Supporting Information

The supporting information is divided into the following sections:

1. An example of computing W_1 and W_2

2. Additional simulations for asymptotic convergence

3. Simulation details and additional results

4. Additional information and results for the yeast cell cycle data

5. Additional information and results for the *Arabidopsis* data

6. Proofs of the theorems

1 An Example of W_1 and W_2

Suppose $\mathbf{x} = (1, 3, 4, 2, 5)$, $\mathbf{y} = (1, 4, 5, 2, 3)$, and we are interested in computing W_1 and W_2 counts for $k = 3$. For W_1, there are three possible positions to start a contiguous subsequence of length 3, and only the ones starting at position 1 and 2 have the same ranks in $\mathbf{x}$ and $\mathbf{y}$. There are no pairs of contiguous subsequences of length 3 with reverse ranks. Hence $W_1 = 2$. To compute W_2, first sort $\mathbf{y}$ in an ascending order with permutation σ. Applying σ to $\mathbf{x}$ we have $\sigma(\mathbf{x}) = (1, 2, 5, 3, 4)$. The total number of increasing subsequences of length 3 in $\sigma(\mathbf{x})$ is 5, and there are no decreasing subsequences of length 3. Hence $W_2 = 5$.

2 Asymptotic Convergence

Figure S1 shows the convergence of the empirical quantiles of T_1 and T_2 toward the theoretical standard normal quantiles as n increases. Note due to the fact that T_1 can only take $n - k + 2$ possible values, it is easy to produce ties. To examine the asymptotic power of the two statistics under alternative distributions described previously, we generated data that i) were partially coupled time series with the length of dependence $m = n/10$; ii) followed an exact functional relationship with six monotonic pieces, and computed the average power at 5% significance level over 500 iterations. The results for different k and n are shown in Table S1. As predicted by the theoretical analysis, larger k results in better power and T_1 is more powerful than T_2 on the time-course data. In all the cases, as n increases the power tends to 1. The table also displays the average power for the corresponding null distributions of i) and ii) when the two data vectors are independent. Some values are slightly larger than 0.05 due to the heavier tails of the empirical distributions.

3 Simulations

The estimates for Hoeffding's D, dCov, the Renyi correlation, MI and MIC were computed using relevant R packages (`Hmisc` [1], `energy` [2], `acepack` [3], `parmigene` [4] and `minerva` [5]). We used the
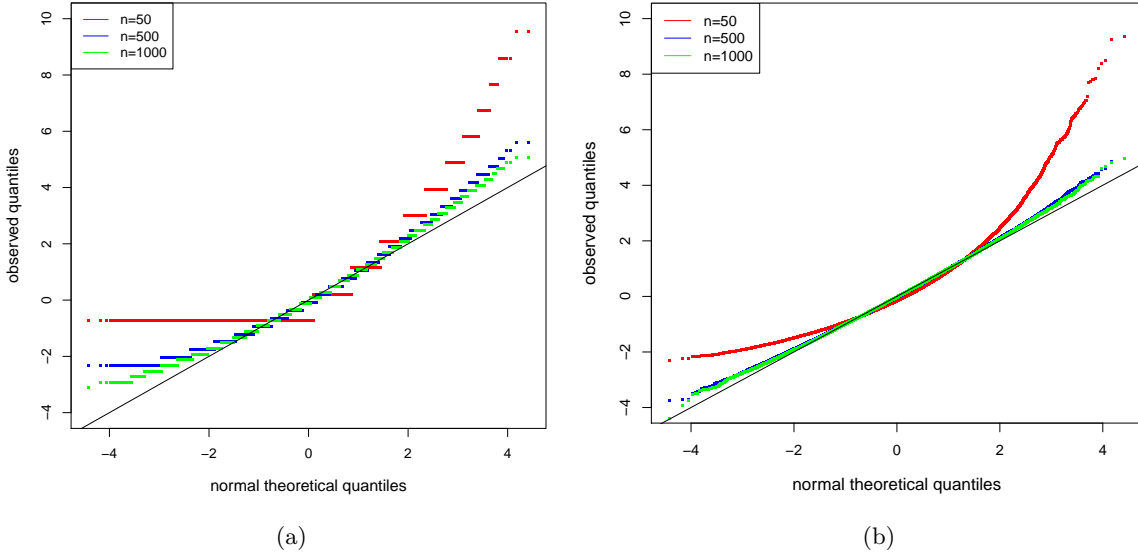
Figure S1: Empirical quantiles for the standardized counts (a) T_1 and (b) T_2 for $n = 50, 500$ and 1000, $k = 5$, from 10^5 simulated random permutations.

k/n	100	200	300	400	500
3	0.332	0.562	0.690	0.812	0.902
4	0.636	0.976	1	1	1
5	1	1	1	1	1
6	1	1	1	1	1

(a) Power of T_1

k/n	100	200	300	400	500
3	0.070	0.050	0.056	0.038	0.036
4	0.060	0.044	0.030	0.054	0.052
5	0.038	0.074	0.064	0.058	0.060
6	0.052	0.042	0.064	0.046	0.076

(b) Power of T_1

k/n	100	200	300	400	500
3	0.302	0.504	0.658	0.796	0.848
4	0.340	0.568	0.726	0.844	0.908
5	0.360	0.650	0.798	0.892	0.952
6	0.392	0.734	0.882	0.924	0.982

(c) Power of T_2

k/n	100	200	300	400	500
3	0.048	0.054	0.076	0.052	0.050
4	0.068	0.060	0.044	0.072	0.040
5	0.042	0.056	0.040	0.070	0.038
6	0.050	0.068	0.056	0.046	0.046

(d) Power of T_2

k/n	100	200	300	400	500
3	0.516	0.784	0.884	0.952	0.972
4	0.710	0.952	0.996	1	1
5	0.866	0.992	1	1	1
6	0.946	1	1	1	1

(e) Power of T_2

k/n	100	200	300	400	500
3	0.058	0.048	0.062	0.044	0.052
4	0.058	0.062	0.044	0.060	0.052
5	0.078	0.040	0.060	0.062	0.030
6	0.070	0.060	0.064	0.062	0.054

(f) Power of T_2

Table S1: Power at 5% significance level for different choices of k and n when $\mathbf{x}$ and $\mathbf{y}$ are: (a), (c) two independent AR(1) time series (with coefficients 0.1 and -0.2 respectively) but $(x_1, \ldots, x_m) = (y_1, \ldots, y_m)$ with $m = n/10$; (e) $x_i \overset{iid}{\sim} Unif(0,1)$, $y_i = \cos(6\pi x_i)$. The right panel shows the power under the corresponding null distributions: (b), (d) $\mathbf{x}$ and $\mathbf{y}$ are two independent AR(1) time series (with coefficients 0.1 and -0.2 respectively); (f) $\mathbf{x}$ and $\mathbf{y}$ are iid $Unif(0,1)$.

2

standard ACE estimate ([6]) for approximating the Renyi correlation. The computation of some of the measures involve tuning parameters. The MI estimates were computed using the kth nearest neighbor (KNN) algorithm of [7]. A number of bandwidth parameters were tried (6, 10 and 20) and the results corresponding to the one with the best power (20) are shown. The MIC estimates were computed using the R package `minerva` with default parameter settings. For statistics with unknown asymptotic distributions (dCov, ACE and MI), the p-values were calculated by a permutation procedure. For each dataset generated the same statistics were calculated on a null dataset obtained by permuting the orders of y_i. The power was taken to be the fraction of datasets with a statistic value more significant than 95% of the values produced by the permuted datasets. Pre-computed p-values of MIC were downloaded from http://www.exploredata.net/Downloads/P-Value-Tables.

