Supplementary material for Optimal transport-based machine learning to match specific patterns: application to the detection of molecular regulation patterns in omics data

Thi Thanh Yen Nguyen[1,†], Warith Harchaoui[1,2],
Lucile Mégret[3], Cloé Mendoza[3],
Olivier Bouaziz[1,*], Christian Neri[3,*,†], Antoine Chambaz[1,*]

[1] Université Paris Cité, CNRS, MAP5, F-75006 Paris, France
[2] DERAISON.ai, France
[3] Sorbonne Université, CNRS UMR 8256, Brain-C Lab, Paris, France
[*] These authors contributed equally to this work.
[†] Correspondence.

October 12, 2023

A A brief data analysis

So as to give a sense of the distribution of the data, we propose two kinds of visual summaries. The first one uses Lloyd's k-means algorithm (Lloyd, 1982) to build synthetic profiles representing the real profiles $x_1, \ldots, x_M$ on the one hand and $y_1, \ldots, y_N$ on the other hand. The second one uses kernel density estimators of the j-th component of $x_1, \ldots, x_M$ on the one hand and of $y_1, \ldots, y_N$ on the other hand, for each $1 \leq j \leq d$.

A.1 Using k-means to cluster the mRNA and miRNA profiles

In Figure 1 we plot the synthetic mRNA profiles $\hat{x}_1, \ldots, \hat{x}_5$ of the 5 centroids obtained by running Lloyd's k-means algorithm on $x_1, \ldots, x_M$ with $k = 5$. Likewise, we plot in Figure 2 the synthetic miRNA profiles $\hat{y}_1, \ldots, \hat{y}_5$ of the 5 centroids obtained by running Lloyd's k-means algorithm on $y_1, \ldots, y_N$ with $k = 5$.

The 5 mRNA centroids correspond to 5319 ($\hat{x}_1$), 2097 ($\hat{x}_2$), 4688 ($\hat{x}_3$), 310 ($\hat{x}_4$) and 1202 ($\hat{x}_5$) mRNA profiles. The first and third centroids ($\hat{x}_1$ and $\hat{x}_3$), which represent 73% of the real mRNA profiles, are rather flat. The second and fourth centroids ($\hat{x}_2$ and $\hat{x}_4$), which represent 18% of the real mRNA profiles, are decreasing in poly Q length and age, in a more pronounced way for the latter than for the former. Finally, the fifth centroid ($\hat{x}_5$), which represents the remaining 9% of real mRNA profiles, is increasing in poly Q length and age.

The 5 miRNA centroids correspond to 872 ($\hat{y}_1$), 7 ($\hat{y}_2$), 80 ($\hat{y}_3$), 81 ($\hat{y}_4$) and 103 ($\hat{y}_5$) miRNA profiles. The first centroid ($\hat{y}_1$), which represents 76% of the real miRNA profiles, is rather flat.

The second and fifth centroids ($\hat{y}_2$ and $\hat{y}_5$), which represent 10% of the real miRNA profiles, are increasing in poly Q length and age, in a more pronounced way for the former than for the latter. The fourth centroid ($\hat{y}_4$), which represents 7% of the real miRNA profiles, is decreasing in poly Q length and age. Finally, the third centroid ($\hat{y}_3$), which represents 7% of the real miRNA profiles, exhibits two peaks.

In Section 1, we stated the following biological hypothesis: if a miRNA induces the degradation of a target mRNA or blocks its translation into proteins, or both, then the profile of the former should be similar to minus the profile of the latter (a particular form of affine relationship). In view of this hypothesis, it is tempting to relate the synthetic miRNA profiles $\hat{y}_2$ and $\hat{y}_5$ to the synthetic mRNA profiles $\hat{x}_4$ and $\hat{x}_2$, respectively, and the synthetic miRNA profile $\hat{y}_4$ to the synthetic mRNA profile $\hat{x}_5$. Our objective is to identify groups of real mRNA and miRNA profiles that interact in this manner.

A.2 Using kernel density estimators to study the marginal distributions of the mRNA and miRNA profiles

For each $1 \leq j \leq d$, we build the kernel density estimator of the j-th component of the mRNA profiles $x_1, \ldots, x_M$, using a Gaussian kernel and the default fine-tuning of the `density` function from the `stats` `R`-package (R Core Team, 2022), see Figure 3. We do the same for the miRNA profiles $y_1, \ldots, y_N$, see Figure 4. Both for mRNA and miRNA the kernel density estimates are systematically more concentrated around their means (all close to 0) than the corresponding Gaussian densities. Moreover, the kernel density estimates obtained from the M mRNA profiles are much smoother than those obtained from N miRNA profiles, a feature that could be simply explained by the fact that $M/N > 11$.

Table 1 reports, for each level of poly Q length (Q80, Q92, Q111, Q140, Q175) and age (2, 6, 10 months), the empirical standard deviation of mRNA (a) and miRNA (b) gene expressions, all normalized by the empirical standard deviation at poly Q length Q80 and 2 months of age (that is, by 0.0475 for mRNA and 0.0660 for miRNA). A clear pattern emerges from sub-Table 1 (a): except for poly Q length Q80, the poly Q length-specific empirical standard deviation increases as age increases. Likewise, except for age 2 months, the age-specific empirical standard deviation increases as poly Q length increases. On the contrary, no clear pattern emerges from sub-Table 1 (b) but the fact that, except for poly Q lengths Q80 and Q92, the poly Q length-specific empirical standard deviation increases as age increases. We do not comment on the empirical means because they are all very small compared to the corresponding empirical standard deviations.

B Simulation study

B.1 Two "Gromov-Wasserstein co-clustering" algorithms

We compare our algorithms with two co-clustering algorithms adapted from (Laclau et al., 2017). For self-containedness, we summarize here how these algorithms work.

The first step of both algorithms consists in computing the similarity matrices $K_X \in$

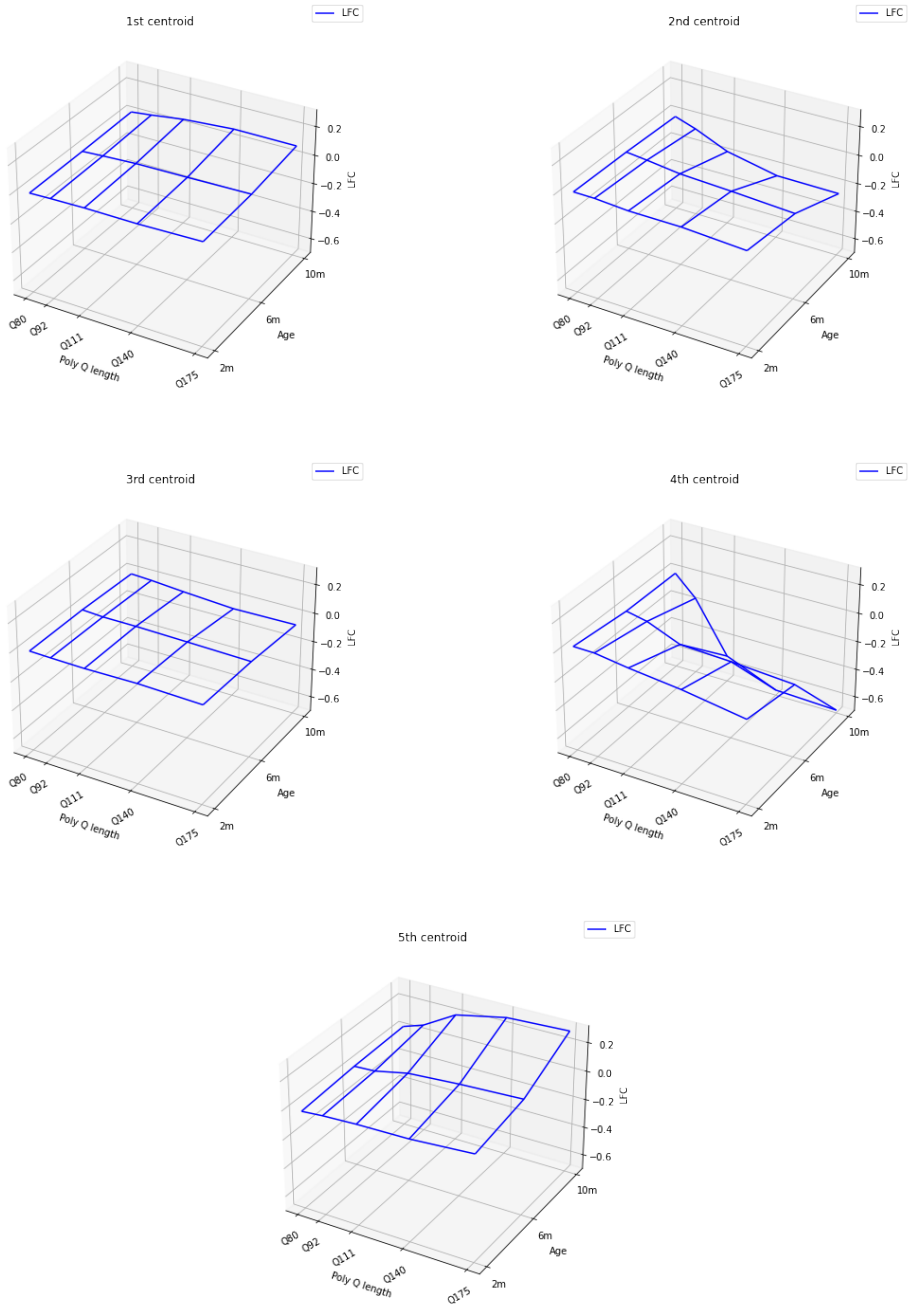

Figure 1: Profiles $\hat{x}_1, \ldots, \hat{x}_5$ of the 5 centroids obtained by Lloyd's k-means algorithm on the mRNA profiles $x_1, \ldots, x_M$.

