Subject Section

Fundamental gene network rewiring at the second order within and across mammalian systems

Ruby Sharma [1], Sajal Kumar [1] and Mingzhou Song [1,2]*

[1]Department of Computer Science, [2]Molecular Biology and Interdisciplinary Life Science Graduate Program
New Mexico State University, Las Cruces, NM 88003, USA

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: Genetic or epigenetic events can rewire molecular networks to induce extraordinary phenotypical divergences. Among the many network rewiring approaches, no model-free statistical methods can differentiate gene-gene pattern changes not attributed to marginal changes. This may obscure fundamental rewiring from superficial changes.

Results: Here we introduce a model-free Sharma-Song test to determine if patterns differ in the second order, meaning that the deviation of the joint distribution from the product of marginal distributions is unequal across conditions. We prove an asymptotic chi-squared null distribution for the test statistic. Simulation studies demonstrate its advantage over alternative methods in detecting second-order differential patterns. Applying the test on three independent mammalian developmental transcriptome datasets, we report a lower frequency of co-expression network rewiring between human and mouse for the same tissue group than the frequency of rewiring between tissue groups within the same species. We also find second-order differential patterns between microRNA promoters and genes contrasting cerebellum and liver development in mice. These patterns are enriched in the spliceosome pathway regulating tissue specificity. Complementary to previous mammalian comparative studies mostly driven by first-order effects, our findings contribute an understanding of system-wide second-order gene network rewiring within and across mammalian systems. Second-order differential patterns constitute evidence for fundamentally rewired biological circuitry due to evolution, environment, or disease.

Availability: The generic Sharma-Song test is available from the R package 'DiffXTables' at https://cran.r-project.org/package=DiffXTables. Other code and data are described in Methods.

Contact: joemsong@cs.nmsu.edu

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

1 Introduction

Molecular network rewiring in cells can induce an extraordinary phenotypical divergence such as pluripotency and differentiation (Boland et al., 2014; Thiagarajan et al., 2014). Rewiring arises from either genetic variations (Van Roey et al., 2013; Lundby et al., 2019) or epigenetic reprogramming (Reik, 2007; Watanabe et al., 2013; Klinakis et al., 2020). Rewired distal enhancer-promoter interactions within conserved topologically associating domains (TADs) are the basis of tissue specificity (Dixon et al., 2012, 2015; Phillips-Cremins et al., 2013; Smith et al., 2016). In diseases, TAD boundaries are disrupted to rewire enhancer-gene interactions (Lupianez et al., 2015). Such biological importances of rewiring motivated many computational methods to infer differential molecular networks from high-throughput omics data. However, we see a lack of statistical foundation to account for observed differences across systems as caused by either changed upstream input or a truly rewired connection—the latter may drown in an ocean of superficial differences, potentially missing a critical biological event.

Here, we define a conceptual framework that distinguishes types of change across conditions. We define differential/conserved relationships between two random variables by the presence/absence of differences in joint distribution across conditions. A relationship is *conserved* if the

pair of random variables share a joint distribution across all conditions; otherwise they are *differential*. If relationships under some conditions differ in the marginal distribution of either variable, they are *first-order differential*. If some relationships differ in joint distribution not attributed to differences in marginal distribution, they are *second-order differential*. Relationships are *full-order* differential if they are both 1st- and 2nd-order differential. In the context of a biological system, we argue that 2nd-order differences be intrinsic in network rewiring. This is because a 1st-order difference does not require a changed biological circuitry; it can be due to different downstream responses by two systems identical in circuitry but with changed input. Transcriptome differences of the same tissue type between human and mouse are reported to be much greater than differences between distinct tissue types within the same species (Lin et al., 2014); this finding provoked a debate on whether it is caused by artifacts of unwanted confounding effects (Gilad and Mizrahi-Man, 2015). However, the reported transcriptome differences can be dominated by 1st-order effects; the extent of mammalian molecular network rewiring at the 2nd-order is not answered.

Differential correlation underlies the majority of methods for network rewiring analysis (Hu et al., 2009; de la Fuente, 2010; Mentzen et al., 2009; Leonardson et al., 2010; McKenzie et al., 2016; Jardim et al., 2019). It can detect 2nd-order differences but only when variables in a relationship are linearly related. Pairwise differential correlation is generalized to differences in parameters in regression (Ouyang et al., 2011; Ha et al., 2015; Kim et al., 2018; He et al., 2019; Xu et al., 2019), but again with a strong assumption on the mathematical form of relationships. Tests of differential patterns *model-free*—without assuming parametric mathematical models for relationships—are scarce. A chi-squared heterogeneity test for goodness of fit is available from statistics textbooks (Zar, 2009). It is not as powerful as a comparative chi-squared test (Song et al., 2014) that was applied to detect rewiring in fruit fly gene networks and yeast metabolic pathways (Zhang et al., 2015). A recent model-free chi-squared test named marginal change test determines the first-order change in a pair of random variables across conditions (Sharma et al., 2020). However, there is no method known to us to exclusively detect 2nd-order differences in relationships.

To detect 2nd-order differential patterns model-free, we present the Sharma-Song test. It uses a contingency table to represent frequencies of discrete row and column random variables. Such a table is the sample joint distribution of the pair, enabling us to infer whether population distributions are 2nd-order differential without a parametric model. We quantify the difference in the extent to which the joint distributions of each pattern deviate from the product of their respective marginal distributions. We show that the test statistic is asymptotically chi-squared distributed under the null hypothesis. Our simulation studies demonstrate marked advantage of the test over alternatives in capturing 2nd-order differential patterns. We also inspected the statistical power of the test via simulation.

To understand the extent of 2nd-order co-expression network rewiring across tissue types within and between human and mouse, we apply the Sharma-Song test on three independent mammalian development transcriptome datasets: the uni-source FANTOM5 CAGE collection (FANTOM Consortium et al., 2014; FANTOM Consortium, 2017), the uni-source Evo-devo RNA-seq collection (Cardoso-Moreira et al., 2019), and a multi-source RNA-seq collection (Yang et al., 2017). All datasets captured gene expression of all major tissue and cell types in both human and mouse. Uni-source refers to data sequenced by the same organization, subject to fewer technical variations than data from multiple organizations (multi-source). Firstly, we compare network rewiring of the same tissue groups between human and mouse. Secondly, we contrast different tissue groups within human and also within mouse. The two comparisons reveal higher frequencies of network rewiring across tissue groups than across species on all three datasets. We also examined network rewiring related to microRNA taking place in the development of cerebellum versus liver in mice on the FANTOM5 data. Second-order differential patterns between microRNA promoters and genes in the spliceosome pathway are highly enriched between cerebellum and liver. Complementary to previous work mostly driven by 1st-order effects, our findings contribute an understanding of system-wide 2nd-order gene network rewiring within and across mammalian systems.

