do not take into account the interactions between views. Rather, all methods, irrespective of the technical differences do consider interactions among the views via the concatenation trick. They only differ in the output. For some methods a joint representation for all views is presented while for others individual projections are provided for each view. Even for some methods like JIVE, both representations are provided simultaneously. Of course, one can exclude methods with non-joint representations from the ensemble. In that case our ensemble would include a limited number of methods and hence the result of our ensemble would be less reliable. Another drawback is that we cannot benefit from the diversity provided from other non-joint representations. Of course, in the future, by development of more multi-view methods one can make an ensemble with such an arrangement.

3.5 TF algorithm

We tried several CP TF algorithms in toolboxes such as TensorLab (https://www.tensorlab.net). CP-ALS (Carroll and Chang, 1970; Harshman,R.A. unpublished work) was chosen due to its simplicity and popularity. We did not find any other approaches with any specific advantages over CP-ALS. The number of components is set to two.

4 Results

The output of our multi-insight approach can be used for different purposes. In the following subsections we attempt to demonstrate some of these purposes, including intuitive interpretation (on the hollow sphere dataset), hypothesis making (on simulated ovarian cancer dataset) and multi-insight visualization (on the breast cancer dataset).

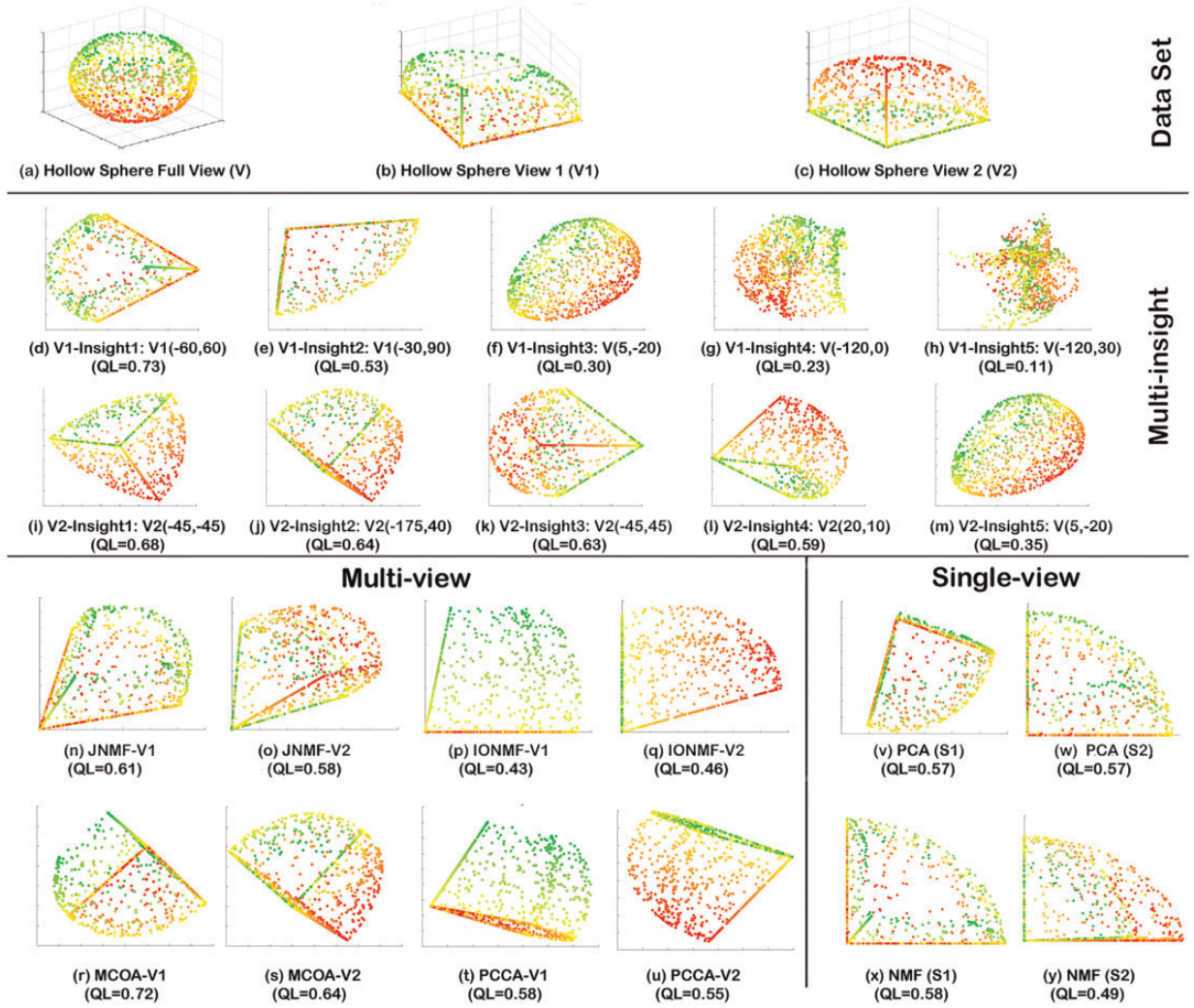

Fig. 3. Visualization results with Hollow Sphere Dataset. The X and Y axis in the figures represents respectively the first and second principal axis in the lower dimensional projection. In order to track the data points in lower-dimension the datapoints are colored from green to red according on their position on the z axis (the top green, bottom red and middle yellow) (Color version of this figure is available at *Bioinformatics* online.)

4.1 Results with hollow sphere dataset