Descriptions of the parameters used for the four types of dependence relationships are given in Table S2. 2000 datasets were generated for every scenario with $i \in \{1, \ldots, 220\}$, $e_i \overset{iid}{\sim} N(0,1)$ for the first three relationships and $e_i \overset{iid}{\sim} N(0, 0.5)$ for the time-course relationship. Outliers were created by randomly choosing a fraction of the data and replacing e_i with η_i.

	x_i	y_i	η_i
Linear	$x_i \overset{iid}{\sim} N(0,1)$	$y_i = x_i + 2e_i$	$\eta_i \overset{iid}{\sim} N(0,5)$
Quadratic	$x_i \overset{iid}{\sim} N(0,1)$	$y_i = x_i^2 + 2e_i$	$\eta_i \overset{iid}{\sim} N(0,5)$
Cross	$x_i \overset{iid}{\sim} N(0,1)$	$y_i = \begin{cases} \frac{1}{2} + x_i + e_i \text{ with probability } \frac{1}{2}, \\ \frac{3}{2} - x_i + e_i \text{ with probability } \frac{1}{2}. \end{cases}$	$\eta_i \overset{iid}{\sim} N(0,3)$
Partially coupled time series	$x_i \sim AR(1)$ with coefficient 0.1	$y_i = \begin{cases} x_i + e_i, \ i \in [1, 30] \\ -x_i + e_i, \ i \in [101, 120] \\ AR(1) \text{ with coefficient -0.2,} \\ \text{independent of } x_i, \text{ otherwise.} \end{cases}$	$\eta_i \overset{iid}{\sim} N(0,3)$

Table S2: Parameters for generating the four types of relationships

Power curves for T_1 and T_2 with different choices of k are shown in Figure S2.

We additionally investigated the power loss on a linear relationship with increasing noise level but no outliers. The results are plotted in Figure S3. The linear relationship was generated with $y = x + \beta e$, where $x \overset{iid}{\sim} N(0,1)$, $e \overset{iid}{\sim} N(0,1)$ and $\beta \in \{1, 2, \ldots, 10\}$. T_2^+ remains the best performing statistic. As expected, Pearson's correlation shows better performance on data with no outliers and is now ranked the second. T_2 still demonstrates less power than Spearman's correlation, dCov and Hoeffding'D, but remains more powerful than Renyi's correlation and MI.

We provide a power comparison between our statistics and the LIS-based statistics ([8]) computed using their R package `LIStest` on simulated data in Figure S4. The four scenarios used the same parameters as described in Table S1, except with $n = 200$ — the largest n allowed by the R package. LIS_Ln represents the LIS; LIS_JLn uses a jackknife version of LIS; and LIS_JLMn uses the longest monotonic subsequence (the maximum of LIS and the longest decreasing subsequence). Overall the power of this class of statistics is not optimal, which can be explained by the relative non-robustness of the length of LIS in the presence of noise and outliers. Furthermore, intuitively LIS-based statistics are better suited to detect global monotonic relationships, which differs from our consideration of potentially changing local dependence patterns. As Figure S4 confirms, their power values are lower for non-monotonic relationships ((b), (c) and (d)).
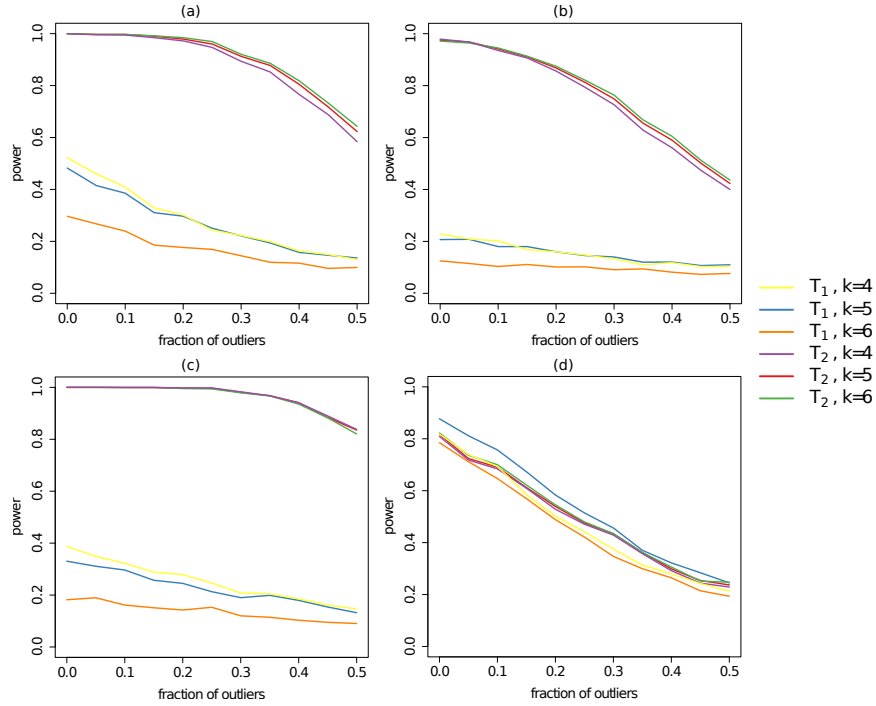
Figure S2: The power of T_1 and T_2 for various k values rejecting at 5% significance level as level of contamination by outliers increases when the bivariate data have (a) a linear relationship; (b) a quadratic relationship; (c) a cross-shaped relationship; (d) are two partially coupled time series.
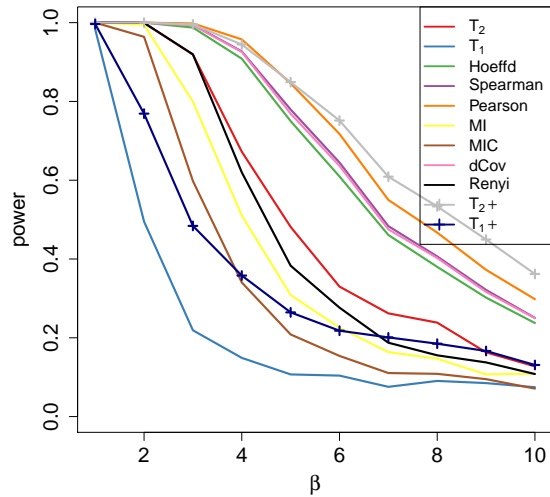

Figure S3: The power of various statistics rejecting at 5% significance level as the noise level increases on a linear relationship.
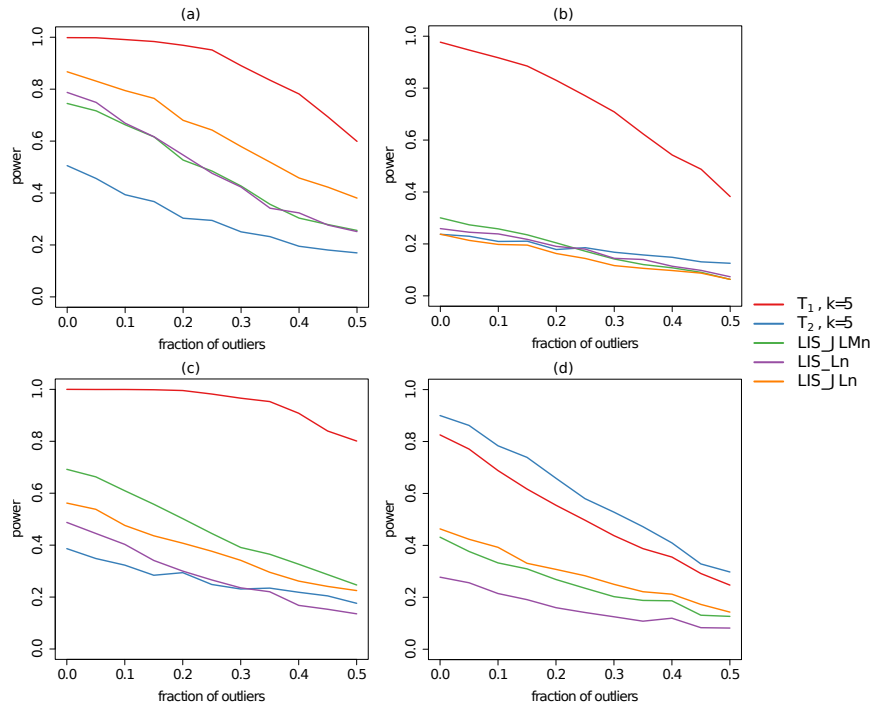
Figure S4: The power of T_1 and T_2 against LIS-based statistics rejecting at 5% significance level as level of contamination by outliers increases when the bivariate data follow (a) a linear relationship; (b) a quadratic relationship; (c) a cross-shaped relationship; (d) two partially coupled time series.