Table 1. **Statistical methods for testing three types of differential pattern.**

Model-free	Differential	First-order differential	Second-order differential
Yes	Comparative chi-squared test Heterogeneity test	Marginal change test Pearson chi-squared test Kolmogorov-Smirnov test	*Sharma-Song test*
No	Parametric regression	Generalized linear models	Differential correlation

2 Methods

2.1 Overview of the Sharma-Song test of second-order change in joint distribution

The Sharma-Song test evaluates statistical evidence for 2nd-order differential patterns across contingency tables. The null hypothesis is that row and column variables are independent of each other in all tables and tables are independently observed. It takes as input K contingency tables $\mathbf{C}_1, \ldots, \mathbf{C}_K$ of the same dimensions $r \times s$. The test outputs test statistic D^2, degrees of freedom ν, and P-value, based on an asymptotically chi-squared null distribution. The P-value is the statistical significance of 2nd-order differential patterns across the K tables. There are four major steps in the test as outlined in Figure 1a. In Step 1, 1st-order effects represented by expected $\bar{\mathbf{C}}_k$ are subtracted from each table $\mathbf{C}_k$ and the remaining counts are normalized to $\mathbf{A}_k$ by squared roots of expected counts in $\bar{\mathbf{C}}_k$. In Step 2, we use Helmert transform to convert matrix $\mathbf{A}_k$ to vector $\mathbf{e}_k$ of dimension $(r-1)(s-1)$. Components in $\mathbf{e}_k$ are i.i.d. standard normal variables under the null hypothesis. In Step 3, vectors Δ_k, deviation from each vector $\mathbf{e}_k$ to the pooled, form columns in matrix $\mathbf{Q}$. The rows of $\mathbf{Q}$ are projected to $\mathbf{S}_+$, the column space of common null covariance matrix of row vectors of $\mathbf{Q}$. In Step 4, the squared Mahalanobis distances from the projected vectors to the origin are summed to give the test statistic D^2, chi-squared distributed with degrees of freedom ν under the null hypothesis. Figure 1b illustrates the test on two genes *NOTCH3* and *PDGFA* from the Evo-devo RNA-seq data (Cardoso-Moreira et al., 2019). *NOTCH3*, a neurogenic locus notch homolog protein 3, regulates the expression of *PDGFRB* (Jin et al., 2008), which in turn interacts with heterodimers formed by PDGFA and PDGFB (Stelzer et al., 2008). Additionally, the *NOTCH3–PGDFA* interaction is predicted (McDowall et al., 2009). Contrasting the discretized co-expression patterns of the pair between human ectoderm and primitive streak derived tissue groups, the *NOTCH3–PDGFA* gene pair is declared statistically significantly 2nd-order differential ($P = 1.58 \times 10^{-31}$) between the two tissue groups.

The relationship of the Sharma-Song test with other methods for detecting differential patterns is summarized in Table 1. Network rewiring methods based on parametric models are widely available but not model-free, carrying potential strong biases for new systems where parametric model assumptions may be violated. Choices of model-free approaches are relatively limited but they are less subject to model biases.

2.2 The Sharma-Song test

The Sharma-Song test examines K contingency tables for second-order differences. Let X and Y be two discrete random variables of r and s levels, respectively. For table $k \in \{1, \ldots, K\}$, let $p_k(X, Y)$ be the

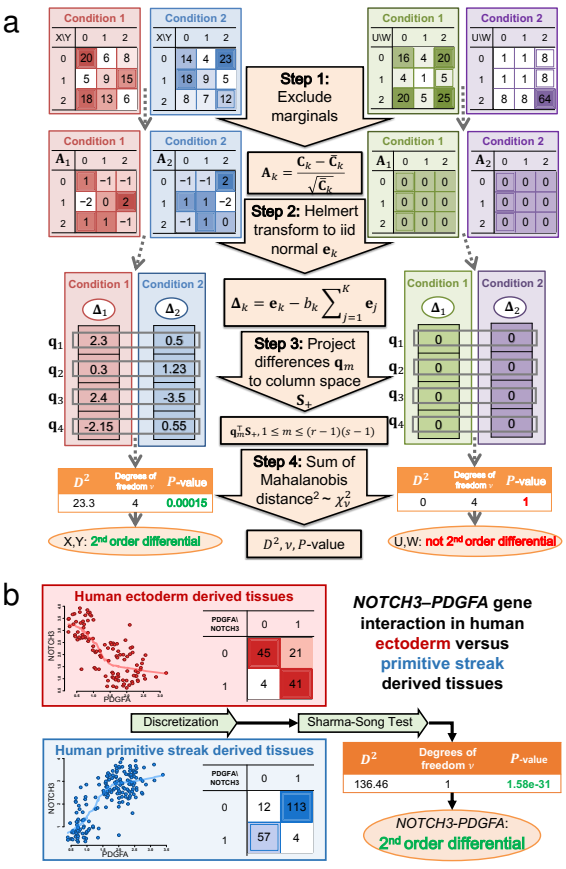

Fig. 1. Overview and examples of the Sharma-Song test to detect 2nd-order differential patterns. (a) The test has four steps, illustrated on two pairs of 3×3 contingency tables. At a significance level of 0.05, the first pair (X, Y) is declared 2nd-order differential but not the second pair (U, W). **(b)** Applying the Sharma-Song test on the genes NOTCH3 and PDGFA using Evo-devo data. Normalized (left) and discretized (right) co-expression patterns of the pair in human ectoderm (red box) and primitive streak (blue box) tissue groups are shown. At a significance level of 0.05, the NOTCH3–PDGFA gene pair is declared 2nd-order differential between the two tissue groups.

joint probability mass function of X and Y; let $p_k(X)$ and $p_k(Y)$ be the marginal distributions of X and Y. The K patterns are *conserved* if $p_1(X,Y) = \cdots = p_K(X,Y)$; otherwise they are *differential*. The K patterns are *first-order conserved* if $p_1(X) = \cdots = p_K(X)$ and $p_1(Y) = \cdots = p_K(Y)$; otherwise they are *first-order differential*. The K patterns are *second-order conserved* if $p_1(X,Y) - p_1(X)p_1(Y) = \cdots = p_K(X,Y) - p_K(X)p_K(Y)$; otherwise, they are *second-order differential*: some pattern deviates from the product of marginal distributions to an extent different from others. Patterns are *full-order differential* if they are both first- and second-order differential.

Next, we define the Helmert transform. Given an $r \times s$ contingency table $\mathbf{C}=[n_{ij}]$ ($n_{ij} \geq 0$), we define its expected matrix $\bar{\mathbf{C}} = [\bar{n}_{ij}]$ by

$$\bar{n}_{ij} = \frac{\sum_{l=1}^{s} n_{il} \sum_{l=1}^{r} n_{lj}}{n}, \text{ where } n = \sum_{i=1}^{r}\sum_{j=1}^{s} n_{ij} \tag{1}$$

where n is the total counts in $\mathbf{C}$. Normalizing $\mathbf{C}$, we get matrix $\mathbf{A} = [a_{ij}]$:

$$a_{ij} = (n_{ij} - \bar{n}_{ij})/\sqrt{\bar{n}_{ij}} \tag{2}$$

Let matrix $[p_{ij}]$ be cell probabilities of a contingency table whose counts follow a multinomial distribution. Let row marginal probability $p_{i\cdot}$ be the sum of row i of $[p_{ij}]$, and column marginal probability $p_{\cdot j}$ the sum of column j of $[p_{ij}]$. The $r \times r$ row-Helmert matrix $\mathbf{V} = [v_{ij}]$ is defined by

$$v_{ij} = \begin{cases} \sqrt{p_{i\cdot}} & i = 1 \\ 0 & j > i, i \neq 1 \\ -\sqrt{\dfrac{p_{1\cdot} + \cdots + p_{i-1\cdot}}{p_{1\cdot} + \cdots + p_{i\cdot}}} & j = i, i \neq 1 \\ \sqrt{\dfrac{p_{i\cdot}p_{j\cdot}}{(p_{1\cdot} + \cdots + p_{i-1\cdot})(p_{1\cdot} + \cdots + p_{i\cdot})}} & j < i \end{cases} \tag{3}$$

and the $s \times s$ column-Helmert matrix $\mathbf{W} = [w_{ij}]$ is defined by replacing $p_{1\cdot}, \ldots, p_{r\cdot}$ by $p_{\cdot 1}, \ldots, p_{\cdot s}$ in Eq. (3). The row- and column-Helmert transform of $\mathbf{A}$ is given by $\mathbf{E} = \mathbf{VAW}^\top$. If the row and column variables in $\mathbf{C}$ are independent, elements in matrix $\mathbf{E}$ have a statistical property as given in Lemma 1 previously established (Lancaster, 1949; Irwin, 1949).