Figure 3 shows the input dataset and the visualization results corresponding to our proposed method, as well as the multi-view and single-view methods on the Hollow Sphere Dataset. V1 (Fig. 3b) corresponds to the first view of data (positive elements) and V2 (Fig. 3c) corresponds to the second view (negative elements). Figure 3a represents the full view of the input dataset (before separation of positive and negative elements). The proposed method automatically finds five insights for each view. Figure 3d–h corresponds the top-five insights for V1 and Figure 3i–m corresponds to the top-five insights for V2.

We are, in particular interested in understanding which of the obtained projections (or insights) corresponds to what viewpoint(s) if we place a hypothetical camera above the Hollow Sphere. An ideal result is when we retrieve completely different viewpoints. A viewpoint (specified in the figure after colon mark for each projection) is defined by the horizontal rotation about the z-axis and vertical elevation of the viewpoint in degrees and obtained in MATLAB by rotating the 3D object in different directions. For instance, the viewpoint of V1-Insight1 (Fig. 3d) is V1(−60, 60). It means that the obtained projection corresponds to a viewpoint of (−60, 60) of V1. As another example, take V1-Insight3 (Fig. 3f). This corresponds to a viewpoint of V(5, −20). The retrieved viewpoints are in agreement with our ideal scenario. The majority of possible viewpoints that makes structural differences to our perception of the 3D object are detected by the proposed method.

One should observe that the majority of patterns obtained by state-of-the-art methods are covered in one of these insights. For instance, JNMF-V1 (Fig. 3n) and JNMF-V2 (Fig. 3o), respectively look like V1-Insight1 (Fig. 3d) and V2-Insight3 (Fig. 3k). Regarding non-NMF based multi-view methods, MCOA-V1 is similar to V2-Insight4 (Fig. 3l) and MCOA-V2 is produced in V2-Insight2 (Fig. 3j). Regarding PCCA, though we do not have any direct equivalent for PCCA-V1, an insight similar to PCCA-V2 can be found in V2-Insight4 (Fig. 3l).

The output of single-view methods are also presented in the figure. For instance, the output of PCA on V1 (Fig. 3v) and on V2 (Fig. 3w) are similar to V1-Insight2 (Fig. 3e) and V1-Insight1 (Fig. 3d), respectively. For NMF it is opposite. NMF on V1 (Fig. 3x) can be