4 Yeast Cell Cycle

The yeast expression data was accessed from http://genome-www.stanford.edu/cellcycle/ and contains the expression levels of 6178 genes from four reasonably long time-course experiments: alpha factor release (18 time points), cdc 15 (24 time points), cdc 28 (17 time points) and elutriation (14 time points). We linearly interpolated some missing data if a point had the two adjacent time points belonging to the same experiment with no missing values. We focused on the coexpression of 133 transcription factors (TFs) with no missing data after interpolation. Since the data has a number of ties, we added small random perturbations for the computation of T_1 and T_2 and took the final results as the maximum counts over 50 iterations.

Figure S5 shows two pairs of TFs (MOT3 and RPN4; PHO2 and SUT1) with genetic interaction identified by T_1 but missed by all the other methods.
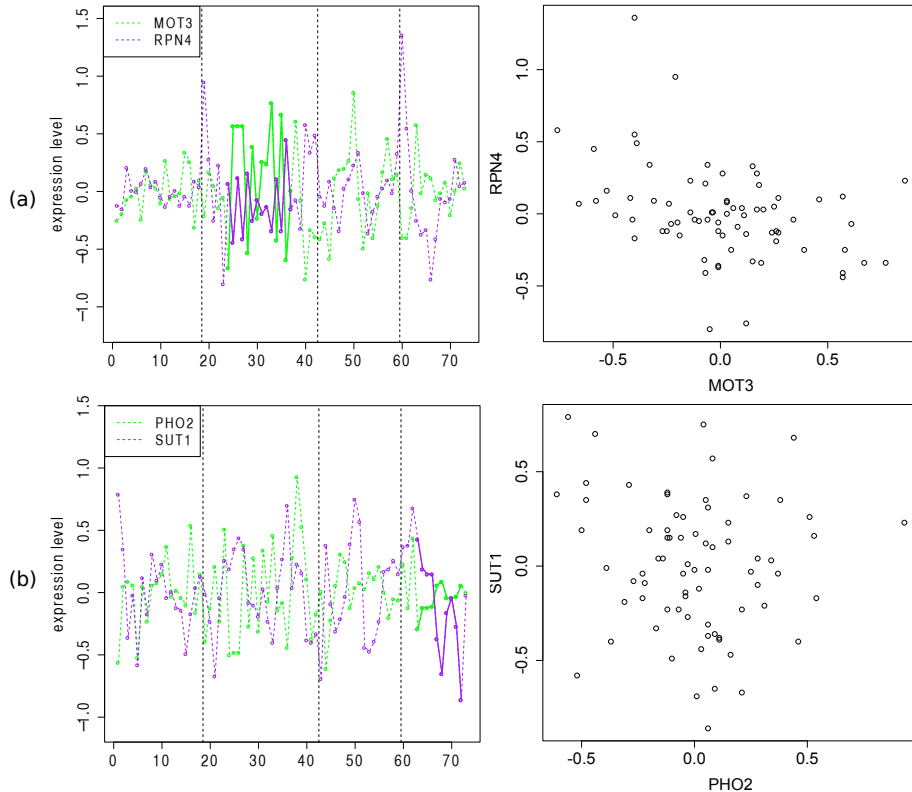

Figure S5: Expression levels of (a) MOT3 and RPN4; (b) PHO2 and SUT1 in four time-course experiments (boundaries indicated by the dashed lines). The solid lines highlight regions contributing to the counts in T_1. Both have reported genetic interactions ([9, 10]), but received low rankings under methods other than T_1.

Table S3 shows the number of known TF interactions among strongly coexpressed pairs as ranked by each method. A number of k values were tried for T_1, while for T_2 only $k = 7$ was shown since the results were quite stable over a range of k values. As T_1 led to many ties, the cutoffs were chosen to include the entire stretches of gene pairs with the same statistic values.

	k = 6				k = 7			k = 8			k = 9		
Top rank	4	7	16	31	4	11	22	5	14	44	3	11	37
Pearson	0	2	2	3	0	2	2	1	2	6	0	2	4
Spearman	0	1	2	2	0	1	2	0	1	2	0	1	2
Hoeffding's D	1	1	1	2	1	1	2	1	1	2	0	1	2
MI	0	1	1	1	0	1	1	1	1	1	0	1	1
MIC	1	1	1	1	1	1	1	1	1	2	1	1	2
dCov	1	1	1	2	1	1	1	1	1	2	1	1	2
Renyi	0	0	2	2	0	2	2	0	2	3	0	2	2
T_1	0	1	3	3	1	3	4	1	3	6	0	1	5
T_2	0	1	2	3	0	2	2	0	2	3	0	2	3

Table S3: Number of known interactions in highly ranked coexpression pairs by various statistics. A range of k values were tested for T_1, and $k = 7$ for T_2.

5 *Arabidopsis* Microarrays

The original CEL files of the microarrays were downloaded from NCBI GEO (GSE 5623, 7636, 7639, 7641, 7642, 8787 and 30166), then normalized using the robust multi-array analysis (RMA) ([11]) function in the Bioconductor package. After normalization, a small fraction of the data were tied. We added small random perturbations for the computation of T_1 and T_2 and took the final results as the maximum count over 20 iterations. We noted that this had negligible influence on all the final results. Asymptotic p-values were computed for T_1, T_2, the Pearson correlation, the Spearman correlation and Hoeffding's D. For dCov, Renyi and MI, null statistic values were calculated by permuting the sample labels of each gene and used as empirical quantiles for determining the significance level of the statistics. Pre-computed p-values for MIC from http://www.exploredata.net/Downloads/P-Value-Tables were used.

Figure S7 shows two pairs of genes in the same pathway, where the bulk of the samples follow a linear trend but they failed to be identified by MI at an unadjusted significance level of 5%. On the other hand, both pairs were assigned significant p-values by T_2 and other statistics including the Pearson and Spearman correlations.

For each pathway, we ranked the coexpression between the pathway genes and all the genes available and chose the top L pairs, where L is the number of total gene pairs in this pathway. We then counted the number of gene pairs belonging to this pathway among the chosen pairs, and kept 20 pathways in which at least one method achieved a significant enrichment of pathway genes using Fisher's exact test. Table S4 tallies the methods with the highest counts of same pathway pairs in these 20 pathways.

6 Proofs

6.1 Running time of the algorithms

Lemma 6.1. *Computing W_1 and W_2 takes $O(k(\log k)n)$ and $O(kn \log n)$ time respectively.*

Proof. Computing W_1 involves ranking and comparing the elements of vectors of length k $O(n)$ number of times, thus the running time is $O(k(\log k)n)$.

W_2 counts the total number of subsequences of length k with matching or reverse rank patterns. For any pair of subsequences with matching rank patterns, permuting the two subsequences simultaneously to sort one of them in an increasing order will also sort the other one in an increasing order. Using this observation, let σ be the permutation that sorts $\mathbf{y}$ in an increasing order and $\mathbf{z} = \sigma(\mathbf{x})$ be that
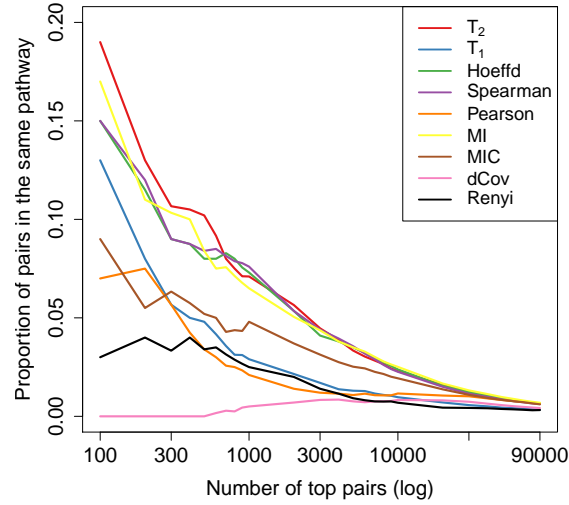
Figure S6: Proportion of gene pairs in the same pathway as a function of the number of highly ranked pairs chosen.
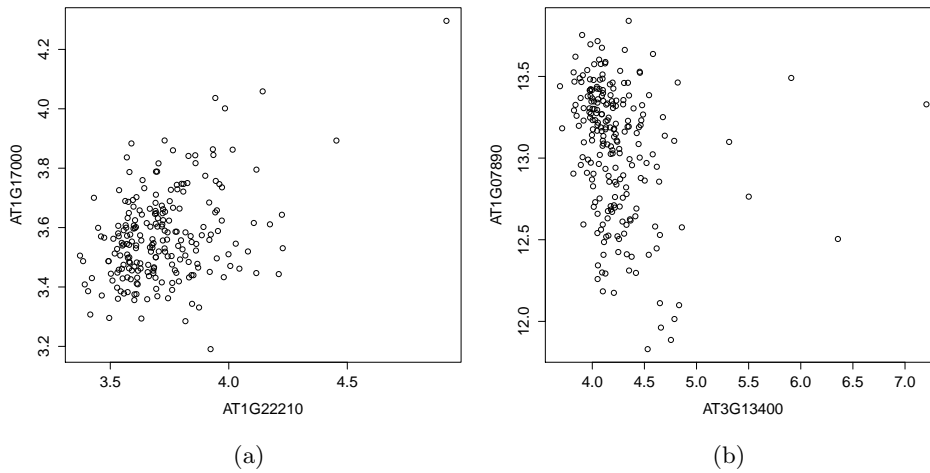