Lemma 1. *A contingency table whose row and column variables are statistically independent can be asymptotically partitioned into a matrix of independent standard normal random variables, by orthogonal transformation of table using an $r \times r$ row-Helmert matrix and an $s \times s$ column-Helmert matrix.*

We use sample marginal probabilities to set $p_{i\cdot} = \sum_{j=1}^{s} n_{ij}/n$ and $p_{\cdot j} = \sum_{i=1}^{r} n_{ij}/n$, resulting in the 1st row and 1st column of $\mathbf{E}$ containing all zeros (Lancaster, 1949).

Now, we set up the Sharma-Song test on $r \times s$ contingency tables $\mathbf{C}_1, \ldots, \mathbf{C}_K$ of sample size $n_1, \ldots, n_K$, respectively. We calculate their Helmert transformed matrices $\mathbf{E}_1, \ldots, \mathbf{E}_K$. Let $\mathbf{e}_k$ be the column-major vector representation of matrix $\mathbf{E}_k$ after removing its 1st row and 1st column. $\mathbf{e}_k$ thus has dimension $M=(r-1)(s-1)$. Let $\mathbf{e}=\sum_{k=1}^{K} \mathbf{e}_k$ be the pooled vector. Deviation vectors $\mathbf{\Delta}_k$ from $\mathbf{e}_k$ to scaled $\mathbf{e}$ are

$$\mathbf{\Delta}_k = \mathbf{e}_k - b_k\mathbf{e} \quad \text{with} \quad b_k = \frac{\sqrt{n_k}}{\sum_{l=1}^{K}\sqrt{n_l}}, \quad k = 1, \ldots, K \tag{4}$$

where the scaling vector $\mathbf{b} = (b_1, \ldots, b_K)^\top$ is chosen such that the deviation vectors are all zero if all input contingency tables differ by a linear multiplicative factor. A high magnitude of $\mathbf{\Delta}_k$ indicates strong heterogeneity between contingency table $\mathbf{C}_k$ and other tables. With $\mathbf{\Delta}_1$, …, $\mathbf{\Delta}_K$ as columns, we form an $M \times K$ matrix $\mathbf{Q} = [\mathbf{\Delta}_1, \ldots, \mathbf{\Delta}_K]$. We define the row vectors of $\mathbf{Q}$ by $\mathbf{q}_m = (\Delta_{1m}, \ldots, \Delta_{Km})^\top$, $m = 1, \ldots, M$, where Δ_{km} is the m-th component in vector $\mathbf{\Delta}_k$. Under the null hypothesis, the covariance matrices of $\mathbf{q}_m$, are all equal to

$$\mathbf{\Sigma_q} = \mathbf{I} - \mathbf{J}_K\text{diag}(\mathbf{b}) - \text{diag}(\mathbf{b})\mathbf{J}_K + K\mathbf{bb}^\top \tag{5}$$

where $\mathbf{I}$ is the identity matrix and $\mathbf{J}_K$ a matrix of all ones, both $K \times K$.

Lemma 2. *The rank of covariance matrix $\mathbf{\Sigma_q}$ is exactly $K - 1$.*

Lemma 2 is proven in Supplement A. As the inverse of rank-deficient matrix $\mathbf{\Sigma_q}$ does not exist, the Mahalanobis distance from each $\mathbf{q}_m$ to the origin is undefined. By eigenvalue decomposition of covariance matrix $\mathbf{\Sigma_q}$, we have $\mathbf{\Sigma_q} = \mathbf{S\Lambda S}^{-1}$, where the columns of $\mathbf{S}$ are the eigenvectors of $\mathbf{\Sigma_q}$ and the diagonal matrix $\mathbf{\Lambda}$ contains the corresponding eigenvalues of $\mathbf{\Sigma_q}$. Keeping only the $K-1$ non-zero eigenvalues and their corresponding eigenvectors, we obtain $\mathbf{\Lambda}_+$ and $\mathbf{S}_+$. Projecting $\mathbf{q}_m$ to $\mathbf{S}_+$, the column space of $\mathbf{\Sigma_q}$, and summing up the squared Mahalanobis distances from each projected vector to the origin, we obtain the Sharma-Song test statistic

$$D^2 = \sum_{m=1}^{(r-1)(s-1)} \left\| \mathbf{q}_m^\top \mathbf{S}_+ \mathbf{\Lambda}_+^{-1/2} \right\|^2 \tag{6}$$

Theorem 1. *Under the null hypothesis of row and column variables being independent in K independent contingency tables of dimension $r \times s$,*

the Sharma-Song test statistic D^2 asymptotically follows a chi-squared distribution χ^2_ν with $\nu = (K-1)(r-1)(s-1)$ degrees of freedom.

Theorem 1 is proven in Supplement B. With the chi-squared null distribution, we can compute a P-value using the upper tail probability of an observed statistic as the significance of 2nd-order difference.

We measure the effect size by normalizing the test statistic by the table size and the total sample size:

$$\varepsilon = \sqrt{D^2 \Big/ \left[(r-1)(s-1)\sum_{k=1}^{K} n_k\right]} \tag{7}$$

Algorithm S1 Sharma-Song-Test in Supplement C computes the test statistic, degrees of freedom, effect size, and statistical significance. It takes as input K contingency tables $\mathbf{C}_1, \ldots, \mathbf{C}_K$ of dimension $r \times s$. Its time complexity is evidently $O(K(r^2 s + rs^2) + K^3)$, constant in the sample size of input tables, favorable for data of large sample sizes.

2.3 The second-order network rewiring pipeline

We introduce a pipeline for second-order network rewiring (Supplementary D). It first finds the median absolute deviation (MAD) of each cognate gene to remove hardly changed genes among the bottom 5% MAD. It then divides the datasets into four species-tissue groups based on developmental origin: human ectoderm, human primitive streak, mouse ectoderm, and mouse primitive streak. Genes with zero variance within each group are removed from further analysis. It runs four comparisons: human ectoderm versus primitive streak, mouse ectoderm versus primitive streak, human versus mouse ectoderm and human versus mouse primitive streak. Each gene expression in each condition is discretized using an optimal univariate clustering algorithm (Wang and Song, 2011; Song and Zhong, 2020) that determines discretization levels by the Bayesian information criterion (Kass and Wasserman, 1995). Each co-expression pattern forms a contingency table. It then builds model-free co-expression networks of each comparison based on Pearson's chi-squared test for each condition (Song et al., 2009). It reports Benjamini and Hochberg (1995) (BH) adjusted P-values and Cramér's V (Cramér, 1999) effect size for each co-expression pattern among all possible gene pairs.