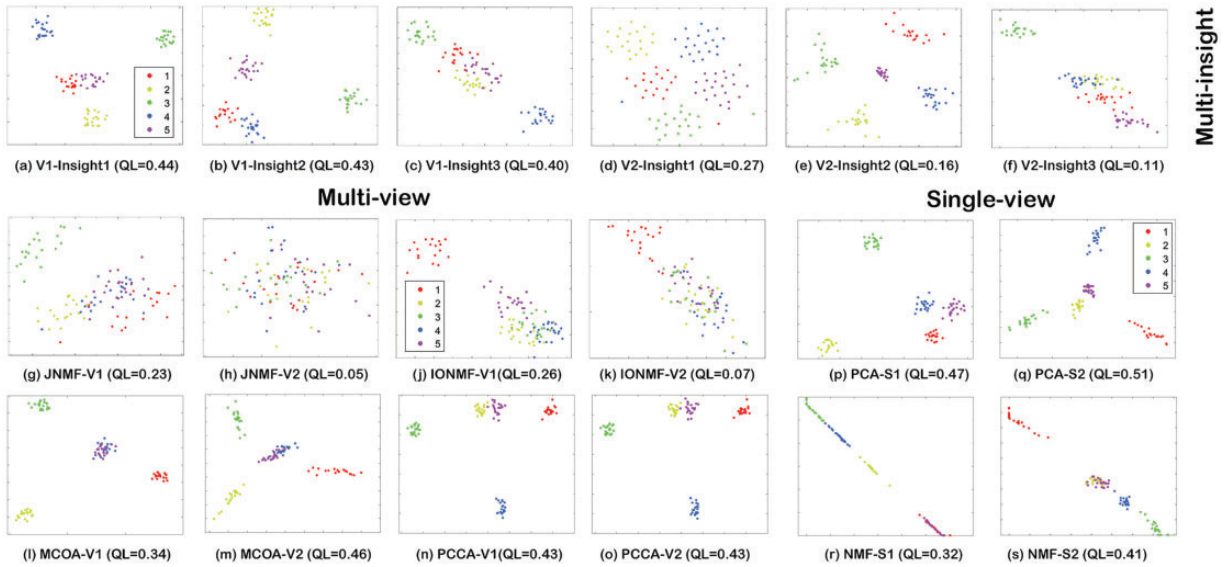

Fig. 4. Visualization results with simulated ovarian cancer. Number of clusters (cancer sub-types) in simulation is 5. V1: mRNA expression, V2: DNA MET. The X and Y axis in the figures represents respectively the first and second principal axis in the lower dimensional projection (Color version of this figure is available at *Bioinformatics* online.)

compared with V1-Insight1 (Fig. 3d) and a similar insight to NMF-S2 can be seen in V1-Insight2 (Fig. 3e).

An interesting point is that the multi-insight approach not only contains the collective knowledge of JNMF, IONMF, MCOA and PCCA as well as PCA and NMF but also has something extra to offer. Sometimes, this extra information or insight remains hidden if we rely solely on a single-insight method. The examples are V1-Insight3 (Fig. 3f) and V2-Insight5 (Fig. 3m). These projections imply that the higher-dimension data has the shape of a sphere, something which cannot be inferred from any of the projections provided by other methods.

The more careful observation of the output of multi-view and single-view methods shows that these approaches generate projections of the most likely narrative (with higher QL). Since they have to present only one snapshot, this should present the most important features of the data. Hence, they have to neglect secondary features that might uncover some hidden knowledge about the data. As we can see the QL corresponding to V1-Insight3 is relatively low (0.43 lower than the top insight) but it still contains an important fact about the input data (the shape). Therefore, maximizing quality (the goal of many DR techniques), although it identifies the most prevailing aspects of the data, does not provide any information about *other main characteristics*.

4.2 Results with simulated ovarian cancer dataset

The visualization results from the simulated ovarian cancer dataset are depicted in Figure 4. The method automatically selects three insights based on our k-means rule. From the simulation, we know that the number of true classes is five and thus a good visualization approach should be able to discriminate between these five clusters. Although the quintuplet cluster structure is the principal hypothesis, we are looking for other secondary hypotheses as well. We can extract the following hypotheses from our multi-insight visualization.

- Hypothesis 1: The data have five clusters. This is supported by V1-Insight1 (Fig. 4a), V1-Insight2 (Fig. 4b), V2-Insight1 (Fig. 4d), and V2-Insight2 (Fig. 4e).
- Hypothesis 2: Types 1 and 5 are more similar than others. This can be supported by V1-Insight1 (Fig. 4a), V1-Insight3 (Fig. 4c) and V2-Insight3 (Fig. 4f).
- Hypothesis 3: Types 1 and 4 are more similar than others. This can be supported by V1-Insight2 (Fig. 4b) and V2-Insight3 (Fig. 4f).
- Hypothesis 4: Types 1 and 2 are more similar than others. This can be supported by V1-Insight3 (Fig. 4c) and V2-Insight3 (Fig. 4f).
- Hypothesis 5: Type 3 is the easiest type to recognize. This is supported by all insights.

As we can see, the first hypothesis with the support of four out of six insights is selected as the principal hypothesis. Of course we know that this is equivalent to the ground truth as this is how we simulated the data. However, a secondary hypothesis is formed, which is supported by three insights (but not strongly); Types 1 and 5 are more similar than others. However, we know that the data are not simulated with such a pattern. Therefore, this apparent finding is probably due to the added noise. The same logic applies to the third and fourth hypothesis. Furthermore, we may sometimes conclude that some types are easier to detect. Our fifth hypothesis suggests that Type 3 is the most distinguishable type. Sometimes such knowledge can be helpful.