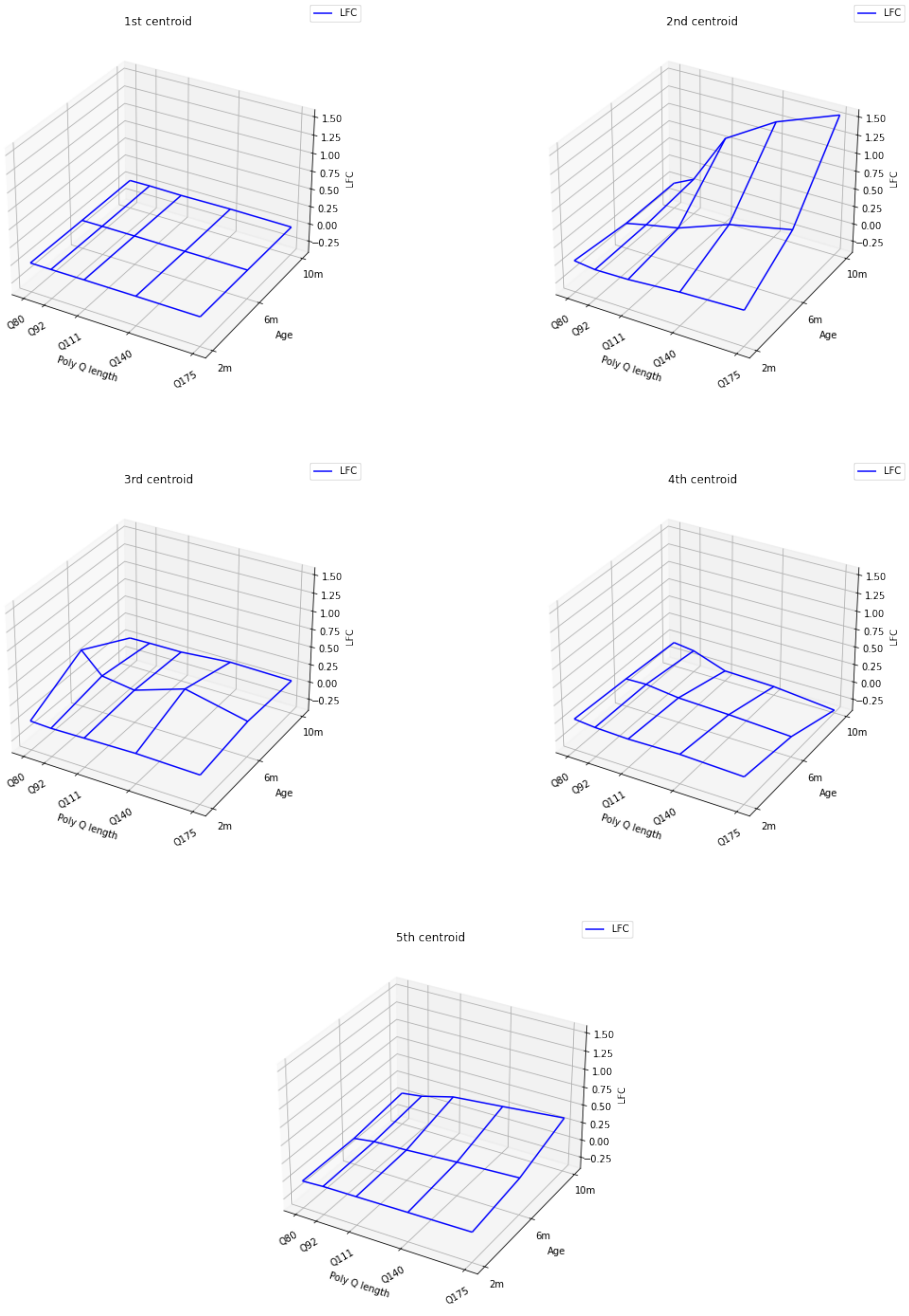

Figure 2: Profiles $\hat{y}_1, \ldots, \hat{y}_5$ of the 5 centroids obtained by running Lloyd's k-means algorithm on the miRNA profiles $y_1, \ldots, y_N$.

poly Q length	Age 2	Age 6	Age 10
Q80	1	0.646	1.39
Q92	0.886	1.02	1.48
Q111	0.964	1.21	3.08
Q140	0.805	1.70	4.11
Q175	1.24	1.86	4.32

(a) mRNA

poly Q length	Age 2	Age 6	Age 10
Q80	1	2.35	1.03
Q92	0.516	1.06	0.956
Q111	0.655	0.722	2.15
Q140	0.698	1.92	2.72
Q175	0.588	1.80	3.34

(b) miRNA

Table 1: For each level of poly Q length (Q80, Q92, Q111, Q140, Q175) and age (2, 6, 10 months) we computed the empirical standard deviation of mRNA (a) and miRNA (b) gene expressions, all normalized by the empirical standard deviation at poly Q length Q80 and 2 months of age (that is, by 0.0475 for mRNA and 0.0660 for miRNA).

$(\mathbb{R}_+)^{M\times M}$ and $K_Y \in (\mathbb{R}_+)^{N\times N}$ given by

$$(K_X)_{mm'} := \exp\left\{-\frac{\|x_m - x_{m'}\|_2^2}{2\ell_X^2}\right\} \quad (m, m' \in [\![M]\!]),$$
$$(K_Y)_{nn'} := \exp\left\{-\frac{\|y_n - y_{n'}\|_2^2}{2\ell_Y^2}\right\} \quad (n, n' \in [\![N]\!])$$

where ℓ_X (respectively, ℓ_Y) is the mean of all pairwise Euclidean distances between elements of X (respectively, of Y). The similarity matrices K_X and K_Y now represent X and Y through the lens of the so called radial basis function kernel.

For any integers $a, b \geq 1$ and pair of matrices $A \in \mathbb{R}^{a\times a}$ and $B \in \mathbb{R}^{b\times b}$, define

$$\begin{aligned}\Pi_{a,b} &:= \left\{P \in (\mathbb{R}_+)^{a\times b} | P\mathbf{1}_b = a^{-1}\mathbf{1}_a, P^\top \mathbf{1}_a = b^1 \mathbf{1}_b\right\},\\ \langle [A,B],[P,P]\rangle_F &:= \sum_{i,k\in[\![a]\!],j,\ell\in[\![b]\!]} (A_{ik} - B_{j\ell})^2 P_{ij}P_{k\ell} \quad (P \in \Pi_{a,b}),\\ \mathcal{GW}_\gamma(A,B) &:= \min_{P\in\Pi_{a,b}} \{\langle [A,B],[P,P]\rangle_F - \gamma E(P) \end{aligned} \tag{1}$$

where $E(P) := -\sum_{(i,j)\in[\![a]\!]\times[\![b]\!]} P_{ij}(\log P_{ij} - 1)$. The quantity $\mathcal{GW}_\gamma(A,B)$ is known in the literature as an entropic Gromov-Wasserstein discrepancy between A and B. It can be used to define an entropic Gromov-Wasserstein barycenter of A and B and its barycenter transport matrices. Specifically, setting $s = \lfloor \frac{1}{2}(a+b) \rfloor$ (one choice among many), $(\hat{\Gamma}, \hat{P}_A, \hat{P}_B) \in (\mathbb{R}_+)^{s\times s} \times \Pi_{s,a} \times \Pi_{s,b}$ that solves

$$\min_{\Gamma, P_A, P_B} \frac{1}{2}\left\{\Big(\langle [\Gamma, A],[P_A,P_A]\rangle_F - \gamma E(P_A)\Big) + \Big(\langle [\Gamma,B],[P_B,P_B]\rangle_F - \gamma E(P_B)\Big)\right\} \tag{2}$$

(where (Γ, P_A, P_B) ranges over $(\mathbb{R}_+)^{s\times s} \times \Pi_{s,a} \times \Pi_{s,b}$) can be interpreted as a barycenter between A and B ($\hat{\Gamma}$) and the optimal transport matrices between $\hat{\Gamma}$ and A ($\hat{P}_A$) and between $\hat{\Gamma}$ and B ($\hat{P}_B$).

The second step of the algorithms consists either in solving numerically (1) with $(A, B) =$