To detect 2nd-order differential patterns with strong dynamics, significant co-expression with BH adjusted P<0.1 and Cramér's V>0.8 in at least one condition is required for selecting a gene pair. All gene expression levels are standardized to zero mean and unit variance in each condition. Continuous expression values of each gene are discretized (Song et al., 2020) on merged samples for that gene. The Sharma-Song test evaluates co-expression patterns as contingency tables for 2nd-order difference across conditions. BH adjusted P-values and Sharma-Song effect sizes are collected for each co-expression pattern. To offer a sufficient strength of differentiality, we further require both a minimal Sharma-Song effect size ε and P-value for each study. To determine the cutoff, we generated an alternative population by simulating 100,000 discrete differential co-expression patterns in two experimental conditions with 100 samples in each. Differential co-expression patterns were simulated using `simulate_tables` (Sharma et al., 2017) in the 'FunChisq' R package (Zhang et al., 2020), where each of the two contingency tables were randomly selected to carry either a functional, a dependent but non functional or an independent relationship. Both the number of rows and columns for each table were independently and randomly selected to be between two and the maximum dimension size (row or column) observed in tables from the study. We applied the Sharma-Song test to determine ε_{60}, the cutoff at 60% of the effect size computed on the alternative population. We provide an R package 'DiffXCoExpNet' and R scripts to analyze second-order rewired networks on omics data in Supplement D. Code and data for omics application are available for download at https://www.cs.nmsu.edu/~joemsong/rewiring

2.4 Preparation of uni-source FANTOM5 CAGE data

Sample selection: The ectoderm-derived tissue group includes cerebellum, diencephalon, hippocampus, medulla oblongata, pituitary gland, skin and spinal cord. The primitive-streak-derived tissue group includes aorta, colon, epididymis, heart, kidney, liver, lung, ovary, pancreas, prostate, spleen, testis, thymus, tongue, and vagina. Adult samples are included. Developmental data, unmatched between human and mouse, are not used except two late neonatal time points N25 and N30 from mouse, as lab mice can reach sexual maturity as early as three weeks post birth (Snell, 1956). Considering only healthy samples, we identified 34 human and 37 mouse samples. *Transcription start sites selection:* The first promoter region (p1) in FANTOM5 usually represents the primary transcript most abundantly expressed. Thus, we only used TSSs labeled as p1 in our analysis. We found p1 of 13,574 cognate gene names between human and mouse. Gene expression in tags per million (TPM) was log transformed after adding one. By principal component analysis (PCA), we removed a sample of human skin tissue as outlier. After preprocessing, we selected 9,757 genes of rich dynamics for 2nd-order network rewiring analysis.

2.5 Preparation of uni-source Evo-devo RNA-seq data

The RNA-seq dataset by Cardoso-Moreira et al. (2019) was acquired from developing tissues of mammals from early organogenesis to adulthood. The time span of human tissues extended from week after conception, infants, juveniles, to adults, whereas mouse tissues were collected from embryo to post neonate developmental stages. Gene expression is already normalized to RPKM. *Sample selection:* The dataset consists of seven tissue types out of which cerebellum and brain are ectoderm derived and heart, kidney, liver, ovary and testis are primitive streak derived. Human and mouse data have 297 and 316 samples, respectively. *Gene selection:* Out of 43,207 human and 35,192 mouse transcripts, we selected 15,992 cognate genes between human and mouse. Gene expression was log transformed after the addition of one. After preprocessing, 13,307 genes of rich dynamics go to 2nd-order differential analysis.

2.6 Preparation of multi-source Yang et al. RNA-seq data

The RNA-seq dataset integrated by Yang et al. (2017) contains gene expression of differentiated tissues in unit of FPKM and is already normalized with batch effects removed. *Sample selection:* Out of 78 mouse and 184 human samples, we selected primary tissue types derived from either primitive-streak or ectoderm. The primitive streak group includes colon, heart, kidney, liver, lung, ovary, placenta, spleen, and testis; the ectoderm group includes brain, cerebellum, and cortex. *Gene selection:* Out of 22,495 mouse and 20,343 human protein coding genes, we selected 15,662 genes cognate between human and mouse. Gene expression was log transformed after the addition of one. After preprocessing, 13,910 genes of rich dynamics enter 2nd-order differential analysis.

3 Results

3.1 Second-order gene network rewiring between tissues opposed to between mammalian species

To learn the extent of gene network rewiring across tissue types versus across mammals, we performed four comparative studies between two tissue groups and between human and mouse on the FANTOM5 CAGE, Yang et al. RNA-seq, and Evo-devo RNA-seq collections. Table 2 and Figure 3 summarize the four comparative studies for all three collections.

Table 2. **Percentages of second-order rewired gene pairs among co-expressed gene pairs across tissue types and across mammalian species.**

Collection	Comparison	Co-expressed	2nd-order rewired (%)
	Across Tissue Type		
FANTOM5	Human: ectoderm (16) vs primitive streak (18)	521,416	112,794 **(22%)**
CAGE	Mouse: ectoderm (10) vs primitive streak (27)	3,507,316	494,585 **(14%)**
Uni-source	*Across Species*		
(n=71)	Ectoderm: Human (16) vs Mouse (10)	3,537,651	416,013 **(12%)**
	Primitive streak: Human (18) vs Mouse (27)	472,845	36,953 **(8%)**
	Across Tissue Type		
Evo-devo	Human: ectoderm (111) vs primitive streak (186)	917,171	362,684 **(40%)**
RNA-seq	Mouse: ectoderm (98) vs primitive streak (218)	4,913,298	1,305,751 **(27%)**
Uni-source	*Across Species*		
(n=613)	Ectoderm: Human (111) vs Mouse (98)	4,480,894	1,150,029 **(26%)**
	Primitive streak: Human (186) vs Mouse (218)	1,216,456	179,417 **(15%)**
	Across Tissue Type		
Yang et al	Human: ectoderm (21) vs primitive streak (52)	2,343,971	736,176 **(31%)**
RNA-seq	Mouse: ectoderm (14) vs primitive streak (37)	2,783,768	639,630 **(23%)**
Multi-source	*Across Species*		
(n=124)	Ectoderm: Human (21) vs Mouse (14)	3,764,211	799,602 **(21%)**
	Primitive streak: Human (52) vs Mouse (37)	1,267,793	195,243 **(15%)**

Co-expression of gene pairs are compared across species-tissue groups. Species are human and mouse. Tissue types are either ectoderm or primitive streak derived. Sample sizes for each group are given inside parentheses. Significantly co-expressed gene pairs are the union of such pairs from both groups being compared, with $P<0.1$ and Cramér's $V>0.8$. Significantly 2nd-order differential pairs with BH adjusted $P<0.05$ and $\varepsilon>\varepsilon_{60}$, obtained by the Sharma-Song test, are reported by total number and percentage among significant co-expression pairs. All P-values are BH adjusted for multiple testing.

Results from the FANTOM5 CAGE collection. In each comparison, all 9,757 × 9,757 co-expression patterns were evaluated to build a condition specific co-expression network including significant patterns from either species-tissue group. Then we used the Sharma-Song test with BH adjusted $P<0.05$ and $\varepsilon>0.456$ to report numbers of 2nd-order differential patterns (Figure 2). We found higher percentages of 2nd-order differences across tissue groups within human (22%) and also across tissue groups within mouse (14%), but lower percentages of 2nd-order differences across species within the primitive-streak tissue group (8%) and across species within the ectoderm tissue group (12%). Four 2nd-order differential gene pairs are shown for their rich nonlinear diversity in Figure 3a,d,g,j, indicating potentially rewired molecular mechanisms.