Now let us take a look at the results of state-of-the-art approaches. JNMF-V1 (Fig. 4g) suggests a triplet cluster structure: A: Type 1, B: Type 3 and C: Types 2, 4 and 5. JNMF-V2 (Fig. 4h) suggests no cluster structure at all which is as expected based on its low QL. From both views of IONMF (i.e. Fig. 4j and k) two clusters are inferred: A: Type 1 and B: Types 2–5. Regarding MCOA we observe that it suggests four clusters in such a way that Types 4 and 5 overlap in both views. Although PCCA-V1 can identify five clusters, it is difficult to distinguish between Types 2 and 5 in both views.

When NMF is applied separately on two views we observe that NMF-S1 identifies three clusters in such a way that Types 2 and 4 form one cluster and Types 5 and 1 are mixed together. NMF-S2 also shows four clusters where Types 5 and 2 are overlapping. However, it appears that PCA when applied on each view separately

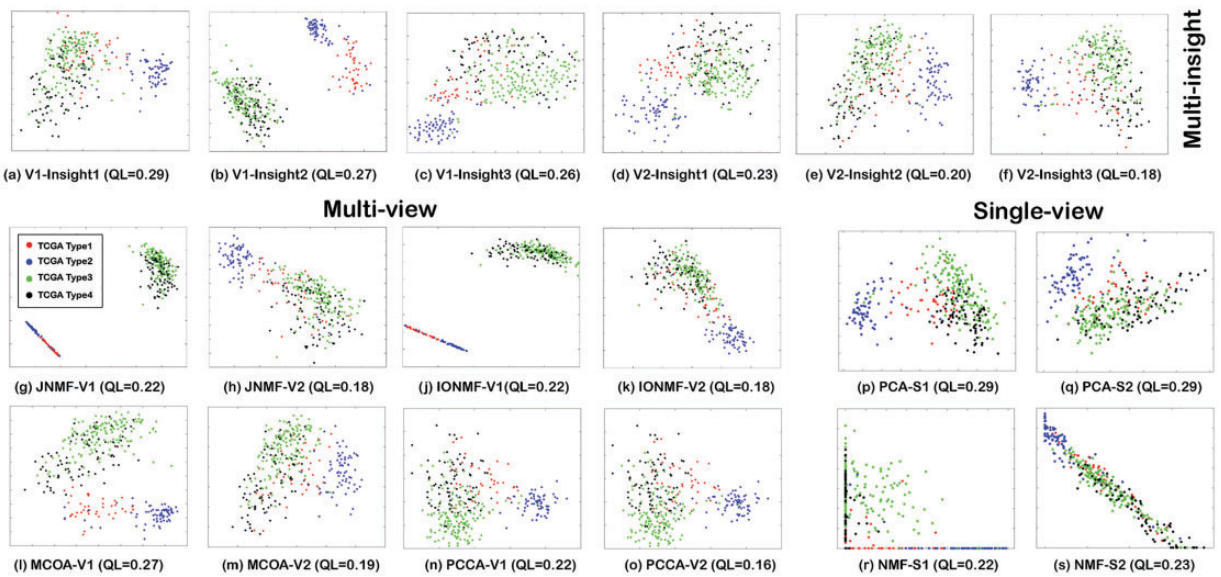

Fig. 5. Visualization result with breast cancer dataset (Red: TCGA Type1, Blue: TCGA Type2, Green: TCGA Type3, Black: TCGA Type 4). V1: mRNA expression, V2: DNA MET. The X and Y axis in the figures represents respectively the first and second principal axis in the lower dimensional projection (Color version of this figure is available at *Bioinformatics* online.)

performs well. Both PCA-S1 and PCA-S2 can identify five clusters. An interesting point is that although PCA provides better average QL compared with the insights provided by our method, the visualizations of some of our methodet output provides more informative and clearer discrimination between clusters. As we can see, V2-Insight1 and V2-Insight2 also identify five clusters like PCA-S1 and PCA-2 but the classes have been discriminated better in our insights. However, this behavior of PCA is still captured by one of our insights. For instance, PCA-S1 is equivalent to V1-Insight2.