(a)

(b)

Figure S7: Expression levels of two gene pairs in the same pathway showing a linear relationship with outliers which were not identified as statistically significant by MI.

permutation applied to $\mathbf{x}$. Then W_2 is the number of increasing (and decreasing) subsequences of length k in $\mathbf{z}$. To compute W_2, it suffices to consider counting the increasing subsequences. One obvious solution is dynamic programing. Let `dp[i,l]` be the number of increasing subsequences of length `l` ending at position `i`, then the matrix `dp[i,l]` can be updated as follows.

```
Initialize dp[i,l] = 0; dp[i,1] = 1
for i = 2 to n
     for j = 1 to i-1
     if z[i] > z[j]
```

	Individual pathways											Tally
Renyi				✓								1
dcov												0
Hoeffd	✓					✓	✓					3
MI	✓						✓	✓			✓	4
MIC												0
Pearson		✓				✓	✓					3
Spearman	✓					✓		✓				3
T_1			✓		✓				✓		✓	4
T_2	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	12

Table S4: Methods with the highest counts of pathway genes pairs in 20 pathways with statistically significant enrichment

```
for l = 2 to k
    dp[i,l] += dp[j,l-1]
```

The final answer is obtained by summing `dp[i,k]` over i. It is easy to see this has a running time of $O(kn^2)$. Note that in the second loop the only entries involved in the update are `z[j]` whose ranks are smaller than that of `z[i]`. Therefore by first ranking the elements in $\mathbf{z}$, a binary indexed tree structure can be implemented to perform the sum and update efficiently, reducing the running time to $O(kn \log n)$ ([12]). □

6.2 Asymptotic distributions of W_1

Throughout the sections, C and C_i denote positive constants which may be different at each appearance. Without loss of generality assume $\mathbf{x}$ satisfies the assumption that it has an exchangeable distribution. Then the ranks of any subsequence of $\mathbf{x}$ can be treated as a random permutation. Denote

$$
\begin{aligned}
\mathbb{I}_i^+ &= \mathbb{I}(\phi(x_i, \ldots, x_{i+k-1}) = \phi(y_i, \ldots, y_{i+k-1})), \\
\mathbb{I}_i^- &= \mathbb{I}(\phi(x_i, \ldots, x_{i+k-1}) = \phi(-y_i, \ldots, -y_{i+k-1})), \\
\mathbb{I}_i &= \mathbb{I}_i^+ + \mathbb{I}_i^-.
\end{aligned}
\tag{S1}
$$

We have

$$
\begin{aligned}
&\mathbb{E}(\mathbb{I}_i^+) \\
&= \sum_{\mathbf{w}} \mathbb{P}(\phi(x_i, \ldots, x_{i+k-1}) = \mathbf{w} \mid \phi(y_i, \ldots, y_{i+k-1}) = \mathbf{w}) \mathbb{P}(\phi(y_i, \ldots, y_{i+k-1}) = \mathbf{w}) \\
&= \frac{1}{k!} \sum_{\mathbf{w}} \mathbb{P}(\phi(y_i, \ldots, y_{i+k-1}) = \mathbf{w}) \\
&= \frac{1}{k!}
\end{aligned}
\tag{S2}
$$

by the independence assumption and the fact that there is only one way to arrange a list of numbers in a given order. Clearly also $\mathbb{E}(\mathbb{I}_i^-) = 1/k!$. In the next lemma, we characterize the behavior of the cross terms $\mathbb{E}(\mathbb{I}_i^+ \mathbb{I}_j^+)$.

Lemma 6.2. *1. When $|j - i| \geq k$, $\mathbb{I}_i^+$ and $\mathbb{I}_j^+$ are independent. So are $(\mathbb{I}_i^+, \mathbb{I}_j^-)$ and $(\mathbb{I}_i^-, \mathbb{I}_j^-)$.*

9

2. *When* $|j - i| = k - l$ *with* $1 \leq l \leq k - 1$,

$$\frac{1}{(2k - l)!} \leq \mathbb{E}(\mathbb{I}_i^+ \mathbb{I}_j^+) \leq \frac{\binom{2k-2l}{k-l}}{(2k - l)!}. \tag{S3}$$

The same conclusions hold for $(\mathbb{I}_i^-, \mathbb{I}_j^-)$, $1 \leq |j - i| < k$, *and* $(\mathbb{I}_i^+, \mathbb{I}_j^-)$, $(\mathbb{I}_i^-, \mathbb{I}_j^+)$, $|i - j| = k - 1$.

3. $\mathbb{E}(\mathbb{I}_i^+ \mathbb{I}_j^-) = \mathbb{E}(\mathbb{I}_i^- \mathbb{I}_j^+) = 0$ *for* $1 \leq |i - j| < k - 1$.

Proof. Note that conditioning on the sequence $\mathbf{y}$,

$$\mathbb{E}(\mathbb{I}_i^+ \mathbb{I}_j^+) = \sum_{\mathbf{w},\mathbf{v}} \mathbb{P}(\phi(x_i, \ldots, x_{i+k-1}) = \mathbf{w}, \phi(x_j, \ldots, x_{j+k-1}) = \mathbf{v})$$

$$\times \mathbb{P}(\phi(y_i, \ldots, y_{i+k-1}) = \mathbf{w}, \phi(y_j, \ldots, y_{j+k-1}) = \mathbf{v}). \tag{S4}$$

For $|j - i| \geq k$, the subsequences $(x_i, \ldots, x_{i+k-1})$ and $(x_j, \ldots, x_{j+k-1})$ do not overlap. Thus their local rank patterns are independent, each having probability $1/k!$ for a given order.

$$\mathbb{E}(\mathbb{I}_i^+ \mathbb{I}_j^+) = \left(\frac{1}{k!}\right)^2 \sum_{\mathbf{w},\mathbf{v}} P(\phi(y_i, \ldots, y_{i+k-1}) = \mathbf{w}, \phi(y_j, \ldots, y_{j+k-1}) = \mathbf{v})$$

$$= \left(\frac{1}{k!}\right)^2 = \mathbb{E}(\mathbb{I}_i^+) \mathbb{E}(\mathbb{I}_j^+). \tag{S5}$$

For $j - i = k - l < k$ (assuming WLOG $j > i$), $(x_i, \ldots, x_{i+k-1})$ and $(x_j, \ldots, x_{j+k-1})$ form a contiguous subsequence $x_i, \ldots, x_j, \ldots, x_{j+k-1}$. Suppose $\phi(x_i, \ldots, x_{j+k-1}) = (u_1, \ldots, u_{2k-l})$, then

$$\phi(u_1, \ldots, u_k) = (w_1, \ldots, w_k),$$
$$\phi(u_{k-l+1}, \ldots, u_{2k-l}) = (v_1, \ldots, v_k),$$
$$\phi(u_{k-l+1}, \ldots, u_k) = \phi(w_{k-l+1}, \ldots, w_k) = \phi(v_1, \ldots, v_l)$$
$$:= (o_1, \ldots, o_l), \qquad \text{say.} \tag{S6}$$