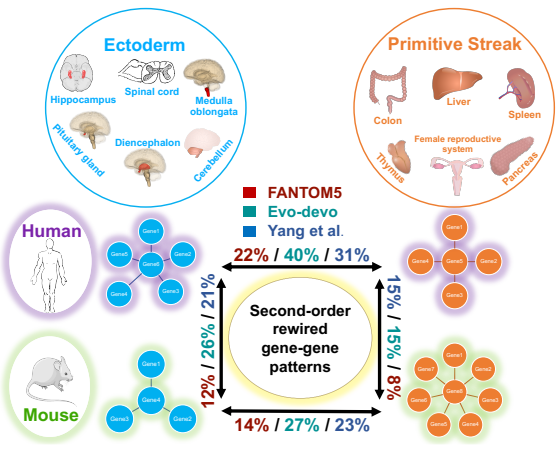

Fig. 2. Second-order gene network rewiring is more frequent between tissue groups than between human and mouse on FANTOM5, Evo-devo, and Yang et al. datasets. Tissue types are grouped by developmental origin of either ectoderm or primitive streak. Tissue group comparisons within specie (horizontal) reveal high percentages of 2nd-order differential patterns within both human (22%, 40%, 31%) and mouse (14%, 27%, 23%). In contrast, the same tissue group across species (vertical) shows low percentages of 2nd-order differential patterns for both ectoderm-derived tissues (12%, 26%, 21%) and primitive-streak-derived tissues (8%, 15%, 15%)

.

Results from the Evo-devo RNA-seq collection. In each comparison, all 13,307×13,307 co-expression patterns were evaluated to build a condition specific co-expression network including significant patterns from either species-tissue group. Table 2 reports 2nd-order differential pattern percentages by the Sharma-Song test with BH adjusted $P<0.05$ and $\varepsilon>0.251$. We also found higher percentages of 2nd-order differences across tissue groups within human (40%) and also across tissue groups within mouse (27%), but lower percentages of 2nd-order differences across species within the primitive-streak tissue group (15%) and across species within the ectoderm tissue group (26%). Four 2nd-order differential gene pairs are shown in Figure 3b,e,h,k.

Results from the Yang et al. RNA-seq collection. In each comparison, all 13,909×13,909 co-expression patterns were evaluated to build a condition specific co-expression network including significant patterns from either species-tissue group. Table 2 reports 2nd-order differential pattern percentages by the Sharma-Song test using BH adjusted $P<0.05$ and $\varepsilon>0.333$. We again found higher percentages of 2nd-order differences across tissue groups within human (31%) and also across tissue groups within mouse (23%), but lower percentages of 2nd-order differences across species within the primitive-streak tissue group (15%) and across species within the ectoderm tissue group (21%). Four 2nd-order differential gene pairs are shown in Figure 3c,f,i,l.

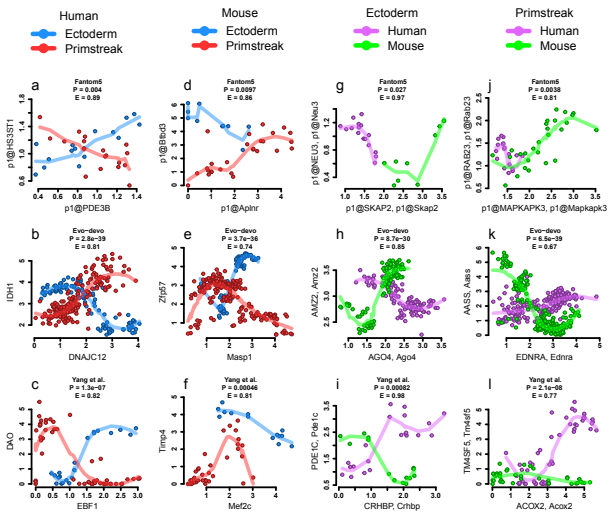

Fig. 3. Diverse second-order differential gene co-expression patterns between tissue groups and across species reported by the Sharma-Song test. Each plot represents a gene pair common to human and mouse. **(a,b,c)** Between human ectoderm- and primitive-streak-derived tissue groups. (a) FANTOM5: *p1@PDE3B–p1@HS3ST1*, (b) Evo-devo: *DNAJC12–IDH1*, (c) Yang et al.: *EBF1–DAO* **(d,e,f)** Between mouse ectoderm- and primitive-streak-derived tissue groups. (d) FANTOM5: *p1@Aplnr–p1@Btbd3*, (e) Evo-devo: *Masp1–Zfp57*, (f) Yang et al.: *Mef2c–Timp4*. **(g,h,i)** Between human and mouse ectoderm-derived tissue groups. (g) FANTOM5: *p1@Skap2, p1@SKAP2–p1@Neu3, p1@NEU3*. (h) Evo-devo: *Ago4, AGO4–Amz2, AMZ2*. (i) Yang et al.: *Crhbp, CRHBP–Pde1c, PDE1C*. **(j,k,l)** Between human and mouse primitive-streak-derived tissue groups. (j) FANTOM5: *p1@Mapkapk3, p1@MAPKAPK3–p1@Rab23, p1@RAB23* (k) Evo-devo: *Ednra, EDNRA–Aass, AASS*. (l) Yang et al.: *Acox2, ACOX2–Tm4sf5, TM4SF5*.

3.2 Reproducible second-order differential interactions across all three datasets

Despite differences in sample source, sample size, and sequencing technology, we observe some level of reproducibility of 2nd-order differential interactions across the three datasets. We found that the

strongest 20-percentile group of 2nd-order differential interactions contains the largest numbers of common high-degree hub genes between Evo-devo and Yang et al. collections in all four comparisons as shown in Supplement E. Figure 4 showcases four gene pairs with consistent patterns across all three collections. A complete list is provided in four tables in Supplement F. In Figure 4a, *CACNA1A*, a gene involved in the production of instructions for creating calcium channel for regulating communication between neurons (Auvin et al., 2009) is positively co-expressed with *P2RX4* in human ectoderm tissue group—the latter is a purinergic receptor highly expressed in central and peripheral neurons, whereas the pair is negatively co-expressed in the primitive streak tissue group. P2X4 endcoded by *P2RX4* mediates Ca^{2+} channels and neuropathic pain (Suurväli et al., 2017; Ozaki et al., 2016; Xu et al., 2016). Both *CACNA1A* and *P2RX4* belong to the calcium signaling pathway (hsa04020) (Kanehisa and Goto, 2000) which plays a crucial role during development. Defects in this pathway could lead to diseases like epilepsy and cardiac malformations (Paudel et al., 2018). In Figure 4b, gene *RELN* is critical for cerebral cortical development as it regulates neuronal migration (Chang et al., 2007), positively co-expressed in mouse ectoderm with *PIK3R3*, a phosphoinositide 3-kinase gene encoding the PI3K enzyme with function in neurotransmitter-regulated neuronal signaling (Gross and Bassell, 2014), whereas the pair is negatively co-expressed in the primitive streak tissue group. Both *RELN* and *PIK3R3* are involved in the PI3K-Akt signaling pathway (Kanehisa and Goto, 2000). The PI3K-Akt-MtoR signaling pathway regulates cellular processes like cell growth, proliferation, autophagy and apoptosis, important for neurodevelopment (Wang et al., 2017).