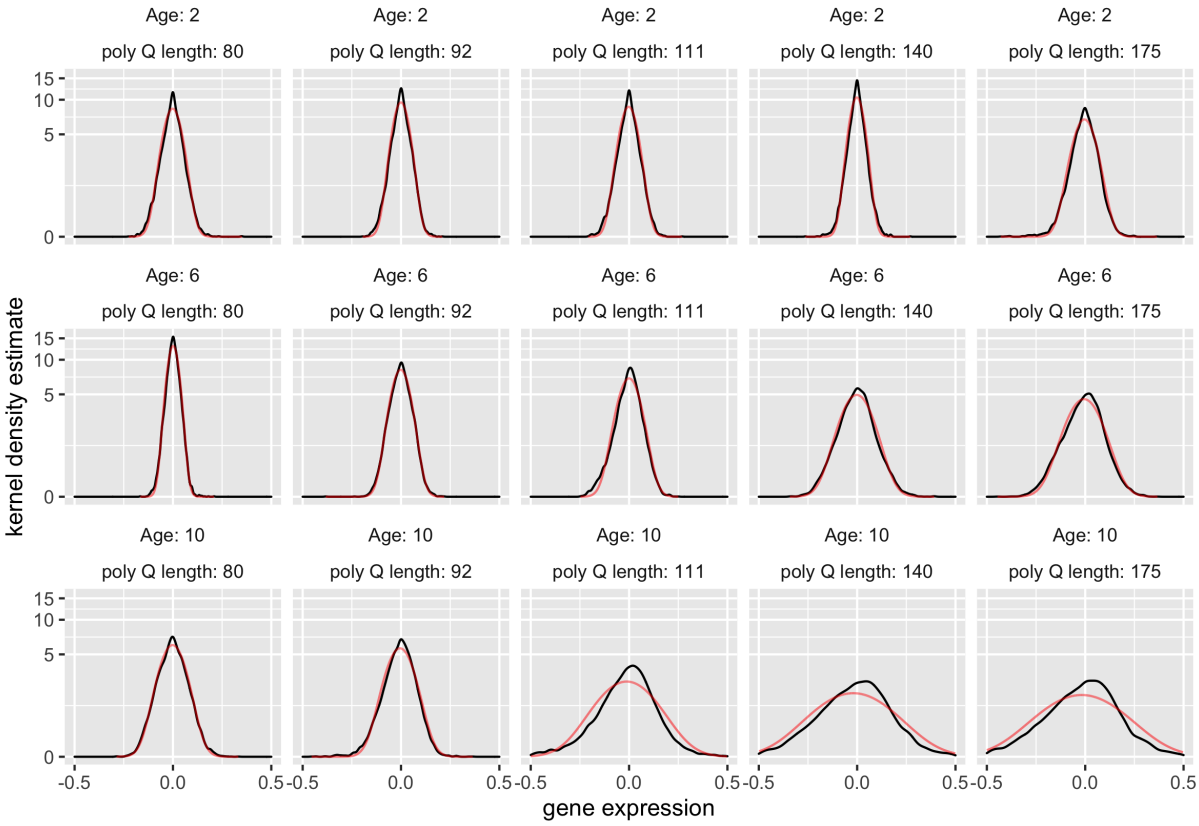

Figure 3: In black, kernel density estimates of the densities of mRNA gene expression for each level of poly Q length (Q80, Q92, Q111, Q140, Q175) and age (2, 6, 10 months), zooming on the interval $[-0.5, 0.5]$ and using a $\log(1 + \cdot)$-scale on the y-axis. In red, densities of the Gaussian laws with a mean and a variance equal to the empirical mean and variance computed in each stratum of data. Systematically, the kernel density estimates are more concentrated around their means than the corresponding Gaussian densities.

(K_X, K_Y), yielding $\tilde{Q}$, or in solving numerically (2) with $(A, B) = (K_X, K_Y)$, yielding in particular the transport matrices $\tilde{Q}_X$ and $\tilde{Q}_Y$. We call CCOT-GWD and CCOT-GWB the corresponding algorithms. In both cases, the Sinkhorn-Knopp algorithm is used and provides solutions that decompose as

$$\begin{aligned}\tilde{Q} &= \mathrm{diag}(\rho)\xi\,\mathrm{diag}(\rho'),\\ \tilde{Q}_X &= \mathrm{diag}(\rho_X)\xi_X\,\mathrm{diag}(\rho'_X),\\ \tilde{Q}_Y &= \mathrm{diag}(\rho_Y)\xi_Y\,\mathrm{diag}(\rho'_Y),\end{aligned}$$

for some $\rho, \rho_X \in \mathbb{R}^M$, $\rho', \rho'_Y \in \mathbb{R}^N$, $\rho_X, \rho_Y \in \mathbb{R}^s$ and $\xi \in \mathbb{R}^{M\times N}, \xi_X \in \mathbb{R}^{s\times M}, \xi_Y \in \mathbb{R}^{s\times N}$ (Peyré et al., 2016).

The third and last step builds upon either (ρ, ρ') or (ρ'_X, ρ'_Y) to derive partitions of X and

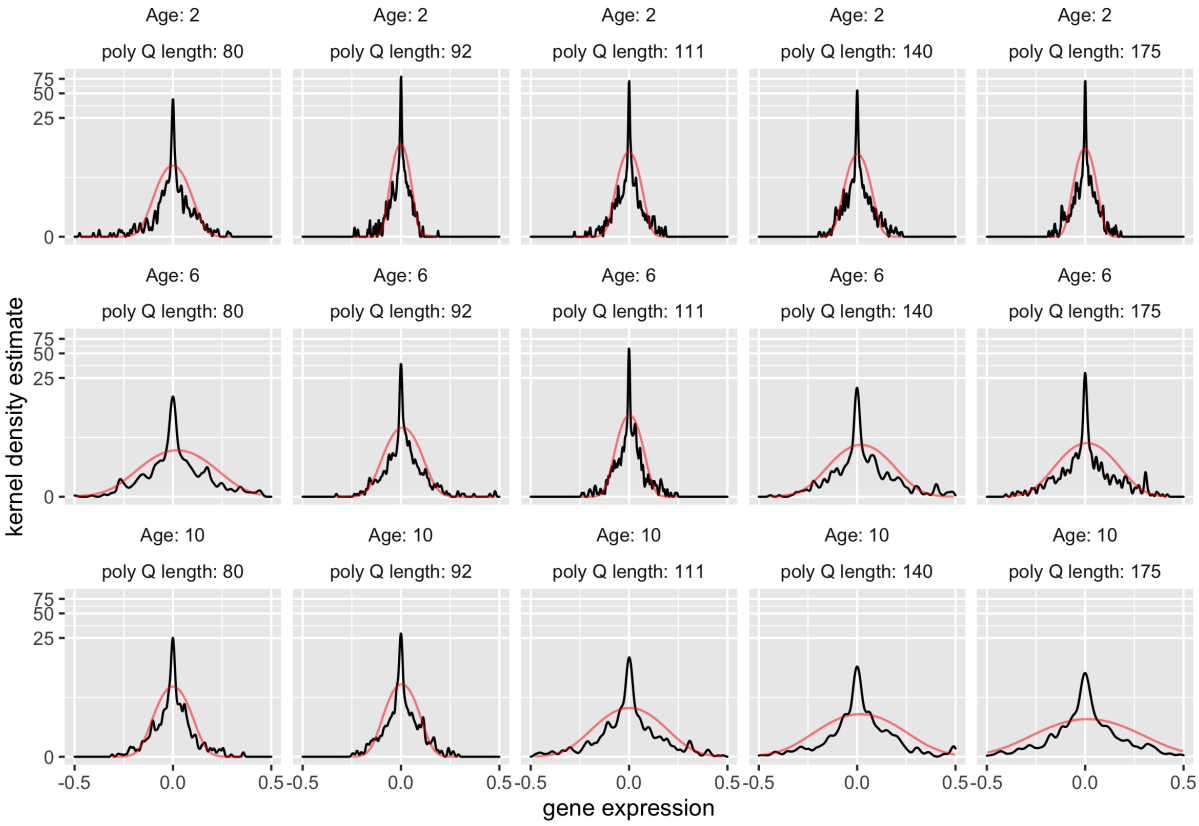

Figure 4: In black, kernel density estimates of the densities of miRNA gene expression for each level of poly Q length (Q80, Q92, Q111, Q140, Q175) and age (2, 6, 10 months), zooming on the interval $[-0.5, 0.5]$ and using a $\log(1 + \cdot)$-scale on the y-axis. In red, densities of the Gaussian laws with a mean and a variance equal to the empirical mean and variance computed in each stratum of data. Systematically, the kernel density estimates are more concentrated around their means than the corresponding Gaussian densities.

Y, by detecting "jumps" along the vectors. The two partitions finally yield a co-clustering.

B.2 First simulation study

Simulation scheme. For four different choices of the hyperparameters $M \geq 200, N \geq 200, K \geq 2, d \geq 2$, $\mu_1, \ldots, \mu_K \in \mathbb{R}^d$, $\sigma \in \mathbb{R}_+^*$, $\alpha \in (\mathbb{R}_+)^K$ such that $\sum_{k \in [\![K]\!]} \alpha_k = 1$, we sample independently $x_1, \ldots, x_M$ from the mixture of Gaussian laws

$$\sum_{k \in [\![K]\!]} \alpha_k N(\mu_k, \sigma^2 \mathrm{Id}_d) \tag{3}$$

and $y_1, \ldots, y_N$ from

$$\sum_{k \in [\![K]\!]} \alpha_k N(-\mu_k, \sigma^2 \mathrm{Id}_d). \tag{4}$$

One way to sample x from the mixture (3) consists in sampling a latent label u in $[\![K]\!]$ from the multinomial law with parameter $(1; \alpha_1, \ldots, \alpha_K)$ then in sampling x from the Gaussian law $N(\mu_u, \sigma^2 \mathrm{Id}_d)$. Similarly, sampling y from the mixture (4) can be carried out by sampling a latent label v in $[\![K]\!]$ from the multinomial law with parameter $(1; \alpha_1, \ldots, \alpha_K)$ then by sampling y from the Gaussian law $N(-\mu_v, \sigma^2 \mathrm{Id}_d)$. We think of x and y as having a mirrored relationship if $u = v$. In this light, the challenge that we tackle consists in finding such relationships without having access to the latent labels.

Table 2 describes the four configurations that we investigate. Note that configuration A2 is more difficult to deal with than A1 because *(i)* the weights in α are balanced in the latter and unbalanced in the former, and *(ii)* because the variance σ^2 is smaller in A1 than in A2. Moreover, configurations A3 and A4 are more challenging than A2 because there is $K = 4$ components in the Gaussian mixture under A3 and A4 and $K = 3$ components under A2.

configuration	(M, N)	K	$\mu_1, \ldots, \mu_K$	σ^2	α
A1	$(200, 200)$	3	$\begin{pmatrix}4.0\\0.5\\1.5\end{pmatrix}, \begin{pmatrix}1.8\\4.5\\1.1\end{pmatrix}, \begin{pmatrix}1.5\\1.5\\5.5\end{pmatrix}$	0.10	$(1/3, 1/3, 1/3)$
A2	$(300, 300)$	3	$\begin{pmatrix}4.0\\0.5\\1.5\end{pmatrix}, \begin{pmatrix}1.8\\4.5\\5.1\end{pmatrix}, \begin{pmatrix}3.5\\1.5\\5.5\end{pmatrix}$	0.15	$(0.2, 0.3, 0.5)$
A3	$(400, 300)$	4	$\begin{pmatrix}4.0\\0.5\end{pmatrix}, \begin{pmatrix}0.5\\3.5\end{pmatrix}, \begin{pmatrix}7.5\\7.8\end{pmatrix}, \begin{pmatrix}0.5\\0.5\end{pmatrix}$	0.20	$(0.4, 0.2, 0.2, 0.2)$
A4	$(300, 300)$	4	$\begin{pmatrix}4.0\\0.5\end{pmatrix}, \begin{pmatrix}0.5\\3.5\end{pmatrix}, \begin{pmatrix}7.5\\7.8\end{pmatrix}, \begin{pmatrix}0.5\\0.5\end{pmatrix}$	0.10	$(0.5, 0.2, 0.1, 0.2)$

Table 2: Four different configurations for the first simulation scheme. Configuration A1 is less challenging than A2 which is itself less challenging than A3 and A4.