As can be seen, there is a large divergence among the results obtained by state-of-the-art methods and the majority of them are not able to find the true structure of data in the low-dimensional projection. Only some of them have been able to pick up some of the secondary hypotheses, which can be considered another evidence for usefulness of the proposed method.

Note that in this case since our dataset is simulated we attributed the secondary hypotheses to the noise. In a real application these secondary hypotheses could refer to some hidden knowledge, and further investigation would be needed to sort this out.

4.3 Results with breast cancer dataset

TCGA identified *four classes* of breast cancer subtypes (TCGA Network, 2012) by integrating five different types of data; mRNA expression, DNA MET, miRNA expression, DNA copy number variation and protein expression. These four classes correlated well with the previously defined mRNA subtypes, which, according to the authors, suggest that information in DNA copy number variation, DNA MET and miRNA expression is captured at the level of gene and protein expression.

As was previously mentioned our objective of this experiment is to understand to what extent we are able to uncover the same pattern as was discovered in this high-dimensional cluster analysis, based on only two data sources; mRNA expression (V1) and DNA MET (V2) and our 2D multi-insight approach.

Figure 5 depicts the visualization of the multi-insight method in comparison to multi- and single-view methods. Regarding the multi-insight approach, in each view the method has automatically picked three insights.

Based on these insights we can relatively easy pick up three clusters; Types 1–4. In particular, V1-Insight2 (Fig. 5b) with the second highest QL clearly identifies the three clusters. We can also observe that is not easy to separate between Types 3 and 4. This is as expected, as these are related to Lum A and B which are typically difficult to separate also in high-dimensional clustering. However, several of our insights point in direction of a separation also of these types (see Fig. 5a, c and d) and by combining cluster membership information from multiple insights, we can come a long way on identifying also these types.

Note that though we base our conclusions on only three insights, each insight represents an output of a rather large set of DR methods. Thus, we do not actually conclude based on one projection, rather our conclusion is based on a summary of multiple methods.

Now, let us look at the results produced by other selected approaches. We can make the following statements.

- JNMF: The first view is dominant due to higher QL. From this view (Fig. 5g) we are able to separate Types 1 and 2 from Types 3 and 4. The second view (Fig. 5h), however cannot help to separate further. Thus, the overall conclusion is two clusters: A: Types 1 and 2 and B: Types 3 and 4. Note that at a first glance it might seem that Type 2 is separable from the rest of the points in Fig. 5h. However, if we plot the data without label information (i.e. all points in black) no clear border between Type 2 and others can be observed and all the points look contiguous.
- IONMF: The same as JNMF
- MCOA: It seems that MCOA-V1 has been able to detect three clusters and this view is similar to V1-Insight2. However, if we look closer, the separation between Types 1 and 2 is better done in V1-Insight2 compared with MCOA-V1. From MCOA-V2 we can infer two clusters, similar to V1-Insight1.
- PCCA: Both views produce an insight similar to V1-Insight1 implying two clusters.
- PCA: Although Type 2 can be separated from the rest, like in V1-Insight2 we can not find any clear border between Types 1

and 2 in PCA-S1 or PCA-S2. Therefore, we can relate both PCA-S1 and PCA-S2 to V1-Insight1.

- NMF: The types are mixed and are not identifiable in neither of the views.

As we can see, the majority of outputs from state-of-the-art methods have appeared in the output of the multi-insight approach. The additional advantage of our approach is that the ideal projection closer to the ground truth has a better chance to appear among the outputs from our method, which might not be the case for other approaches. For instance, in the example above MCOA-V1 provides a close to ideal projection, but we saw in the previous section that this was not case here and MCOA was not able to detect five clusters in the simulated dataset. We can conclude that our approach is more robust in identification of the ground truth and also provides a summary of major important insights.

5 Conclusion

We propose a new framework based on tensor decomposition for combining various DR techniques for multi-insight data visualization of multi-omics data. Via experimental evaluation with synthetic, simulated and real datasets we demonstrate that the new method is a stronger data-driven approach and a more effective way to gain insight into the data.

One aspect that we have not yet mentioned is that our method also provides a measure of the credibility of the solution. For example, if out of 100 DR methods, 95 of them suggest that the data has two clusters we can be more certain to make that conclusion compared with the situation where we have only one output. However, our method comes with a price. The multi-insight method needs to run many algorithms and therefore can be computationally expensive. Although the core method for combining the projections does not impose more computational complexity than any of the individual algorithms, still running a high number of algorithms in large-scale remains a challenge. However, fortunately this can be addressed by parallelization which theoretically is considered easy in the proposed framework. Besides, as a result of the decreasing trend of computing costs, data-driven approaches like ours will become more and more feasible.