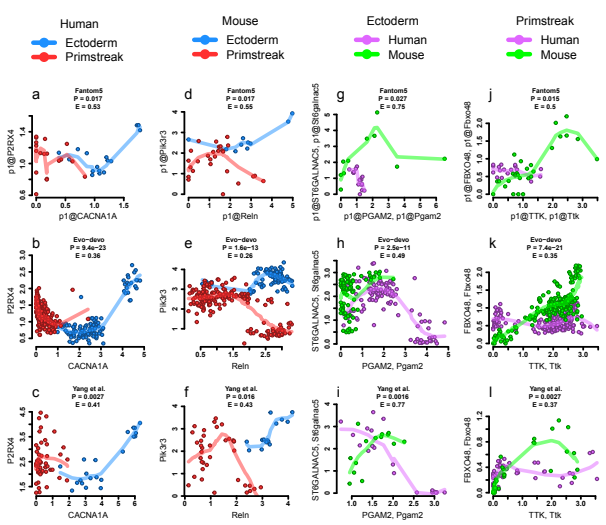

Fig. 4. Reproducible second-order differential patterns across all three datasets. Each plot represents a 2nd-order differential pattern between two genes common to human and mouse. **(a,b,c)** Reproducible 2nd-order differential patterns of *CACNA1A–P2RX4* between human ectoderm-derived tissue group and primitive-streak-derived tissue group. **(d,e,f)** Reproducible 2nd-order differential patterns of *Reln–Pik3r3* between mouse ectoderm-derived tissue group and primitive-streak-derived tissue group. **(g,h,i)** Reproducible 2nd-order differential patterns of *Pgam2–St6galnac5* from the three datasets between human and mouse ectoderm-derived tissue groups. **(j,k,l)** Reproducible 2nd-order differential patterns of *Ttk–Fbxo48* between human and mouse primitive-streak-derived tissue groups.

3.3 Pathways enriched with genes in second-order differential interactions across all three datasets

We found pathways that are highly enriched in genes involved in 2nd-order rewired gene interactions in all four species-tissue comparisons. From 2nd-order differential patterns of each comparison, we selected from all three datasets common genes to perform SIGORA pathway analysis (Foroushani et al., 2013). *Human ectoderm vs. primitive streak comparison:* We found 4,599 common genes among all the three datasets and 56 enriched pathways (Supplement G.1 Table S5) post Bonferroni correction (Dunnett, 1955). *Mouse ectoderm vs. primitive streak comparison*: We found 7,345 common genes enriching 89 pathways (Supplement G.2 Table S6). *Human vs. mouse ectoderm comparison*: We found 7,037 coinciding genes enriching 23 pathways (Supplement G.3 Table S7) that were common in both human and mouse. For these three comparisons, the most enriched include MAPK signaling (hsa04010), axon guidance (hsa04360), Wnt signaling (hsa04310) and Hippo signaling (hsa04390) pathways, all having a major role in development. MAPK signaling is involved in cell proliferation, differentiation and migration; it also activates the JNK pathway essential for embryonic development (Zhang and Liu, 2002; Maekawa et al., 2005). Axon guidance is related to neural development, the process by which an axon reaches its targets and leads to neural circuit (Stoeckli, 2018). Wnt signaling is crucial for stem cell proliferation and differentiation during embryo genesis and adult tissue homeostasis (Steinhart and Angers, 2018). Hippo signaling controls the organ size and also regulates tissue homeostasis during development (Pan, 2010). *Human vs. mouse primitive streak comparison*: We found 3,357 coinciding genes enriching six pathways (Supplement G.4 Table S8) common between human and mouse. The top pathways are axon guidance (hsa04360) and cell cycle (hsa04110). Cell cycles are closely coupled with cellular differentiation where developmental signals often determine a cell cycle mode specific to cell type (Jakoby and Schnittger, 2004), subject to tissue specific rewiring. For example, in some brain regions, the orientation of granule cell precursor division can determine the production of more granule cell precursors or granule cells (Miyashita et al., 2017).

3.4 Detecting rewired miRNA-gene patterns between developing mouse cerebellum and liver

To evaluate if the Sharma-Song test can return 2nd-order differential patterns that are biologically relevant, we scrutinized network rewiring between developing mouse cerebellum and liver. Both tissue types have vast literature to provide a basis for biological justification of our methodology development. They also have relatively large sample sizes, long time courses, and rich dynamic in the FANTOM5 mouse data. After preprocessing (Supplement H.1), we obtain 12,937 TSSs and 502 pri-miRNAs for 36 cerebellar samples and 15 liver samples.

Using the 2nd-order network rewiring pipeline, we evaluated $502 \times 12{,}937$ gene pairs to find 20,577 co-expressed patterns in cerebellum and 197,347 in liver. A union of 216,709 unique patterns were supplied to the Sharma-Song test that returned 42,352 significant (BH adjusted $P < 0.05$ and $\varepsilon > 0.456$) 2nd-order differential patterns, five of which are shown in Supplement H.1 Figure S2. *Tpx2* and *Ctnna1* are positively co-expressed in the cerebellum whereas negatively co-expressed in the liver. *mmu-miR-5622-5p* promoted by *Tpx2* targets *Ctnna1*. *Ctnna1* is important to the Hippo and cancer pathways (Vlachos et al., 2015) by inhibiting *Yap1*, which in high concentration disrupts the Hippo pathway and can result in cancer (Herr et al., 2014; Silvis et al., 2011). *Tpx2*, hosting *mmu-miR-5622-5p*, is over-expressed in liver cancer and promotes its growth (Hsu et al., 2017). We observe a strong negative relationship between *Ctnna1* and *Tpx2* in liver, lung, heart, and kidney, which can be interpreted as *Tpx2* or *mmu-miR-5622-5p* suppressing *Ctnna1*. In the cerebellum, however, the relationship is rewired so that *Ctnna1* is promoted by *Tpx2*,

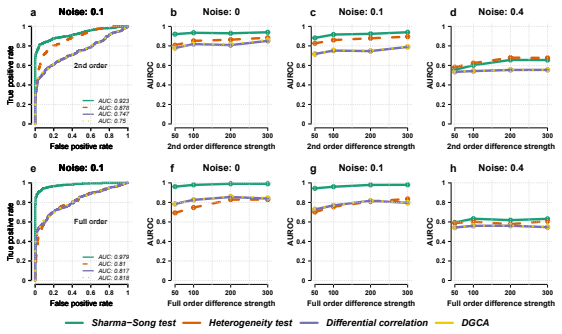

Fig. 5. Benchmarking the Sharma-Song test and three other methods. The top row is for detecting 2nd- from 1st-order differential patterns: **(a)** ROC curve at noise level 0.1. **(b, c, d)** AUROC as a function of 2nd-order difference at three noise levels (0, 0.1, 0.4). The bottom row is for detecting full- from 1st-order differential patterns: **(e)** ROC curve at noise level 0.1. **(f, g, h)** AUROC as a function of full-order difference at noise levels 0, 0.1, 0.4.

opposite to other tissue types. Using 5,888 unique TSSs from second-order differential interactions, we peformed SIGORA (Foroushani et al., 2013) analysis to obtain 12 enriched pathways (Supplement H.2) post Bonferroni correction (Dunnett, 1955). The top pathways of spliceosome (mmu03040), RNA-transport (mmu03013), and cell-cycle (mmu04110) are all ubiquitous in development (Supplement H.2). Applying our method on the topology of the spliceosome pathway (mmu03040), we created a 2nd-order rewired spliceosome network between developing cerebellum and liver (Supplement H.3).

3.5 Benchmarking the Sharma-Song test

We conducted simulation studies to verify the performance of the Sharma-Song test in identifying 2nd-order differential patterns in contrast to three other methods: differential Pearson's correlation (Hu et al., 2009; Leonardson et al., 2010; Mentzen et al., 2009), a Z-score on scaled differential Spearman's correlation from DGCA (McKenzie et al., 2016), and the chi-squared heterogeneity test (Zar, 2009). We measured their performance by receiver operating characteristic (ROC) curves.