Results. Thirty times, independently, we simulated synthetic data sets X and Y under the simulation scheme described above, then we applied the various algorithms as presented in Section 5.2. We summarize the results in Tables 4, 5, and 6. Table 4 summarizes the results of the seven algorithms listed in Section 5.2 that rely on *bona fide* co-clustering algorithms (see Section 4.2.1), that is, of our algorithms WTOT-SCC1*, WTOT-SCC1, WTOT-SCC2*, WTOT-SCC2, WTOT-BC* and of algorithms CCOT-GWD and CCOT-GWB. As for Tables 5 and 6, they summarize the results of our algorithm that relies on matching (see Section 4.2.2).

Table 4. Except in configuration A1, where they perform equally well, our algorithms WTOT-SCC1, WTOT-SCC2 outperform their competitors CCOT-GWD and CCOT-GWB.

Recall that WTOT-SCC1 and WTOT-SCC2 learn the number of co-clusters. When they underestimate it, they pay a high price, partly explaining why the standard deviations

are rather large. In order to assess how well they work relative to their counterparts which benefit from knowing in advance the true number of co-clusters, we can compare their measures of performance to those of algorithms WTOT-SCC1* and WTOT-SCC2*. In configurations A1 and A2, algorithms WTOT-SCC1, WTOT-SCC2 perform almost as well as WTOT-SCC1* and WTOT-SCC2*, respectively. In configuration A3, they are clearly outperformed. In configuration A4, algorithm WTOT-SCC1 performs better in average but not in standard deviation.

Finally, we note that algorithm WTOT-BC* outperforms all our other algorithms. Unfortunately, its counterpart that learns the number of co-clusters performs poorly (results not shown).

Tables 5 and 6. Table 5 illustrates the influence of $k = k'$ on the performances of algorithm WTOT-matching. In configuration A1, specificity is not impacted much by the value of $k = k'$, whereas precision decreases and sensitivity increases as $k = k'$ grows. More specifically, precision does not change much when one goes from $k = k' = 10$ to $k = k' = 75$ but it drops for larger values of $k = k'$. As for sensitivity, it increases dramatically when one goes from $k = k' = 10$ to $k = k' = 75$ and slightly for higher values of $k = k'$. Furthermore we note that, in configuration A1, when $k = k'$ equal either 65 or 75 and are thus closest to $N\alpha_\ell = M\alpha_\ell \approx 67$, $\tilde{k}_r$ is close to 67 and precision, sensitivity and specificity are quite satisfying. In configuration A4 (as in configuration A1), specificity is not impacted much by the value of $k = k'$; on the contrary, precision decreases and sensitivity increases steadily as $k = k'$ grows. The best performances are achieved for $k = k' = 95$ and $k = k' = 150$, that is, when $k = k'$ get closer to $M \max_{i \leq 4}\{\alpha_i\} = N \max_{i \leq 4}\{\alpha_i\}$. As emphasized earlier, deriving relevant matchings is more difficult in configuration A4 than in configuration A1 because the weights given in parameter α are unbalanced in the former and balanced in the latter.

Table 6 summarizes the results of WTOT-matching in all configurations for a specific choice of $k = k'$ in terms of the row- and column-specific averages $\tilde{k}_r$ and $\tilde{k}_c$, precision, sensitivity and specificity. In each configuration, we chose the value of $k = k'$ among many retrospectively, so that the overall performance (in terms of precision, sensitivity and specificity) is good. The left-hand-side (m-specific) and right-hand-side (n-specific) tables in Table 6 are very similar. This does not come as a surprise because the first simulation scheme imposes symmetry.

B.3 Second simulation study

Simulation scheme. The second simulation scheme also relies on mixtures of Gaussian laws, but the means and weights are generated randomly from a Gaussian determinantal point process (DPP) for the former and from a Dirichlet law for the latter. More specifically, given the hyperparameters $M \geq 200, N \geq 200, K \geq L \geq 3$, $\sigma \in \mathbb{R}_+^*$,

1. we sample $\mu_1, \ldots, \mu_K$ from a Gaussian DPP on $[0,1]^2$ with a kernel proportional to $x \mapsto \exp(-\|x/0.05\|_2^2)$ conditionally on obtaining exactly K points Lavancier et al. (2015); Baddeley and Turner (2005);

2. independently, we sample $\alpha \in (\mathbb{R}_+)^K$ and $\beta \in (\mathbb{R}_+)^L$ from the Dirichlet laws with parameters $7\,\mathbf{1}_K$ and $7\,\mathbf{1}_L$;

3. we sample independently $x_1, \ldots, x_M$ from the mixture of Gaussian laws

$$\sum_{k \in [\![K]\!]} \alpha_k N(\mu_k, \sigma^2 \mathrm{Id}_2)$$

and $y_1, \ldots, y_N$ from

$$\sum_{k \in [\![L]\!]} \beta_k N(-\mu_k, \sigma^2 \mathrm{Id}_2).$$

We use a DPP to generate $\mu_1, \ldots, \mu_K$ to avoid the arbitrary choice of the mean parameters in such a way that the randomly picked $\mu_1, \ldots, \mu_K$ are dispersed in $[0,1]^2$ (because the DPP is a repulsive point process).

Table 3 describes the four configurations that we investigate. The larger L is the more challenging the configuration is. In configurations B2, B3, B4, it holds that $K = L + 1$, hence the data points from the Kth cluster should not be matched. Moreover, for given (K, L) and (M, N), a configuration gets more challenging as its σ^2 parameter increases. It is noteworthy that the values of σ^2 as reported in Table 3 cannot be compared straightforwardly to those reported in Table 2, because $\mu_1, \ldots, \mu_K$ live in $[0,1]^2$ in the present simulation study whereas they do not in the simulation study of Section B.2.

configuration	(M, N)	(K, L)	σ^2
B1	$(200, 200)$	$(3, 3)$	5×10^{-4}
B2	$(300, 300)$	$(7, 6)$	10^{-4}
B3	$(300, 300)$	$(16, 15)$	10^{-5}
B4	$(300, 300)$	$(16, 15)$	10^{-4}

Table 3: Four different configurations for the second simulation scheme. The larger $\ell \in [\![4]\!]$ is the more challenging configuration Bℓ is.

Results. Thirty times, independently, we simulated synthetic data sets X and Y under the simulation scheme described above, then we applied the various algorithms as presented in Section 5.2. Table 7 summarizes the results of the seven algorithms listed in Section 5.2 that rely on *bona fide* co-clustering algorithms (see Section 4.2.1). Tables 8 and 9 summarize the results of our algorithm that relies on matching (see Section 4.2.2).

Table 7. We first note that WTOT-SCC1, WTOT-SCC2 and CCOT-GWD perform similarly in configurations B1 and B2, much better than CCOT-GWB, but less well than the oracular algorithms WTOT-SCC1*, WTOT-SCC2* and WTOT-BC*. More generally, across configurations B1, B2, B3, B4, the oracular algorithms WTOT-SCC1* and WTOT-SCC2* perform much better than the other algorithms (and WTOT-BC* fails to find a partition with the given number of co-clusters in B3 and B4). Moreover, WTOT-SCC1 and WTOT-SCC2 perform poorly in configurations B2, B3 and B4 though not as poorly as CCOT-

GWD and CCOT-GWB in configurations B3 and B4. It seems that WTOT-SCC1 and WTOT-SCC2 fail to learn a "practical" number of co-clusters from $\tilde{P}$, in part because of those among $x_1, \ldots, x_M$ that are drawn from the Gaussian law $N(\mu_K, \sigma^2 \mathrm{Id}_2)$ when $K = L + 1$ (these data points should not be matched at all). The fact that WTOT-SCC1 and WTOT-SCC2 perform similarly in configurations B3 and B4 although σ^2 is 10 times larger in B4 than in B3 gives credit to the previous interpretation.

Tables 8 and 9. Table 8 illustrates the influence of $k = k'$ on the performances of algorithm WTOT-matching in configurations B1 and B4. In each configuration, the values of $k = k'$ are chosen in the vicinity of M/K (67 in configuration B1, 11 in configuration B4). We observe the same patterns in configurations B1 and B4: precision decreases (gradually) and specificity decreases (slightly) as $k = k'$ grows, while sensitivity increases (strongly in B1 and dramatically in B4).

Table 9 summarizes the results of WTOT-matching in configurations B1, B2, B3, B4 for a specific choice of $k = k'$ in terms of the row- and column-specific averages $\tilde{k}_r$ and $\tilde{k}_c$, precision, sensitivity and specificity. In each configuration, we chose the value of $k = k'$ among many retrospectively so that the overall performance (in terms of precision, sensitivity and specificity) is good. The left-hand-side (m-specific) and right-hand-side (n-specific) tables in Table 9 are very similar although $K > L$ in configuration B3 and B4. Interestingly, the fact that σ^2 is 10 times larger in configuration B4 than in B3 does not affect much the performance of the matching algorithm.

C Biological analysis of the results

In this section, we expose the full-blown biological analysis of the results.