Last but not least, the multi-insight approach can be used for making new hypotheses and more powerful exploratory data analysis, again based on the measure of credibility. If 60 out of 100 methods agree that the data has two clusters and 35 methods suggest that the data has three clusters, a secondary hypothesis of a triple cluster structure should be formed for further investigations.

Note that it is also possible to obtain some sort of multi-insight projections by plotting extra components obtained from a single-insight approach like PCA (e.g. third vs. fourth component). However, this is different from and less powerful than our proposed approach, because projections obtained by lower order components do not reflect the major patterns of the data.

Another issue that should be pointed out is that in this work we consider only datasets with two views (two sources of data). This choice is made due to technical limitations of some DR methods in the ensemble (especially coupled factorization methods) that accept only two matrices. The method as such does not have any limitations on the number of views and any arbitrary number of datasources can be processed with it.

In conclusion, to explore different projections may lead to an increased insight about the high-dimensional features of the data, and hence, increased knowledge.

Funding

This work is funded by Norwegian Research Council through the project 'National training initiative to make better use of biobanks and health registry data' [Project Number 248804].

Conflict of Interest: none declared.

References

Berry,M.W. *et al*. (2007) Algorithms and applications for approximate nonnegative matrix factorization. *Comput. Stat. Data Anal.*, **52**, 155–173.

Carroll,J.D. and Chang,J.-J. (1970) Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition. *Psychometrika*, **35**, 283–319.

Chalise,P. *et al*. (2016) InterSIM: simulation tool for multiple integrative omic datasets. *Comput. Methods Prog. Biomed.*, **128**, 69–74.

Chen,L. and Buja,A. (2009) Local multidimensional scaling for nonlinear dimension reduction, graph drawing, and proximity analysis. *J. Am. Stat. Assoc.*, **104**, 209–219.

Chessel,D. and Hanafi,M. (1996) Analyses de la co-inertie de *k* nuages de points. *Rev. Stat. Appl.*, **44**, 35–60.

Fanaee-T,H. and Gama,J. (2016) Tensor-based anomaly detection: an interdisciplinary survey. *Knowledge-Based Syst.*, **98**, 130–147.

Klami,A. *et al*. (2013) Bayesian canonical correlation analysis. *J. Mach. Learn. Res.*, **14**, 965–1003.

Kriegel,H.-P. *et al*. (2008) Angle-based outlier detection in high-dimensional data. In: *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 444–452. ACM, New York, NY, USA.

Lee,J.A. and Verleysen,M. (2010) Scale-independent quality criteria for dimensionality reduction. *Pattern Recogn. Lett.*, **31**, 2248–2257.

Li,Y. *et al*. (2018) A review on machine learning principles for multi-view biological data integration. *Brief. Bioinformatics*, **19**, 325–340.

Lock,E.F. and Dunson,D.B. (2013) Bayesian consensus clustering. *Bioinformatics*, **29**, 2610–2616.

Meng,C. *et al*. (2016) Dimension reduction techniques for the integrative analysis of multi-omics data. *Brief. Bioinformatics*, **17**, 628–641.

Mokbel,B. *et al*. (2013) Visualizing the quality of dimensionality reduction. *Neurocomputing*, **112**, 109–123.

Papalexakis,E.E. *et al*. (2016) Tensors for data mining and data fusion: models, applications, and scalable algorithms. *ACM Trans Intell. Syst. Technol.*, **8**, 1.

Silva,V.D. and Tenenbaum,J.B. (2003) Global versus local methods in nonlinear dimensionality reduction. In: *Advances in Neural Information Processing Systems*, Conference Paper, pp. 721–728, https://www.scopus.com/inward/record.uri?eid=2-s2.0-84899009769&partnerID=40&md5=9e8e0ca9e9e3cf0f21b2e3fd326f2d0f.

Stražar,M. *et al*. (2016) Orthogonal matrix factorization enables integrative analysis of multiple RNA binding proteins. *Bioinformatics*, **32**, 1527–1535.

TCGA Network (2012) Comprehensive molecular portraits of human breast tumours. *Nature*, **490**, 61–70.

Zhang,S. *et al*. (2012) Discovery of multi-dimensional modules by integrative analysis of cancer genomic data. *Nucleic Acids Res.*, **40**, 9379–9391.