Focusing on the overlapping part (u_{k-l+s}) for $1 \leq s \leq l$, the numbers of elements smaller than u_{k-l+s} in the subsequences $(u_1, \ldots, u_k)$, $(u_{k-l+1}, \ldots, u_{2k-l})$ and $(u_{k-l+1}, \ldots, u_k)$ are $w_{k-l+s} - 1$, $v_s - 1$, and $o_s - 1$, respectively. Given the overall rank u_{k-l+s} in the sequence $(u_1, \ldots, u_{k-l+1}, \ldots, u_k, \ldots, u_{2k-l})$, we have

$$u_{k-l+s} - 1 = (w_{k-l+s} - 1) + (v_s - 1) - (o_s - 1), \tag{S7}$$

since the elements in the overlapping part are counted twice. In other words, the overlapping part (u_{k-l+s}) for $1 \leq s \leq l$ is fixed, and there are at most $\binom{2k-2l}{k-l}$ ways of arranging the rest $2k - 2l$ numbers. Thus we arrive at the upper bound in (S3). The lower bound is trivial. The same arguments hold for $(\mathbb{I}_i^-, \mathbb{I}_j^-)$, $1 \leq |j - i| < k$, and $(\mathbb{I}_i^+, \mathbb{I}_j^-)$, $(\mathbb{I}_i^-, \mathbb{I}_j^+)$, $|i - j| = k - 1$.

Lastly, for $1 \leq |i - j| < k - 1$, $\mathbb{E}(\mathbb{I}_i^+ \mathbb{I}_j^-) = \mathbb{E}(\mathbb{I}_i^+ \mathbb{I}_j^-) = 0$ since no such arrangements of the elements are possible. $\square$

Let N_i denote the dependency neighborhood of $\mathbb{I}_i$, the next lemma tries to bound a key quantity in the variance calculation.

Lemma 6.3. *For all* $k \geq 3$,

$$4(n - 2k + 2) \left(\sum_{l=2}^{k-1} \frac{1}{(2k - l)!} + \frac{2}{(2k - 1)!}\right) \leq \sum_{i=1}^{n-k+1} \sum_{j \in N_i \setminus \{i\}} \mathbb{E}(\mathbb{I}_i \mathbb{I}_j) \leq \frac{C(n - k + 1)}{(k + 1)!} \tag{S8}$$

for some $C > 0$.

Proof. First note that

$$\sum_{i=1}^{n-k+1} \sum_{j \in N_i \setminus \{i\}} \mathbb{E}(\mathbb{I}_i \mathbb{I}_j) = 2 \sum_{i=1}^{n-k+1} \sum_{j \in N_i \setminus \{i\}} \mathbb{E}(\mathbb{I}_i^+ \mathbb{I}_j^+) + 2 \sum_{i=1}^{n-k+1} \sum_{|j-i|=k-1} \mathbb{E}(\mathbb{I}_i^+ \mathbb{I}_j^-)$$

$$\leq 8(n-k+1) \sum_{l=1}^{k-1} \gamma_l \tag{S9}$$

by (S3), where

$$\gamma_l = \frac{\binom{2k-2l}{k-l}}{(2k-l)!}. \tag{S10}$$

It remains to bound $\sum_{l=1}^{k-1} \gamma_l$. Taking the ratio of successive terms,

$$\begin{aligned} r_l = \frac{\gamma_{l+1}}{\gamma_l} &= \frac{\binom{2k-2l-2}{k-l-1}}{(2k-l-1)!} \cdot \frac{(2k-l)!}{\binom{2k-2l}{k-l}} \\ &= \frac{(k-l)^2 (2k-l)}{(2k-2l)(2k-2l-1)} \\ &= \frac{(k-l)(2k-l)}{2(2k-2l-1)}, \quad l = 1, \dots, k-2. \end{aligned} \tag{S11}$$

For all $k \geq 3$, there exists positive constant C_1 and C_2 (independent of k) such that

$$C_1 k \leq r_l \leq C_2 k, \quad l = 1, \dots, k-2. \tag{S12}$$

Therefore $\sum_{l=1}^{k-1} \gamma_l$ is upper bounded by

$$\begin{aligned} \sum_{l=1}^{k-1} \gamma_l &\leq \gamma_{k-1} \sum_{l=0}^{k-2} \left(\frac{1}{C_1 k} \right)^l \\ &= \gamma_{k-1} \cdot \frac{1 - \left(\frac{1}{C_1 k} \right)^{k-1}}{1 - \frac{1}{C_1 k}} \\ &\leq \frac{C}{(k+1)!} \end{aligned} \tag{S13}$$

for some $C > 0$. Equations (S13) and (S9) give the required upper bound.

For the lower bound, it is easy to see

$$\sum_{i=1}^{n-k+1} \sum_{j \in N_i \setminus \{i\}} \mathbb{E}(\mathbb{I}_i \mathbb{I}_j) = 2 \sum_{i=1}^{n-k+1} \sum_{j \in N_i \setminus \{i\}} \mathbb{E}(\mathbb{I}_i^+ \mathbb{I}_j^+) + 2 \sum_{i=1}^{n-k+1} \sum_{|j-i|=k-1} \mathbb{E}(\mathbb{I}_i^+ \mathbb{I}_j^-)$$

$$\geq 2 \left[2(n-k+1) - 2(k-1) \right] \left(\sum_{l=1}^{k-1} \frac{1}{(2k-l)!} + \frac{1}{(2k-1)!} \right)$$

$$\geq 4(n-2k+2) \left(\sum_{l=2}^{k-1} \frac{1}{(2k-l)!} + \frac{2}{(2k-1)!} \right) \tag{S14}$$

by the lower bound in (S3). $\qquad \square$

11

With the above bounds we can now prove Theorem 1.

Proof of Theorem 1. In order to use Stein's method for normal approximation, we first give a lower bound of the variance. Note that

$$
\begin{aligned}
\sigma_{1,n}^2 &= \sum_{i=1}^{n-k+1} \sum_{j \in N_i} \left(\mathbb{E}(\mathbb{I}_i \mathbb{I}_j) - (\mathbb{E}\mathbb{I}_i)(\mathbb{E}\mathbb{I}_j) \right) \\
&= \sum_{i=1}^{n-k+1} \sum_{j \in N_i \setminus \{i\}} \mathbb{E}(\mathbb{I}_i \mathbb{I}_j) + \frac{2(n-k+1)}{k!} - \sum_{i=1}^{n-k+1} \sum_{j \in N_i} \mathbb{E}(\mathbb{I}_i)\mathbb{E}(\mathbb{I}_j) \\
&\geq 4(n-2k+2)\left(\sum_{l=2}^{k-1} \frac{1}{(2k-l)!} + \frac{2}{(2k-1)!} \right) + \frac{2(n-k+1)}{k!} - \frac{4(n-k+1)(2k-1)}{(k!)^2}.
\end{aligned} \tag{S15}
$$

by (S8). For k such that $k/n \to 0$, when n is sufficiently large, $\sigma_{1,n}^2$ is lower bounded by the dominating terms

$$
\begin{aligned}
\sigma_{1,n}^2 &\geq C_1 \left(4n \left(\sum_{l=2}^{k-1} \frac{1}{(2k-l)!} + \frac{2}{(2k-1)!} \right) + \frac{2n}{k!} - \frac{4n(2k-1)}{(k!)^2} \right) \\
&= \frac{2C_1 n}{k!} \left(2\left(\frac{1}{k+1} + \frac{1}{(k+2)(k+1)} + \cdots + \frac{2}{(2k-1)\cdots(k+1)} \right) + 1 - \frac{2(2k-1)}{k!} \right) \\
&\geq \frac{C_2 n}{k!}
\end{aligned} \tag{S16}
$$

for some $C_1, C_2 > 0$ and all $k \geq 3$. One version of Stein's method gives the following error bound for normal approximation ([13]),