C.1 Analyzing the overlaps

Figure 5 presents two Venn diagrams summarizing the overlaps between the sets of miRNAs (respectively, mRNAs) which belong to a pair output by the WGCNA, MiRAMINT and WTOT-matching algorithms. On the one hand, focusing on miRNAs, 13/14 (respectively, 29/46) miRNAs involved in a mRNA-miRNA pair output by MiRAMINT (respectively, WGCNA) are among the miRNAs involved in a mRNA-miRNA pair output by WTOT-matching. On the other hand, focusing on mRNAs, 1/20 (respectively, $100/1,583$) miRNAs involved in a mRNA-miRNA pair output by MiRAMINT (respectively, WGCNA) are among the miRNAs involved in a mRNA-miRNA pair output by WTOT-matching. We carry out one-sided Fisher's exact tests to quantify to what extent the overlaps reflect an agreement between two algorithms (using the 1,478 miRNAs and 27,355 mRNAs appearing in the TargetScan, MicroCosm and miRDB databases as reference populations). The p-value of the test comparing WTOT-matching and MiRAMINT in mRNAs equals 0.45. The other p-values are smaller than 10^{-6}.

It is desirable to identify miRNAs that are particularly susceptible to play a distinct role in HD in mice. To do so, we evaluate two simple criteria on the mRNAs associated to each miRNA (the miRNAs with no matched mRNAs are obviously less interesting in our study). The criteria assess to what extent a mRNA profile is "monotonic" and, on the contrary, to what extent it

is "peaked", accounting for the amplitude of log-fold change. Formally, rewriting each profile $x \in \mathbb{R}^{15}$ as a matrix $(\tilde{x}_{tq})_{t\in[\![3]\!],q\in[\![5]\!]}$, the first criterion is the minimum (relative to time t) of the absolute values of the slopes of the regression lines of the sets $\{(q, \tilde{x}_{tq}) : q \in [\![5]\!]\}$ and the second criterion is $\max_{q\in[\![5]\!]}(\tilde{x}_{1q} - \tilde{x}_{2q}) \times (\tilde{x}_{2q} - \tilde{x}_{3q})$. By convention, a miRNA profile is labeled monotonic (respectively, peaked) if at least one of its associated mRNA profiles is such that its first (respectively, second) criterion is larger than 95% (respectively, smaller than 99%) of the similar criteria. Moreover, all mRNA profiles x appearing in a pair (x, y) are labeled like y. We stress that no mRNA labeling conflicts occur.

Below, we reproduce the same analysis as above focusing in turn on mRNA-miRNA matchings labeled as peaked, monotonic, and neither peaked nor monotonic.

Peaked profiles. Figure 6 presents two Venn diagrams summarizing the overlaps between the sets of miRNAs (respectively, mRNAs) which belong to a pair output by the WGCNA, MiRAMINT and WTOT-matching algorithms, *looking at the WTOT-matching matchings labeled as peaked*. None of the 17 miRNAs and none of the 12 mRNAs involved in a mRNA-miRNA pair output by WTOT-matching are involved in a mRNA-miRNA pair output by the WGCNA or MiRAMINT algorithms.

The take-home message is that the WTOT-algorithm retains mRNA-miRNA matchings that we label as peaked whereas neither the WGCNA nor the MiRAMINT algorithms do.

Monotonic profiles. Figure 7 presents two Venn diagrams summarizing the overlaps between the sets of miRNAs (respectively, mRNAs) which belong to a pair output by the WGCNA, MiRAMINT and WTOT-matching algorithms, *looking at the WTOT-matching matchings labeled as monotonic*. On the one hand, focusing on miRNAs, 8/14 (respectively, 9/46) miRNAs involved in a mRNA-miRNA pair output by MiRAMINT (respectively, WGCNA) are among the miRNAs involved in a mRNA-miRNA pair output by WTOT-matching. On the other hand, focusing on mRNAs, 0/20 (respectively, $14/1,583$) miRNAs involved in a mRNA-miRNA pair output by MiRAMINT (respectively, WGCNA) are among the miRNAs involved in a mRNA-miRNA pair output by WTOT-matching. We carry out one-sided Fisher's exact tests to quantify to what extent the overlaps reflect an agreement between two algorithms (using the 1,478 miRNAs and 27,355 mRNAs appearing in the TargetScan, MicroCosm and miRDB databases as reference populations), excluding the comparison of the MiRAMINT and WTOT-matching algorithms in mRNAs (due to an empty intersection). The p-values are smaller than 10^{-5}.

The take-home message is that, in matchings that we label as monotonic, the agreement between the WTOT-matching and WGCNA algorithms is better than that between the WTOT-matching and MiRAMINT algorithms.

Neither peaked nor monotonic profiles. Finally, Figure 8 presents two Venn diagrams summarizing the overlaps between the sets of miRNAs (respectively, mRNAs) which belong to a pair output by the WGCNA, MiRAMINT and WTOT-matching algorithms and labeled neither as peaked nor monotonic. On the one hand, focusing on miRNAs, 12/14 (respectively, 28/46) miRNAs involved in a mRNA-miRNA pair output by MiRAMINT (respectively, WGCNA) are among the miRNAs involved in a mRNA-miRNA pair output by WTOT-matching. On the other hand, focusing on mRNAs, 1/20 (respectively,

86/1,583) miRNAs involved in a mRNA-miRNA pair output by MiRAMINT (respectively, WGCNA) are among the miRNAs involved in a mRNA-miRNA pair output by WTOT-matching. We carry out one-sided Fisher's exact tests to quantify to what extent the overlaps reflect an agreement between two algorithms (using the 1,478 miRNAs and 27,355 mRNAs appearing in the TargetScan, MicroCosm and miRDB databases as reference populations), excluding the comparison of the MiRAMINT and WTOT-matching algorithms in mRNAs (due to an intersection reduced to a singleton). The p-value are smaller than 10^{-5}.

The take-home message is that, in matchings that we label as neither peaked nor monotonic, the agreement between the WTOT-matching and WGCNA algorithms is better than that between the WTOT-matching and MiRAMINT algorithms.

C.2 Enrichment analysis

We interpret here the results of the enrichment analysis exposed in Section 6.3.2 from a biological viewpoint. Recall that the peaked and monotonic profiles are especially interesting because they are more susceptible to correspond to mRNAs and miRNAs that play a distinct role in HD in mice. Extracellular matrix organization (the primary annotation of the matchings output by the WTOT-matching algorithm, driven by the mRNA-miRNA matchings labeled as neither peaked nor monotonic) is known to be regulated by miRNAs (Rutnam et al., 2013) and HD mutations are known to strongly affect neuronal identity via down-regulating a large number of cell identity genes (Achour et al., 2015). Mitigation of host antiviral defense response (the first secondary annotation of the matchings output by the WTOT-matching algorithm, due to the mRNA-miRNA matchings labeled monotonic) is similar to the primary annotation of the matchings output by the MiRAMINT algorithm. Finally, conventional motile cilium (the second secondary annotation of the matchings output by the WTOT-matching algorithm, due to the mRNA-miRNA matchings labeled peaked) is a new finding.

Additionally, although miRNA levels and regulation in response to mutant huntingtin is anticipated to be dependent on cellular context and could be differentially influenced across murine models of HD, it is noticeable that the analysis of miRNA regulation in the striatum of HD knock-in mice based on the WTOT-matching algorithm retained several miRNAs that are altered in the striatum of other types of HD mice such as BACHD (Olmo et al., 2021) or altered in the human HD caudate nucleus (Petry et al., 2022) such as for example Mir100, Mir127, Mir132, Mir 212 and Mir133, supporting the relevance of our findings for the study of molecular regulation in mouse and human HD.

We believe that these facts substantiate our claim that the WTOT-matching algorithm strikes a good balance between the low and high selectivity of the WGCNA and MiRAMINT algorithms. Moreover, our findings related to striatal alterations in HD mice lead to reconsidering the formerly-expressed view on a limited role of miRNA regulation in the striatum of HD mice on a systems level (Mégret et al., 2020).

D Parametric model Θ

Introduced in Section 4.1, the parametric model Θ consists of affine mappings $\theta : \mathbb{R}^d \to \mathbb{R}^d$ of the form $x \mapsto \theta_1 x + \theta_2$, where θ_1 takes its values in a subset T_1 of $\mathbb{R}^{d\times d}$ and θ_2 takes its values in $\mathbb{R}^d$ (without any constraint). It is easier to describe the set of linear mappings $\{x \mapsto \theta_1 x : \theta_1 \in T_1\}$ after a reparametrization.

In the rest of this section only, we rewrite the mRNA and miRNA profiles $x, y \in \mathbb{R}^d$ under the form of $d_1 \times d_2$ matrices $\tilde{x} = (\tilde{x}_{tq})_{t\in[\![d_1]\!],q\in[\![d_2]\!]}$ and $\tilde{y} = (\tilde{y}_{tq})_{t\in[\![d_1]\!],q\in[\![d_2]\!]}$. For each $t \in [\![d_1]\!]$ and $q \in [\![d_2]\!]$, $\tilde{x}_{t\bullet}$ and $\tilde{x}_{\bullet q}$ are the tth row and qth column of $\tilde{x}$. Here, indices t and q correspond to the age and CAG lengths of the mice whose RNA sequencing yielded $\tilde{x}_{tq}$ and $\tilde{y}_{tq}$.

The definition of T_1 should formalize what we consider to be a (plausible) mirroring relationship. The simplest mirroring relationship is $y = -x$ or, equivalently, $\tilde{y} = -\tilde{x}$. The equality is of course too stringent/rigid, and the definition of T_1 is driven by our wish to relax it.