We designed three algorithms to generate 500 1st-, 500 2nd-, and 500 full-order differential table tuples. The algorithms are implemented as options in the `simulate_diff_tables` function from the 'DiffXTables' R package (Sharma and Song, 2020). Each tuple contains two tables (K=2). Each table contains a random number of points from 100 to 300. We applied noise at levels of 0.0, 0.1, 0.2, 0.3, 0.4 and 0.5 to the tables. Figure 5 shows ROC curves with area under ROC (AUROC) for each method. In the first setup, each method detects 2nd-from 1st-order differential patterns. In the second setup, full-order patterns are detected from 1st-order patterns. With increasing differentiality across tables, the performance of all methods also increases as one would expect. At noise levels of 0.0 and 0.1, the Sharma-Song test is about 10–20% better than differential correlation and DGCA, and 5% better than the heterogeneity test. At noise level 0.4, all methods amount to random guessing. Supplement I benchmarked the methods for K=3 conditions with similar results, confirming the heterogeneity test being unspecific to 2nd-order, differential correlation not recognizing differences in complex nonlinear patterns, and the Sharma-Song test being most effective.

3.6 The statistical power of Sharma-Song test

We studied the statistical power of Sharma-Song test in relation to sample size, given false positive rate, effect size, and table size. The result is shown as Supplement J Figure S6. The false positive rate was fixed at 0.05. The 60%-power effect size was given as 0.235. Three table sizes of 3×3, 4×4 and 5×5 were evaluated.

At a smaller sample size, the test can achieve a greater statistical power on table pairs of smaller sizes than larger sizes. At a larger sample size, the relation is reversed. To attain a power level of about 60%, the required sample size for all table sizes is 12 for each table in the pair. Full details are given in Supplement J.

4 Discussion

Lin et al. (2014) reported that the difference in genetic landscape between distinct tissues types is underwhelming compared to that between species. Their conclusions are concerned with 1st-order differences in transcript abundance based on their own data and ENCODE Project Consortium (2004) data. The Sharma-Song test is designed to capture second-order differential patterns, offering an orthogonal perspective complementary to first-order analysis. The persistence of second-order effects neither proves nor disproves first-order effects. On three independent mammalian development transcriptome collections, we exclusively looked for 2nd-order differential patterns by comparing tissue types with different germline lineage within human and mouse and across the two species. Not in the same analogy to (Lin et al., 2014), we find on all three datasets that 2nd-order differences between tissues of distinct origins are higher than those from same tissue but between distinct species, hinting at tissue-specific circuitry being conserved across mammals.

We studied co-expression patterns that are 2nd-order rewired across tissue groups derived from different development origin but conserved within each group of heterogeneous tissue types from the same development origin. Grouping by origin was due to limited sample availability for each homogeneous tissue type in two of the three collections. We do expect 2nd-order differential co-expression patterns to arise between heart and liver tissues, for example, both developed from the primitive streak.

Shifted and scaled patterns can arise either from truly rewired dynamical systems, or from artifacts due to batch or library size effects. In our studies, we separately scaled and shifted the continuous data in each condition. Although we may have demoted shifted or scaled patterns, such pre-processing makes our analysis robust to unwanted confounding effects.

The Sharma-Song test is applicable to a network topology generated by other network inference tools. On a given network with data observed under multiple conditions, we can ask whether an edge carries any 2nd-order change across conditions. On a directed network, we can test whether a many-to-one interaction from parents to their child has rewired in the 2nd order. Here we can collapse all parent nodes into a single compound node and then apply the test on the table formed by the compound node and the child node.

As high-throughput biology experiments are generating data with increasing cellular, temporal, and spatial resolutions, we expect that the 2nd-order network rewiring instrument offered here will be useful in characterizing fundamental biological circuitry changes due to evolution, environment, or disease.

Acknowledgments

This work was partially supported by National Science Foundation grant 1661331 and USDA grant 2016-51181-25408.

References

Auvin, S., Holder-Espinasse, M., Lamblin, M.-D., and Andrieux, J. (2009). Array-CGH detection of a de novo 0.7-Mb deletion in 19p13.13 including CACNA1A associated with mental retardation and epilepsy with infantile spasms. *Epilepsia*, 50(11):2501–2503.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300.

Boland, M. J., Nazor, K. L., and Loring, J. F. (2014). Epigenetic regulation of pluripotency and differentiation. *Circ Res*, 115(2):311–324.

Cardoso-Moreira, M., Halbert, J., Valloton, D., Velten, B., Chen, C., Shao, Y., Liechti, A., Ascenção, K., Rummel, C., Ovchinnikova, S., Mazin, P. V., Xenarios, I., Harshman, K., Mort, M., Cooper, D. N., Sandi, C., Soares, M. J., Ferreira, P. G., Afonso, S., Carneiro, M., Turner, J. M. A., VandeBerg, J. L., Fallahshahroudi, A., Jensen, P., Behr, R., Lisgo, S., Lindsay, S., Khaitovich, P., Huber, W., Baker, J., Anders, S., Zhang, Y. E., and Kaessmann, H. (2019). Gene expression across mammalian organ development. *Nature*, 571(7766):505–509.

Chang, B. S., Duzcan, F., Kim, S., Cinbis, M., Aggarwal, A., Apse, K. A., Ozdel, O., Atmaca, M., Zencir, S., Bagci, H., et al. (2007). The role of reln in lissencephaly and neuropsychiatric disease. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 144(1):58–63.

Cramér, H. (1999). *Mathematical Methods of Statistics*. Princeton University Press, Princeton, New Jersey.

de la Fuente, A. (2010). From 'differential expression'to 'differential networking'–identification of dysfunctional regulatory networks in diseases. *Trends in Genetics*, 26(7):326–333.

Dixon, J. R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J. E., Lee, A. Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., Diao, Y., Liang, J., Zhao, H., Lobanenkov, V. V., Ecker, J. R., Thomson, J. A., and Ren, B. (2015). Chromatin architecture reorganization during stem cell differentiation. *Nature*, 518(7539):331–336.

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398):376–380.

Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *J Am Stat Assoc*, 50(272):1096–1121.

ENCODE Project Consortium (2004). The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science*, 306(5696):636–640.

FANTOM Consortium (2017). An integrated expression atlas of miRNAs and their promoters in human and mouse. *Nat Biotechnol*, 35(9):872–878.

FANTOM Consortium, the RIKEN PMI, and CLST, (DGT) (2014). A promoter-level mammalian expression atlas. *Nature*, 507(7493):462–470.

Foroushani, A. B., Brinkman, F. S., and Lynn, D. J. (2013). Pathway-GPS and SIGORA: identifying relevant pathways based on the over-representation of their gene-pair signatures. *PeerJ*, 1:e229.

Gilad, Y. and Mizrahi-Man, O. (2015). A reanalysis of mouse ENCODE comparative gene expression data [v1; ref status: indexed, http://f1000r.es/5ez]. *F1000Research*, 4:121.

Gross, C. and Bassell, G. J. (2014). Neuron-specific regulation of class I PI3K catalytic subunits and their dysfunction in brain disorders. *Front Mol Neurosci*, 7:12.

Ha, M. J., Baladandayuthapani, V., and Do, K.-A. (2015). DINGO: Differential network analysis in genomics. *Bioinform*, 31(21):3413–3420.

He, H., Cao, S., Zhang, J.-G., Shen, H., Wang, Y.-P., and Deng, H.-W. (2019). A statistical test for differential network analysis based on inference of Gaussian graphical model. *Sci Rep*, 9:10863.