$$
d_W(T_1, Y) \leq \frac{D^2}{\sigma_{1,n}^3} \sum_{i=1}^{n-k+1} \mathbb{E}|\mathbb{I}_i - 2/k!|^3 + \frac{\sqrt{26}D^{3/2}}{\sqrt{\pi}\sigma_{1,n}^2} \sqrt{\sum_{i=1}^{n-k+1} \mathbb{E}|\mathbb{I}_i - 2/k!|^4} \tag{S17}
$$

where d_W is the Wasserstein metric, $Y \sim N(0,1)$ and $D = \max_i N_i = 2k-1$. This can be further bounded by

$$
\begin{aligned}
&C_1 \cdot \frac{D^2 \mu_{1,n}}{\sigma_{1,n}^3} + C_2 \cdot \frac{D^{3/2} \mu_{1,n}^{1/2}}{\sigma_{1,n}^2} \\
&\leq C \cdot \frac{D^2 \mu_{1,n}}{\sigma_{1,n}^3} \\
&\leq C \cdot \frac{k^2 \sqrt{k!}}{\sqrt{n}} \to 0
\end{aligned} \tag{S18}
$$

using (S16) for $k/(\log n)^\alpha \to 0$, $\alpha < 1$.

The Chen-Stein method yields the following error bound for Poisson approximation,

$$d_{TV}(W_1, Z) \leq \min\{1, \mu_{1,n}^{-1}\} \left(\sum_{i=1}^{n-k+1} \sum_{j \in N_i} \mathbb{E}(\mathbb{I}_i)\mathbb{E}(\mathbb{I}_j) + \sum_{i=1}^{n-k+1} \sum_{j \in N_i \setminus \{i\}} \mathbb{E}(\mathbb{I}_i\mathbb{I}_j) \right)$$

$$\leq \sum_{i=1}^{n-k+1} \sum_{j \in N_i} \mathbb{E}(\mathbb{I}_i)\mathbb{E}(\mathbb{I}_j) + \sum_{i=1}^{n-k+1} \sum_{j \in N_i \setminus \{i\}} \mathbb{E}(\mathbb{I}_i\mathbb{I}_j)$$

$$\leq \frac{4(n-k+1)(2k-1)}{(k!)^2} + \frac{C(n-k+1)}{(k+1)!}$$

$$\leq \frac{C(n-k+1)}{(k+1)!} \tag{S19}$$

for some $C > 0$ and k sufficiently large. For k growing fast enough such that $\mu_{1,n} = O(1)$, the above bound goes to 0. In particular, using Stirling's approximation one can show in the regime $\log n / k = O(1)$ this condition is satisfied. $\qquad\square$

6.3 Asymptotic distribution of W_2

Assuming $\mathbf{x}$ has an exchangeable distribution, the permuted sequence $\sigma(\mathbf{x})$ also has an exchangeable distribution, and its ranks can be treated as a random permutation. For notational simplicity, take $\mathbf{z}$ as a random permutation of $\{1, \ldots, n\}$. For integers $\{i_1, \ldots, i_k\}$ satisfying $1 \leq i_1 < \cdots < i_k \leq n$, define indicators $\mathbb{I}_{i_1, \ldots, i_k}^+(\mathbf{z})$ such that

$$\mathbb{I}_{i_1, \ldots, i_k}^+(\mathbf{z}) = \begin{cases} 1 & (i_1, \ldots, i_k) \text{ is a subsequence of } \mathbf{z}, \\ 0 & \text{otherwise.} \end{cases} \tag{S20}$$

Similarly define

$$\mathbb{I}_{i_1, \ldots, i_k}^-(\mathbf{z}) = \begin{cases} 1 & (i_k, \ldots, i_1) \text{ is a subsequence of } \mathbf{z}, \\ 0 & \text{otherwise.} \end{cases} \tag{S21}$$

Then W_2 can be written as the sum of

$$W_2 = \sum_{1 \leq i_1 < \cdots < i_k \leq n} \mathbb{I}_{i_1, \ldots, i_k}(\mathbf{z}), \tag{S22}$$

where

$$\mathbb{I}_{i_1, \ldots, i_k}(\mathbf{z}) = \mathbb{I}_{i_1, \ldots, i_k}^+(\mathbf{z}) + \mathbb{I}_{i_1, \ldots, i_k}^-(\mathbf{z}). \tag{S23}$$

It is easy to see that if $\{i_1, \ldots, i_k\} \cap \{j_1, \ldots, j_k\} = \emptyset$, $\mathbb{I}_{i_1, \ldots, i_k}(\mathbf{z})$ and $\mathbb{I}_{j_1, \ldots, j_k}(\mathbf{z})$ are independent. The variance of W_2 becomes

$$\text{Var}(W_2)$$
$$= \sum_{\{i_1, \ldots, i_k\} \cap \{j_1, \ldots, j_k\} \neq \emptyset} \{\mathbb{E}(\mathbb{I}_{i_1, \ldots, i_k}(\mathbf{z})\mathbb{I}_{j_1, \ldots, j_k}(\mathbf{z})) - \mathbb{E}(\mathbb{I}_{i_1, \ldots, i_k}(\mathbf{z}))\mathbb{E}(\mathbb{I}_{j_1, \ldots, j_k}(\mathbf{z}))\}$$
$$= 2 \sum_{\{i_1, \ldots, i_k\} \cap \{j_1, \ldots, j_k\} \neq \emptyset} \mathbb{E}(\mathbb{I}_{i_1, \ldots, i_k}^+(\mathbf{z})\mathbb{I}_{j_1, \ldots, j_k}^+(\mathbf{z}))$$
$$+ 2 \sum_{\{i_1, \ldots, i_k\} \cap \{j_1, \ldots, j_k\} \neq \emptyset} \mathbb{E}(\mathbb{I}_{i_1, \ldots, i_k}^+(\mathbf{z})\mathbb{I}_{j_1, \ldots, j_k}^-(\mathbf{z})) - \frac{4D\binom{n}{k}}{(k!)^2}, \tag{S24}$$

since $\mathbb{E}(\mathbb{I}^+(z_{i_1}, \ldots, z_{i_k})) = 1/k!$. Here D is the size of the dependency neighborhood and equals $\binom{n}{k} - \binom{n-k}{k}$.

The sum of the first cross terms can be written as (Proposition 2 in [14])

$$\sum_{\{i_1,\ldots,i_k\} \cap \{j_1,\ldots,j_k\} \neq \emptyset} \mathbb{E}(\mathbb{I}^+_{i_1,\ldots,i_k}(\mathbf{z})\mathbb{I}^+_{j_1,\ldots,j_k}(\mathbf{z}))$$

$$= \sum_{j=1}^{k} \binom{n}{2k-j} \frac{1}{(2k-j)!} A(k-j, j), \tag{S25}$$

where

$$A(N, j) = \sum_{\substack{\sum_{r=0}^{j} l_r = N \\ \sum_{r=0}^{j} m_r = N}} \prod_{r=0}^{j} \left(\frac{(l_r + m_r)!}{l_r! m_r!} \right)^2. \tag{S26}$$

We will be using the following fact about the constants $A(N, j)$ from Lemma 3 in [14].

Fact 6.4. For sufficiently large k, there exists $C > 0$ such that

$$A(k-1, 1) \geq C k^{1/2} \binom{2k-2}{k-1}^2. \tag{S27}$$

It is easy to see for all $k \geq 2$, $A(k-1, 1) > \binom{2k-2}{k-1}^2$.