Biological arguments encourage us to consider that y and x exhibit a (plausible) mirroring relationship if, for each (t, q) ($t \in [\![d_1]\!]$, $q \in [\![d_2]\!]$), $\tilde{y}_{tq}$ is strongly negatively correlated with $\tilde{x}_{tq}$, mainly, and (positively or negatively) correlated with $\tilde{x}_{(t-1)q}$ (if $t > 1$) and/or with $\tilde{x}_{t(q-1)}$ (if $q > 1$), secondarily. We thus formalize $\{x \mapsto \theta_1 x : \theta_1 \in T_1\}$ as the set of all linear mappings of the form

$$x \mapsto \tilde{\theta}_1^a \odot \tilde{x} + \tilde{\theta}_1^b \odot \begin{pmatrix} \mathbf{0}_{d_2}^\top \\ \tilde{x}_{1\bullet} \\ \vdots \\ \tilde{x}_{(d_1-1)\bullet} \end{pmatrix} + \tilde{\theta}_1^c \odot \left(\mathbf{0}_{d_1} \tilde{x}_{\bullet 1} \ \cdots \tilde{x}_{\bullet(d_2-1)}\right)$$

where $\tilde{\theta}_1^a$ and $\tilde{\theta}_1^b, \tilde{\theta}_1^c$ are $d_1 \times d_2$ matrices (here, $\odot$ is the componentwise multiplication). The entries of $\tilde{\theta}_1^a$ correspond to comparisons between $\tilde{x}_{tq}$ and $\tilde{y}_{tq}$ (same poly Q length q and age t). The entries of $\tilde{\theta}_1^b$ (whose first row consists of 0s) correspond to comparisons between $\tilde{x}_{(t-1)q}$ and $\tilde{y}_{tq}$ (same poly Q length q, different age t). The entries of $\tilde{\theta}_1^c$ (whose first column consists of 0s) correspond to comparisons between $\tilde{x}_{t(q-1)}$ and $\tilde{y}_{tq}$ (different poly Q length q, same age t).

In the simulation study presented in Section B, the entries of $\tilde{\theta}_1^a$ are constrained to take their values in the interval $]-5, 0[$ while those of $\tilde{\theta}_1^b, \tilde{\theta}_1^c$ are constrained to take their values in $]-1/2, 1/2[$. The initial mapping is drawn randomly by sampling the entries of $\tilde{\theta}_1^a$ independently and uniformly in $]-5, 0[$ and, independently, by sampling the entries of $\tilde{\theta}_1^b$ and $\tilde{\theta}_1^c$ independently and uniformly in $]-1/2, 1/2[$.

In the illustration of the WTOT-matching algorithm presented in Section 6.2, the mapping $\widehat{\theta}$ is parametrized by $\tilde{\theta}$ given by

$$\tilde{\theta}_1^a = \begin{pmatrix} -0.88 & -1.47 & -0.73 \\ -0.59 & -0.90 & -0.89 \\ -0.62 & -0.70 & -1.17 \\ -0.97 & -1.30 & -0.95 \\ -0.56 & -1.16 & -1.24 \end{pmatrix}, \quad \tilde{\theta}_1^b = \begin{pmatrix} 0 & 0 & 0 \\ 0.13 & -0.19 & 0.13 \\ 0.17 & 0.09 & 0.13 \\ 0.19 & 0.09 & -0.00 \\ 0.18 & 0.15 & 0.08 \end{pmatrix},$$

$$\tilde{\theta}_1^c = \begin{pmatrix} 0 & 0.18 & -0.18 \\ 0 & 0.19 & 0.17 \\ 0 & 0.04 & 0.15 \\ 0 & 0.05 & 0.11 \\ 0 & 0.18 & 0.14 \end{pmatrix}, \quad \theta_2 = \begin{pmatrix} -0.01 & 0.01 & -0.00 \\ 0.00 & 0.01 & 0.01 \\ 0.00 & 0.01 & 0.00 \\ 0.01 & 0.01 & 0.01 \\ -0.01 & 0.01 & 0.01 \end{pmatrix}$$

(the numbers are rounded to two decimal places). We note that:

- On the one hand, the entries of $\tilde{\theta}_1^a$ are distributed around -1. On the other hand, the

entries of θ_2 are small. This is in line with the *strong* biological hypothesis (that is, if a miRNA induces the degradation of a target mRNA or blocks its translation into proteins, or both, then the profile of the former should be similar to minus the profile of the latter).

- The entries of $\tilde{\theta}_1^b$ and $\tilde{\theta}_1^c$ are small.

Procedure 1 *Master optimal transport algorithm.*

Input: X, Y, minibatch sizes $\widetilde{M}, \widetilde{N}$, decay rate $\eta \in]0,1]$, initial regularization parameter γ_0, initial mapping $\theta_0 \in \Theta$, maximal number of iterations T

Output: Transport coupling $\tilde{P}_T \in (\mathbb{R}_+)^{M \times N}$, mapping $\theta_T \in \Theta$, weight ω_T

Compute:

- $\underline{\gamma} = \text{mean}\{\|x - x'\|_2 : x, x' \in X\}$ {for entropy regularization}
- $h = \text{mean}\{\|y - y'\|_2 : y, y' \in Y\}$ {for window calibration}

Set $t \leftarrow 0$
Set stop $\leftarrow$ FALSE
while $\neg$ stop or $t < T$ **do**
$\gamma_t \leftarrow \max(\gamma_0 \times \eta^t, \underline{\gamma})$
Sample uniformly a minibatch of $\widetilde{M}$ observations $\tilde{x}_{1:\widetilde{M}} := (\tilde{x}_1, \ldots, \tilde{x}_{\widetilde{M}})$ from X
Sample uniformly a minibatch of $\widetilde{N}$ observations $\tilde{y}_{1:\widetilde{N}} := (\tilde{y}_1, \ldots, \tilde{y}_{\widetilde{N}})$ from Y
Define and compute $\theta_t(\tilde{x}_{1:\widetilde{M}}) := \big(\theta_t(\tilde{x}_1), \ldots, \theta_t(\tilde{x}_{\widetilde{M}})\big)$
Define and compute $\omega_t \in (\mathbb{R}_+)^{\widetilde{M}}$ such that $\sum_{m \in [\![\widetilde{M}]\!]} (\omega_t)_m = 1$ by setting

$$(\omega_t)_m \propto \sum_{n \in [\![\widetilde{N}]\!]} \varphi\left(\frac{\tilde{y}_n - \theta_t(\tilde{x}_m)}{h}\right) \quad (\text{all } m \in [\![\widetilde{M}]\!])$$

where φ is the standard normal density
Define $\mu^{\omega_t}_{\theta_t(\tilde{x}_{1:\widetilde{M}})}$, the ω_t-weighted empirical measure attached to $\theta_t(\tilde{x}_{1:\widetilde{M}})$, and $\nu_{\tilde{y}_{1:\widetilde{N}}}$, the empirical measure attached to $\tilde{y}_{1:\widetilde{N}}$
Compute $\text{Loss}_t = \bar{\mathcal{W}}_{\gamma_t}\left(\mu^{\omega_t}_{\theta_t(\tilde{x}_{1:\widetilde{M}})}, \nu_{\tilde{y}_{1:\widetilde{N}}}\right)$ and ∇Loss_t, the gradient of Loss_t relative to the parameter defining θ_t {relies on the Sinkhorn-Knopp algorithm}
Update the parameter defining θ_t by performing one step of stochastic gradient descent, yielding θ_{t+1}
Check stopping criterion and update stop variable accordingly
$t \leftarrow t + 1$
end while
Set $\theta_T \leftarrow \theta_{t-1}$
Set $\gamma_T \leftarrow \gamma_{t-1}$
Define and compute $\omega_T \in (\mathbb{R}_+)^M$ such that $\sum_{m \in [\![M]\!]} (\omega_T)_m = 1$ by setting

$$(\omega_T)_m \propto \sum_{n \in [\![N]\!]} \varphi\left(\frac{y_n - \theta_T(x_m)}{h}\right) \quad (\text{all } m \in [\![M]\!])$$

Compute $\tilde{P}_T \in \Pi(\omega_T)$ solving $\min_{P \in \Pi(\omega_T)} \mathcal{W}_{\gamma_T}\left(\mu^{\omega_T}_{\theta_T(X)}, \nu_Y\right)$

	the WTOT(...) algorithms					the CCOT(...) algorithms	
	WTOT-SCC1*	WTOT-SCC1	WTOT-SCC2*	WTOT-SCC2	WTOT-BC*	CCOT-GWD	CCOT-GWB
A1	0	0.068 ± 0.126	0	0.068 ± 0.126	0	0.054 ± 0.14	0.092 ± 0.15
A2	0 ± 0.001	0.014 ± 0.029	0 ± 0.001	0.016 ± 0.035	0.033 ± 0.125	0.105 ± 0.13	0.121 ± 0.146
A3	0.005 ± 0.005	0.189 ± 0.175	0.0182 ± 0.033	0.233 ± 0.179	0.029 ± 0.087	0.612 ± 0.03	0.532 ± 0.068
A4	0.326 ± 0.064	0.282 ± 0.232	0.257 ± 0.256	0.393 ± 0.164	0.05 ± 0.093	0.507 ± 0.123	0.522 ± 0.116

Table 4: Mean ($\pm$ standard deviation) computed across the 30 independent replications of the co-clustering discrepancy obtained for configurations A1, A2, A3, A4.

	$k = k'$	$\tilde{k}_r$	precision	sensitivity	specificity
A1	10	7.825 ± 0.091	1.0 ± 0.0	0.118 ± 0.001	1.0 ± 0.0
A1	35	29.373 ± 0.261	1.0 ± 0.0	0.442 ± 0.003	1.0 ± 0.0
A1	65	60.649 ± 0.998	0.999 ± 0.002	0.913 ± 0.014	1.0 ± 0.0
A1	75	67.418 ± 0.9	0.981 ± 0.006	0.991 ± 0.013	0.994 ± 0.002
A1	95	76.335 ± 1.282	0.888 ± 0.014	1.0 ± 0.0	0.957 ± 0.005
A1	150	97.049 ± 1.182	0.727 ± 0.012	1.0 ± 0.0	0.879 ± 0.005

	$k = k'$	$\tilde{k}_r$	precision	sensitivity	specificity
A4	10	6.964 ± 0.161	0.998 ± 0.003	0.089 ± 0.003	1.0 ± 0.0
A4	35	28.632 ± 0.668	0.995 ± 0.009	0.374 ± 0.01	1.0 ± 0.0
A4	65	54.653 ± 0.927	0.986 ± 0.011	0.668 ± 0.018	0.998 ± 0.002
A4	75	61.193 ± 0.724	0.963 ± 0.016	0.709 ± 0.022	0.993 ± 0.003
A4	95	75.856 ± 0.749	0.893 ± 0.017	0.768 ± 0.022	0.975 ± 0.003
A4	150	121.273 ± 3.63	0.783 ± 0.025	0.976 ± 0.023	0.936 ± 0.011

Table 5: Mean ($\pm$ standard deviation) computed across the 30 independent replications of $\tilde{k}_r$, precision, sensitivity and specificity of the m-specific matchings averaged across all mRNAs for configuration A1 (left) and A4 (right).

	$k = k'$	$\tilde{k}_r$	precision	sensitivity	specificity
A1	75	67.418 ± 0.9	0.981 ± 0.006	0.991 ± 0.013	0.994 ± 0.002
A2	130	100.217 ± 2.127	0.976 ± 0.017	0.894 ± 0.027	0.995 ± 0.004
A3	120	82.764 ± 1.105	0.881 ± 0.015	0.902 ± 0.025	0.968 ± 0.004
A4	120	97.561 ± 1.836	0.821 ± 0.015	0.853 ± 0.025	0.95 ± 0.005

	$k = k'$	$\tilde{k}_c$	precision	sensitivity	specificity
A1	75	67.418 ± 0.9	0.982 ± 0.006	0.991 ± 0.015	0.994 ± 0.002
A2	130	100.217 ± 2.127	0.984 ± 0.012	0.894 ± 0.028	0.995 ± 0.004
A3	120	110.352 ± 1.473	0.878 ± 0.017	0.9 ± 0.024	0.967 ± 0.004
A4	120	97.561 ± 1.836	0.84 ± 0.018	0.853 ± 0.026	0.951 ± 0.004

Table 6: Mean ($\pm$ standard deviation) computed across the 30 independent replications of $\tilde{k}_r$ or $\tilde{k}_c$, precision, sensitivity and specificity of the m-specific matchings (left) and n-specific matchings (right) averaged across all mRNAs (left) and all miRNAs (right).

	the WTOT(...) algorithms					the CCOT(...) algorithms	
	WTOT-SCC1*	WTOT-SCC1	WTOT-SCC2*	WTOT-SCC2	WTOT-BC*	CCOT-GWD	CCOT-GWB
B1	0.062 ± 0.151	0.204 ± 0.221	0.082 ± 0.161	0.204 ± 0.221	0.049 ± 0.125	0.276 ± 0.204	0.53 ± 0.168
B2	0.114 ± 0.108	0.418 ± 0.265	0.178 ± 0.207	0.455 ± 0.258	0.382 ± 0.121	0.477 ± 0.14	0.523 ± 0.115
B3	0.175 ± 0.086	0.724 ± 0.236	0.163 ± 0.082	0.775 ± 0.176	–	0.858 ± 0.042	0.867 ± 0.044
B4	0.174 ± 0.092	0.747 ± 0.196	0.171 ± 0.112	0.782 ± 0.159	–	0.882 ± 0.041	0.883 ± 0.04

Table 7: Mean ($\pm$ standard deviation) computed across the 30 independent replications of the co-clustering discrepancy obtained for configurations B1, B2, B3, B4.

	$k = k'$	$\tilde{k}_r$	precision	sensitivity	specificity
B1	60	48.578 ± 5.201	0.885 ± 0.209	0.658 ± 0.191	0.985 ± 0.025
B1	80	63.96 ± 6.126	0.851 ± 0.199	0.816 ± 0.222	0.968 ± 0.03
B1	85	67.537 ± 6.193	0.837 ± 0.193	0.842 ± 0.222	0.961 ± 0.031
B1	90	71.214 ± 6.208	0.823 ± 0.186	0.864 ± 0.219	0.953 ± 0.031
B1	110	85.833 ± 6.358	0.753 ± 0.156	0.918 ± 0.202	0.913 ± 0.029

	$k = k'$	$\tilde{k}_r$	precision	sensitivity	specificity
B4	10	6.78 ± 0.259	0.926 ± 0.102	0.321 ± 0.046	0.999 ± 0.001
B4	20	15.163 ± 0.619	0.873 ± 0.091	0.72 ± 0.087	0.996 ± 0.003
B4	25	19.033 ± 0.784	0.817 ± 0.084	0.837 ± 0.084	0.991 ± 0.004
B4	30	22.889 ± 0.997	0.754 ± 0.076	0.907 ± 0.077	0.984 ± 0.005
B4	40	31.118 ± 1.086	0.618 ± 0.053	0.969 ± 0.049	0.963 ± 0.005

Table 8: Mean ($\pm$ standard deviation) computed across the 30 independent replications of $\tilde{k}_r$, precision, sensitivity and specificity of the m-specific matchings averaged across all mRNAs for configurations B1 (left) and B4 (right).

	$k = k'$	$\tilde{k}_r$	precision	sensitivity	specificity
B1	85	67.537 ± 6.193	0.837 ± 0.193	0.842 ± 0.222	0.961 ± 0.031
B2	60	48.282 ± 3.449	0.751 ± 0.194	0.838 ± 0.2	0.979 ± 0.022
B3	25	19.546 ± 1.151	0.833 ± 0.136	0.837 ± 0.152	0.992 ± 0.006
B4	25	19.033 ± 0.784	0.817 ± 0.084	0.837 ± 0.084	0.991 ± 0.004

	$k = k'$	$\tilde{k}_c$	precision	sensitivity	specificity
B1	85	63.732 ± 8.642	0.844 ± 0.175	0.836 ± 0.229	0.96 ± 0.033
B2	60	44.349 ± 2.495	0.792 ± 0.218	0.819 ± 0.227	0.971 ± 0.024
B3	25	18.766 ± 0.97	0.847 ± 0.125	0.833 ± 0.152	0.991 ± 0.005
B4	25	18.833 ± 0.793	0.834 ± 0.087	0.827 ± 0.099	0.99 ± 0.005

Table 9: Mean ($\pm$ standard deviation) computed across the 30 independent replications of $\tilde{k}_r$ or $\tilde{k}_c$, precision, sensitivity and specificity of the m-specific matchings (left) and n-specific matchings (right) averaged across all mRNAs (left) and all miRNAs (right).

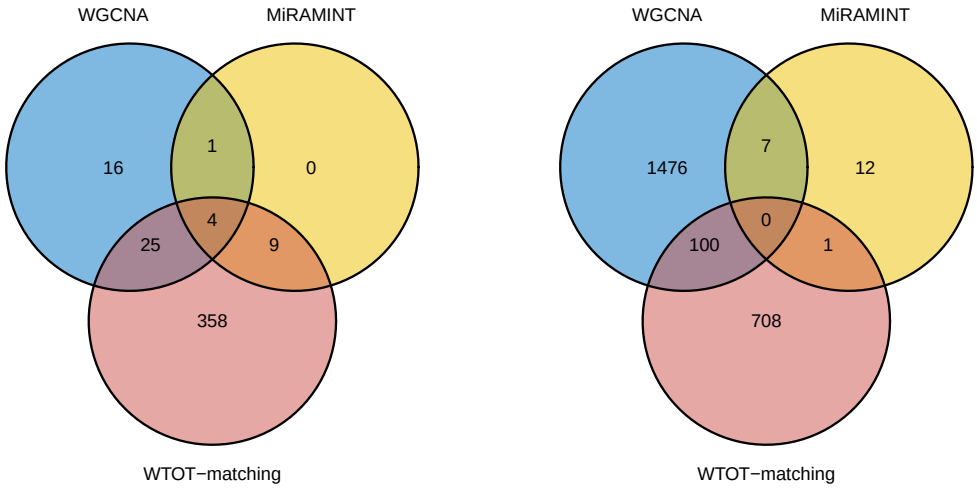

Figure 5: Venn diagrams summarizing the overlaps between the sets of miRNAs (left) and mRNAs (right) which belong to a pair output by the WGCNA, MiRAMINT and WTOT-matching algorithms.

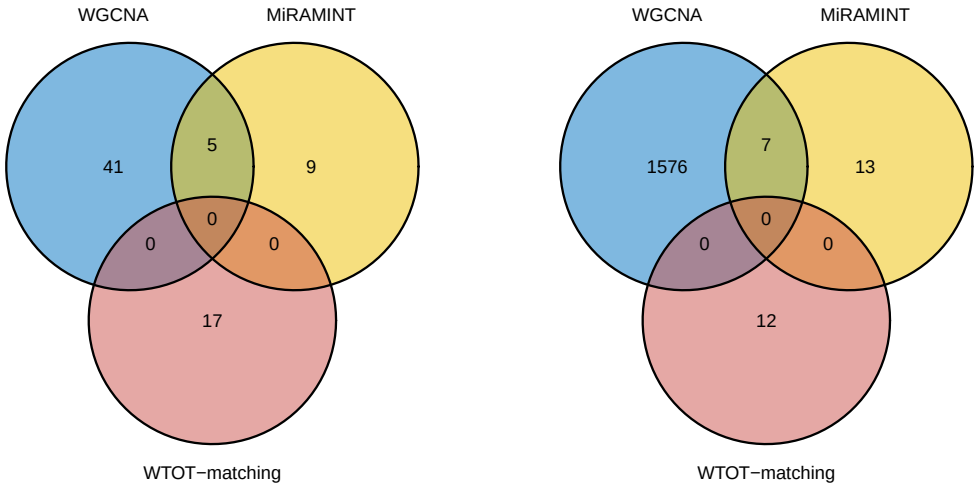

Figure 6: Venn diagrams summarizing the overlaps between the sets of miRNAs (left) and mRNAs (right) which belong to a pair output by the WGCNA, MiRAMINT and WTOT-matching algorithms, *focusing on the WTOT-matching matchings labeled as peaked.*