The sum of the second cross terms reduces to

$$\sum_{|\{i_1,\ldots,i_k\} \cap \{j_1,\ldots,j_k\}|=1} \mathbb{E}(\mathbb{I}^+_{i_1,\ldots,i_k}(\mathbf{z})\mathbb{I}^-_{j_1,\ldots,j_k}(\mathbf{z})),$$

since when the size of the intersection is greater than one, it is impossible to find a permutation $\mathbf{z}$ satisfying both conditions specified by the indicators. Using arguments similar to the proof of Proposition 2 in [14], we can show

$$\sum_{|\{i_1,\ldots,i_k\} \cap \{j_1,\ldots,j_k\}|=1} \mathbb{E}(\mathbb{I}^+_{i_1,\ldots,i_k}(\mathbf{z})\mathbb{I}^-_{j_1,\ldots,j_k}(\mathbf{z})) = \binom{n}{2k-1} \frac{1}{(2k-1)!} B(k), \tag{S28}$$

where

$$B(k) = \sum_{\substack{l_0+l_1=k-1 \\ m_0+m_1=k-1}} \binom{l_0+m_0}{l_0} \binom{l_1+m_1}{l_1} \binom{l_0+m_1}{l_0} \binom{l_1+m_0}{l_1} \tag{S29}$$

Now we can obtain a lower bound on the variance and use the Stein method to prove Theorem 2.

Proof of Theorem 2. From equations (S24), (S25) and (S28), we have

$$\frac{\sigma^2_{2,n}}{\mu^2_{2,n}} \geq \frac{\binom{n}{2k-1}(k!)^2}{2\binom{n}{k}^2(2k-1)!} \left(A(k-1,1) + B(k) \right) - \frac{D}{\binom{n}{k}}$$

$$\geq \frac{k^2}{2n} \cdot \left(1 - \frac{k-1}{n-k+1} \right)^{k-1} \binom{2k-1}{k-1}^{-2} \left(A(k-1,1) + B(k) \right) - \frac{D}{\binom{n}{k}}. \tag{S30}$$

14

For $k \to \infty$ and $k = o(n^{1/2})$, it is easy to check $D/\binom{n}{k} = O(k^2/n)$. Applying Fact 6.4,

$$
\begin{aligned}
\frac{\sigma_{2,n}^2}{\mu_{2,n}^2} &\geq C \cdot \frac{k^{5/2}}{2n} \left(1 - \frac{k-1}{n-k+1}\right)^{k-1} \left[\frac{(2k-2)\cdots k}{(2k-1)\cdots(k+1)}\right]^2 + O(k^2/n) \\
&= C \cdot \frac{k^{5/2}}{2n}(1 + O(k^2/n))\left(\frac{k}{2k-1}\right)^2 + O(k^2/n) \\
&\geq C \cdot \frac{k^{5/2}}{n}
\end{aligned}
\tag{S31}
$$

for some $C > 0$ and sufficiently large k and n. Applying the bound from the Stein method as in equation (S17), we have

$$
\begin{aligned}
d_W(T_2, Y) &\leq C_1 \cdot \frac{D^2 \mu_{2,n}}{\sigma_{2,n}^3} + C_2 \cdot \frac{D^{3/2} \mu_{2,n}^{1/2}}{\sigma_{2,n}^2} \\
&\leq C_1 \cdot \frac{k^{1/4}(k!)^2}{n^{1/2}} + C_2 \cdot \frac{k^{1/2}(k!)^{3/2}}{n^{1/2}} \to 0
\end{aligned}
\tag{S32}
$$

for $k/(\log n)^\alpha \to 0$.

For k fixed, $D/\binom{n}{k} \leq k^2/(n-k+1) + o(1/n)$. (S30) becomes

$$
\begin{aligned}
\frac{\sigma_{2,n}^2}{\mu_{2,n}^2} &\geq \frac{k^2}{2n}(1 + O(1/n))\binom{2k-1}{k-1}^{-2}(A(k-1,1) + B(k)) - \frac{k^2}{n-k+1} + o(1/n) \\
&= \left\{\frac{1}{2}(A(k-1,1) + B(k))\binom{2k-1}{k-1}^{-2} - 1\right\}\frac{k^2}{n} + o(1/n) \\
&:= C(k) \cdot \frac{k^2}{n} + o(1/n), \quad \text{say.}
\end{aligned}
\tag{S33}
$$

When $k = 3$, we can check that $C(3) > 0$ and thus $\sigma_{2,n}^2/\mu_{2,n}^2 \geq C/n$. For other fixed k, the same order lower bound holds. Applying (S17),

$$
d_W(T_2, Y) \leq O(n^{-1/2}) \to 0.
\tag{S34}
$$

$\square$

6.4 Power analysis

First we prove a lemma upper bounding the variances of T_1 and T_2.

Lemma 6.5. • $\sigma_{1,n}^2 = O(n)$ for fixed k; $\sigma_{1,n}^2 = O(n/k!)$ for $k \to \infty$ and $k/(\log n)^\alpha \to 0$.

• $\sigma_{2,n}^2 = O(n^{2k-1})$ for fixed k; $\sigma_{2,n}^2 = O(\mu_{2,n}^2 k^{5/2}/n)$ for $k \to \infty$ and $k/(\log n)^\alpha \to 0$.

Proof. By the upper bound in (S8),

$$
\begin{aligned}
\sigma_{1,n}^2 &= \sum_{i=1}^{n-k+1} \sum_{j \in N_i \setminus \{i\}} \mathbb{E}(\mathbb{I}_i \mathbb{I}_j) + \frac{2(n-k+1)}{k!} - \sum_{i=1}^{n-k+1} \sum_{j \in N_i} \mathbb{E}(\mathbb{I}_i)\mathbb{E}(\mathbb{I}_j) \\
&\leq \frac{C(n-k+1)}{(k+1)!} + \frac{2(n-k+1)}{k!} \\
&= \begin{cases} O(n) & \text{for fixed } k; \\ O(n/k!) & \text{for } k \to \infty, \ k/(\log n)^\alpha \to 0. \end{cases}
\end{aligned}
\tag{S35}
$$

15

To bound $\sigma_{2,n}^2$, first note that $B(k) \le A(k-1,1)$ for all $k \ge 2$. This holds because for every pair of (l_0, l_1) and (m_0, m_1) such that $l_0 + l_1 = k - 1$ and $m_0 + m_1 = k - 1$, we have

$$\binom{l_0 + m_0}{l_0}^2 \binom{l_1 + m_1}{l_1}^2 + \binom{l_0 + m_1}{l_0}^2 \binom{l_1 + m_0}{l_1}^2 \ge 2 \binom{l_0 + m_0}{l_0} \binom{l_1 + m_1}{l_1} \binom{l_0 + m_1}{l_0} \binom{l_1 + m_0}{l_1}.$$

By equations (S24), (S25) and (S28),

$$
\begin{aligned}
\sigma_{2,n}^2 &= 2 \sum_{j=1}^{k} \binom{n}{2k-j} \frac{1}{(2k-j)!} A(k-j, j) + 2 \binom{n}{2k-1} \frac{1}{(2k-1)!} B(k) - \frac{4D\binom{n}{k}}{(k!)^2} \\
&\le 4 \sum_{j=1}^{k} \binom{n}{2k-j} \frac{1}{(2k-j)!} A(k-j, j) - \frac{4D\binom{n}{k}}{(k!)^2} \\
&= O\left(\frac{\mu_{2,n}^2 k^{5/2}}{n} \right)
\end{aligned}
\tag{S36}
$$

by Theorem 1 in [14]. The first part of the lemma holds since $\mu_{2,n} = O(n^k)$ for k fixed. $\qquad\square$

Proof of Theorem 3. It is easy to see the count W_1 is bounded below by $m - k + 1$. By the first part of Lemma 6.5,

$$
\begin{aligned}
T_1 &\ge \frac{m - k + 1 - \mu_{1,n}}{\sigma_{1,n}} \\
&\ge C\sqrt{n} \left(\frac{m}{n} - \frac{2}{k!} \right),
\end{aligned}
\tag{S37}
$$

for some $C > 0$, fixed k and m, n sufficiently large. In this case, m has to grow at the same rate as n, that is $m \sim a_1 n$ and $a_1 > 2/k!$. It follows then $T_1 = \Omega(\sqrt{n})$.

When $k \to \infty$ and $k/(\log n)^\alpha \to 0$, for n large enough,

$$
\begin{aligned}
T_1 &\ge C \sqrt{\frac{n}{k!}} \left(\frac{k!(m - k + 1)}{n - k + 1} - 2 \right) \\
&\ge C \sqrt{\frac{n}{k!}} \left(\frac{a_2 n - k! \cdot k + k!}{n - k + 1} - 2 \right) \\
&= \Omega\left(\sqrt{\frac{n}{k!}} \right)
\end{aligned}
\tag{S38}
$$

for $m \ge a_2 n / k!$, $a_2 > 2$. If m grows at the rate of $a_3 n$, $a_3 \in (0, 1]$,

$$
\begin{aligned}
T_1 &\ge C\sqrt{nk!} \left(\frac{m - k + 1}{n - k + 1} - \frac{2}{k!} \right) \\
&\ge C\sqrt{nk!} a_3 = \Omega(\sqrt{nk!}).
\end{aligned}
\tag{S39}
$$

Similarly, the count W_2 is lower bounded by $\binom{m}{k}$, using the second part of Lemma 6.5, for fixed k,

$$
\begin{aligned}
T_2 &\ge C \cdot \frac{\binom{m}{k} - 2\binom{n}{k}/k!}{n^{k-1/2}} \\
&= C\sqrt{n} \left(\frac{m \cdots (m - k + 1)}{n \cdots (n - k + 1)} - \frac{2}{k!} \right) \\
&\ge C\sqrt{n} \left(\left(\frac{m}{n} \right)^k - \frac{2}{k!} \right)
\end{aligned}
\tag{S40}
$$

16

for sufficiently large m and n. m has to grow at the rate of $b_1 n$ for the lower bound to go to infinity, and $b_1^k > 2/k!$. We have $T_2 = \Omega(\sqrt{n})$.

When $k \to \infty$ and $k/(\log n)^\alpha \to 0$, again by Lemma 6.5,

$$
\begin{aligned}
T_2 &\geq C\sqrt{\frac{n}{k^{5/2}}}\left(k!\frac{m\cdots(m-k+1)}{n\cdots(n-k+1)} - 2\right) \\
&\geq C\sqrt{\frac{n}{k^{5/2}}}\left(k!\left(\frac{m}{n}\right)^k - 2\right) \\
&\geq C\sqrt{\frac{n}{k^{3/2}}},
\end{aligned}
\tag{S41}
$$

for $m \geq en/k$. When $m \sim b_2 n$, $b_2 \in (0,1]$,

$$
\begin{aligned}
T_2 &\geq C\sqrt{\frac{n}{k^{5/2}}}\left(k!\left(\frac{m}{n}\right)^k - 2\right) \\
&\geq Cb_2^k k!\sqrt{\frac{n}{k^{5/2}}}.
\end{aligned}
\tag{S42}
$$

$\square$

Proof of Theorem 4. Let $n_1, \ldots, n_d$ denote the number of points in $(\mathbf{x}, \mathbf{y})$ falling on to each monotonic piece, then W_2 is lower bounded by $\sum_{t=1}^d \binom{n_t}{k}$. For fixed d and k, ,

$$
\frac{\sum_{t=1}^d n_t \cdots (n_t - k + 1)}{n \cdots (n - k + 1)} \xrightarrow{P} \sum_{t=1}^d \ell_t^k
\tag{S43}
$$

Since by Lemma 6.5,

$$
T_2 \geq C\sqrt{n}\left(\frac{\sum_{t=1}^d n_t \cdots (n_t - k + 1)}{n \cdots (n - k + 1)} - \frac{2}{k!}\right)
\tag{S44}
$$

for some $C > 0$, it follows

$$
\mathbb{P}\left(T_2 \geq C\sqrt{n}(d^{-(k-1)} - 2/k!)\right) \to 1
\tag{S45}
$$

using Hölder's inequality and the fact $\sum_{t=1}^d \ell_t = 1$. Thus T_2 is lower bounded by $C\sqrt{n}$ with probability tending to 1 when $d^{k-1} < k!/2$.

When $k \to \infty$ and $k/(\log n)^\alpha \to 0$, it is easy to check

$$
\frac{n_t \cdots (n_t - k + 1)}{n \cdots (n - k + 1)} \cdot \left(\frac{n}{n_t}\right)^k \xrightarrow{P} 1.
\tag{S46}
$$

Also,

$$
\begin{aligned}
&\mathbb{P}\left(\left|\left(\frac{n_t}{n\ell_t}\right)^k - 1\right| \geq \epsilon\right) \\
&\leq \mathbb{P}\left(\frac{n_t}{n\ell_t} - 1 \geq (1+\epsilon)^{1/k} - 1\right) + \mathbb{P}\left(\frac{n_t}{n\ell_t} - 1 \leq (1-\epsilon)^{1/k} - 1\right) \\
&\leq \exp(-2n\ell_t^2((1+\epsilon)^{1/k} - 1)^2) + \exp(-2n\ell_t^2((1-\epsilon)^{1/k} - 1)^2) \to 0
\end{aligned}
\tag{S47}
$$

17

by Hoeffding's inequality. It follows then

$$\sum_{t=1}^{d} \frac{n_t \cdots (n_t - k + 1)}{n \cdots (n - k + 1)} \cdot \left(\sum_{t=1}^{d} \ell_t^k \right)^{-1} \xrightarrow{P} 1. \tag{S48}$$

Now noting that

$$T_2 \geq C \sqrt{\frac{n}{k^{5/2}}} \left(k! \frac{\sum_{t=1}^{d} n_t \cdots (n_t - k + 1)}{n \cdots (n - k + 1)} - 2 \right), \tag{S49}$$

we have

$$\mathbb{P} \left(T_2 \geq C \frac{k!}{d^{k-1}} \sqrt{\frac{n}{k^{5/2}}} \right) \to 1 \tag{S50}$$

again by Hölder's inequality. □

References

[1] Frank E Harrell. Hmisc: Harrell miscellaneous. `http://cran.r-project.org/web/packages/Hmisc/index.html`, 2014.

[2] Maria L Rizzo and Gabor J Szekely. energy: E-statistics (energy statistics). `http://cran.r-project.org/web/packages/energy/index.html`, 2014.

[3] Phil Spector, Jerome Friedman, Robert Tibshirani, and Thomas Lumley. acepack: ace() and avas() for selecting regression transformations. `http://cran.r-project.org/web/packages/acepack/index.html`, 2013.

[4] Gabriele Sales and Chiara Romualdi. parmigene – a parallel R package for mutual information estimation and gene network reconstruction. *Bioinformatics*, 27(13):1876–1877, 2011.

[5] Davide Albanese, Michele Filosi, Roberto Visintainer, Samantha Riccadonna, Giuseppe Jurman, and Cesare Furlanello. minerva and minepy: a C engine for the MINE suite and its R, Python and MATLAB wrappers. *Bioinformatics*, 29(3):407–408, 2013.

[6] Leo Breiman and Jerome H Friedman. Estimating optimal transformations for multiple regression and correlation. *J Am Stat Assoc*, 80(391):580–598, 1985.

[7] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Phys Rev E*, 69(6):066138, 2004.

[8] Jesús E García and V. A. González-López. Independence tests for continuous random variables based on the longest increasing subsequence. *Journal of Multivariate Analysis*, 127:126–146, 2014.

[9] Xuewen Pan, Ping Ye, Daniel S Yuan, Xiaoling Wang, Joel S Bader, and Jef D Boeke. A DNA integrity network in the yeast Saccharomyces cerevisiae. *Cell*, 124(5):1069–1081, 2006.

[10] Michael Costanzo, Anastasia Baryshnikova, Jeremy Bellay, Yungil Kim, Eric D Spear, Carolyn S Sevier, Huiming Ding, Judice LY Koh, Kiana Toufighi, and Sara Mostafavi. The genetic landscape of a cell. *Science*, 327(5964):425–431, 2010.

[11] Rafael A Irizarry, Bridget Hobbs, Francois Collin, Yasmin D Beazer-Barclay, Kristen J Antonellis, Uwe Scherf, and Terence P Speed. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, 4(2):249–264, 2003.

[12] Peter M Fenwick. A new data structure for cumulative frequency tables. *Software: Practice and Experience*, 24(3):327–336, 1994.

[13] Nathan Ross et al. Fundamentals of Stein's method. *Prob. Surv*, 8:210–293, 2011.

[14] Ross G Pinsky. Law of large numbers for increasing subsequences of random permutations. *Random Struct Algorithms*, 29(3):277–295, 2006.