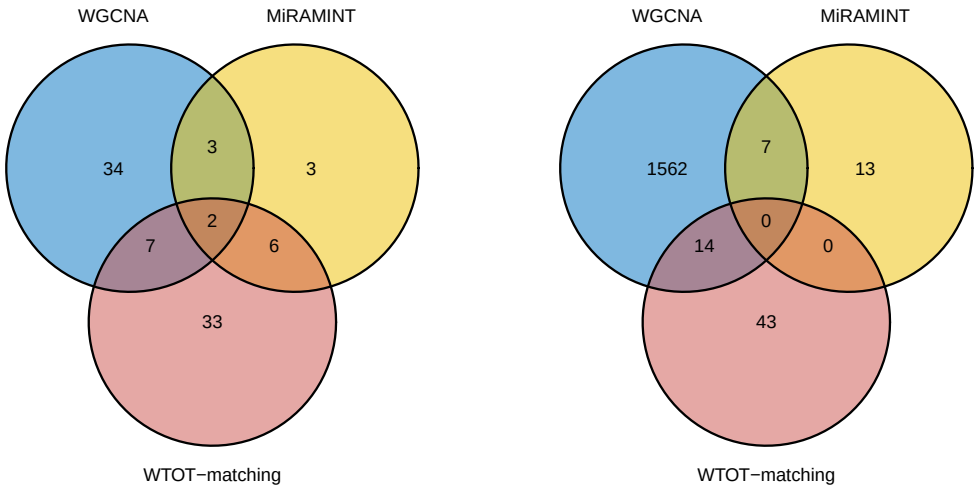

Figure 7: Venn diagrams summarizing the overlaps between the sets of miRNAs (left) and mRNAs (right) which belong to a pair output by the WGCNA, MiRAMINT and WTOT-matching algorithms, *focusing on the WTOT-matching matchings labeled as monotonic.*

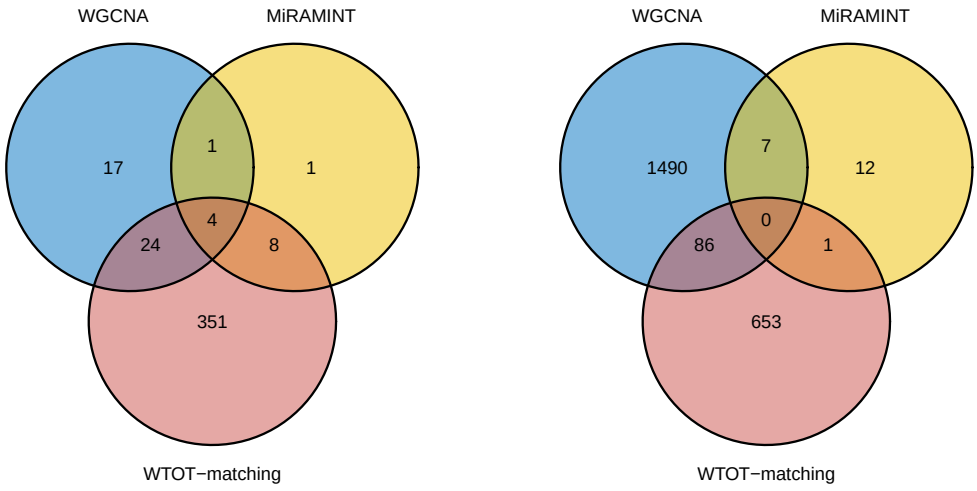

Figure 8: Venn diagrams summarizing the overlaps between the sets of miRNAs (left) and mRNAs (right) which belong to a pair output by the WGCNA, MiRAMINT and WTOT-matching algorithms, *focusing on the WTOT-matching matchings which are labeled as neither peaked nor monotonic.*

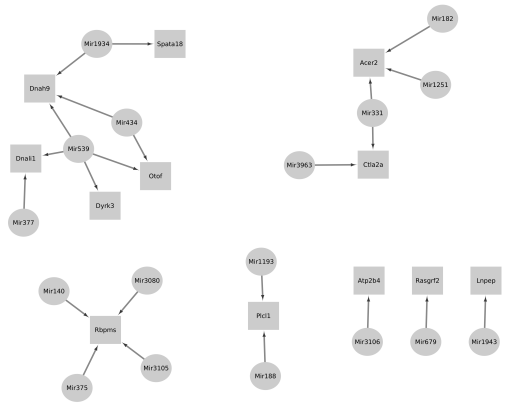

Figure 9: The mRNA-miRNA networks based on the mRNA-miRNA matchings output by the WTOT-matching algorithm, *focusing on the matchings which are labeled as peaked*. Disks correspond to miRNAs and squares to mRNAs. The top annotation is *conventional motile cilium* (GO:0097729, 3 hits).

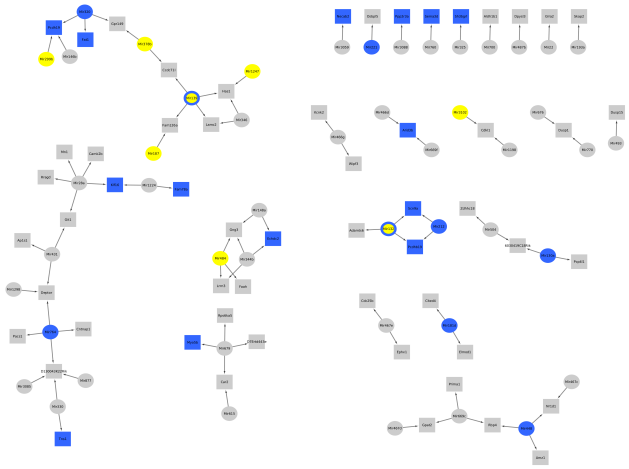

Figure 10: The mRNA-miRNA networks based on the mRNA-miRNA matchings output by the WTOT-matching algorithm, *focusing on the matchings which are labeled as monotonic*. Disks correspond to miRNAs and squares to mRNAs. Elements also retained by the WGCNA algorithm (respectively, the MiRAMINT algorithm) are indicated in blue (respectively, yellow). The top annotation is *mitigation of host antiviral defense response* (GO:0050690, 2 hits).

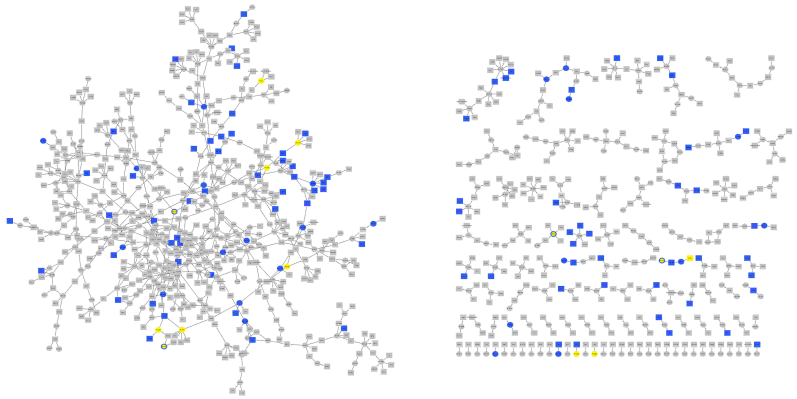

Figure 11: The mRNA-miRNA networks based on the mRNA-miRNA matchings output by the WTOT-matching algorithm, *focusing on the matchings which are labeled as neither peaked nor monotonic*. Disks correspond to miRNAs and squares to mRNAs. Elements also retained by the WGCNA algorithm (respectively, the MiRAMINT algorithm) are indicated in blue (respectively, yellow). The top annotation is *extracellular matrix organization* (GO:0030198, 22 hits).

References

M. Achour, S. Le Gras, C. Keime, F. Parmentier, F.-X. Lejeune, A.-L. Boutillier, C. Neri, I. Davidson, and K. Merienne. Neuronal identity genes regulated by super-enhancers are preferentially down-regulated in the striatum of Huntington's disease mice. *Human Molecular Genetics*, 24(12):3481–3496, 2015.

A. Baddeley and R. Turner. spatstat: An R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12(6):1–42, 2005.

C. Laclau, I. Redko, B. Matei, Y. Bennani, and V. Brault. Co-clustering through Optimal Transport. In *34th International Conference on Machine Learning*, volume 70, pages 1955–1964, Sydney, Australia, Aug. 2017.

F. Lavancier, J. Møller, and E. Rubak. Determinantal point process models and statistical inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 77(4):853–877, 2015.

S. P. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28:129–137, 1982.

L. Mégret, S. Sasidharan Nair, J. Dancourt, J. Aaronson, J. Rosinski, and C. Neri. Combining feature selection and shape analysis uncovers precise rules for miRNA regulation in Huntington's disease mice. *BMC Bioinformatics*, 21(1):75, 2020.

I. G. Olmo, R. P. Olmo, A. N. A. Gonçalves, R. G. W. Pires, J. T. Marques, and F. M. Ribeiro. High-throughput sequencing of BACHD mice reveals upregulation of neuroprotective miRNAs at the pre-symptomatic stage of Huntington's disease. *ASN Neuro*, 13:17590914211009857, 2021.

S. Petry, R. Keraudren, B. Nateghi, A. Loiselle, K. Pircs, J. Jakobsson, C. Sephton, M. Langlois, I. St-Amour, and S. S. Hébert. Widespread alterations in microRNA biogenesis in human Huntington's disease putamen. *Acta Neuropathologica Communications*, 10(1):1–11, 2022.

G. Peyré, M. Cuturi, and J. Solomon. Gromov-Wasserstein Averaging of Kernel and Distance Matrices. In *ICML 2016*, Proc. 33rd International Conference on Machine Learning, New-York, United States, June 2016.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2022. URL https://www.R-project.org/.

Z. J. Rutnam, T. N. Wight, and B. B. Yang. miRNAs regulate expression and function of extracellular matrix molecules. *Matrix Biology*, 32(2):74–85, 2013.

List of Figures