Herr, K. J., Tsang, Y.-h. N., Ong, J. W. E., Li, Q., Yap, L. L., Yu, W., Yin, H., Bogorad, R. L., Dahlman, J. E., Chan, Y. G., Bay, B. H., Singaraja, R., Anderson, D. G., Koteliansky, V., Viasnoff, V., and Thiery, J. P. (2014). Loss of α-catenin elicits a cholestatic response and impairs liver regeneration. *Sci Rep*, 4:6835.

Hsu, C.-W., Chen, Y.-C., Su, H.-H., Huang, G.-J., Shu, C.-W., Wu, T. T.-L., and Pan, H.-W. (2017). Targeting TPX2 suppresses the tumorigenesis of hepatocellular carcinoma cells resulting in arrested mitotic phase progression and increased genomic instability. *J Cancer*, 8(8):1378–1394.

Hu, R., Qiu, X., Glazko, G., Klebanov, L., and Yakovlev, A. (2009). Detecting intergene correlation changes in microarray analysis: a new approach to gene selection. *BMC Bioinform*, 10(1):20.

Irwin, J. O. (1949). A note on the subdivision of χ^2 into components. *Biometrika*, 36(1/2):130–134.

Jakoby, M. and Schnittger, A. (2004). Cell cycle and differentiation. *Current Opinion in Plant Biology*, 7(6):661–669.

Jardim, V. C., Santos, S. d. S., Fujita, A., and Buckeridge, M. S. (2019). BioNetStat: A tool for biological networks differential analysis. *Front Genet*, 10:594.

Jin, S., Hansson, E. M., Tikka, S., Lanner, F., Sahlgren, C., Farnebo, F., Baumann, M., Kalimo, H., and Lendahl, U. (2008). Notch signaling regulates platelet-derived growth factor receptor-β expression in vascular smooth muscle cells. *Circ Res*, 102(12):1483–1491.

Kanehisa, M. and Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1):27–30.

Kass, R. E. and Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *J Am Stat Assoc*, 90(431):928–934.

Kim, Y., Hao, J., Gautam, Y., Mersha, T. B., and Kang, M. (2018). DiffGRN: differential gene regulatory network analysis. *Int J Data Min Bioinform*, 20(4):362–379.

Klinakis, A., Karagiannis, D., and Rampias, T. (2020). Targeting DNA repair in cancer: current state and novel approaches. *Cell Mol Life Sci*, 77:677–703.

Lancaster, H. O. (1949). The derivation and partition of χ^2 in certain discrete distributions. *Biometrika*, 36(1/2):117–129.

Leonardson, A. S., Zhu, J., Chen, Y., Wang, K., Lamb, J. R., Reitman, M., Emilsson, V., and Schadt, E. E. (2010). The effect of food intake on gene expression in human peripheral blood. *Hum Mol Genet*, 19(1):159–169.

Lin, S., Lin, Y., Nery, J. R., Urich, M. A., Breschi, A., Davis, C. A., Dobin, A., Zaleski, C., Beer, M. A., Chapman, W. C., Gingeras, T. R., Ecker, J. R., and Snyder, M. P. (2014). Comparison of the transcriptional landscapes between human and mouse tissues. *PNAS*, 111(48):17224–17229.

Lundby, A., Franciosa, G., Emdal, K. B., Refsgaard, J. C., Gnosa, S. P., Bekker-Jensen, D. B., Secher, A., Maurya, S. R., Paul, I., Mendez, B. L., Kelstrup, C. D., Francavilla, C., Kveiborg, M., Montoya, G., Jensen, L. J., and Olsen, J. V. (2019). Oncogenic mutations rewire signaling pathways by switching protein recruitment to phosphotyrosine sites. *Cell*, 179(2):543–560.

Lupianez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J. M., Laxova, R., Santos-Simarro, F., Gilbert-Dussardier, B., Wittler, L., Borschiwer, M., Haas, S. A., Osterwalder, M., Franke, M., Timmermann, B., Hecht, J., Spielmann, M., Visel, A., and Mundlos, S. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, 161(5):1012–1025.

Maekawa, M., Yamamoto, T., Tanoue, T., Yuasa, Y., Chisaka, O., and Nishida, E. (2005). Requirement of the map kinase signaling pathways for mouse preimplantation development. *Dev*, 132(8):1773–1783.

McDowall, M. D., Scott, M. S., and Barton, G. J. (2009). Pips: human protein–protein interaction prediction database. *Nucleic Acids Research*, 37(suppl_1):D651–D656.

McKenzie, A. T., Katsyv, I., Song, W.-M., Wang, M., and Zhang, B. (2016). DGCA: a comprehensive R package for differential gene correlation analysis. *BMC Syst Biol*, 10(1):106.

Mentzen, W. I., Floris, M., and de la Fuente, A. (2009). Dissecting the dynamics of dysregulation of cellular processes in mouse mammary gland tumor. *BMC Genomics*, 10(1):601.

Miyashita, S., Adachi, T., Yamashita, M., Sota, T., and Hoshino, M. (2017). Dynamics of the cell division orientation of granule cell precursors during cerebellar development. *Mech Dev*, 147:1–7.

Ouyang, Z., Song, M., Güth, R., Ha, T. J., Larouche, M., and Goldowitz, D. (2011). Conserved and differential gene interactions in dynamical biological systems. *Bioinform*, 27(20):2851–2858.

Ozaki, T., Muramatsu, R., Sasai, M., Yamamoto, M., Kubota, Y., Fujinaka, T., Yoshimine, T., and Yamashita, T. (2016). The p2x4 receptor is required for neuroprotection via ischemic preconditioning. *Sci Rep*, 6:25893.

Pan, D. (2010). The hippo signaling pathway in development and cancer. *Developmental Cell*, 19(4):491–505.

Paudel, S., Sindelar, R., and Saha, M. (2018). Calcium signaling in vertebrate development and its role in disease. *Int J Mol Sci*, 19(11):3390.

Phillips-Cremins, J. E., Sauria, M. E. G., Sanyal, A., Gerasimova, T. I., Lajoie, B. R., Bell, J. S. K., Ong, C.-T., Hookway, T. A., Guo, C., Sun, Y., Bland, M. J., Wagstaff, W., Dalton, S., McDevitt, T. C., Sen, R., Dekker, J., Taylor, J., and Corces, V. G. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell*, 153(6):1281–1295.

Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature*, 447(7143):425–432.

Sharma, R., Kumar, S., Zhong, H., and Song, M. (2017). Simulating noisy, nonparametric, and multivariate discrete patterns. *R J*, 9(2):366–377.

Sharma, R., Luo, X., Kumar, S., and Song, M. (2020). Three co-expression pattern types across microbial transcriptional networks of plankton in two oceanic waters. In *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*. Article No.: 14.

Sharma, R. and Song, M. (2020). *DiffXTables: Pattern Analysis Across Contingency Tables*. Comprehensive R Archive Network. R package.

Silvis, M. R., Kreger, B. T., Lien, W.-H., Klezovitch, O., Rudakova, G. M., Camargo, F. D., Lantz, D. M., Seykora, J. T., and Vasioukhin, V. (2011). α-catenin is a tumor suppressor that controls cell accumulation by regulating the localization and activity of the transcriptional coactivator YAP1. *Sci Signal*, 4(174):ra33–ra33.

Smith, E. M., Lajoie, B. R., Jain, G., and Dekker, J. (2016). Invariant TAD boundaries constrain cell-type-specific looping interactions between promoters and distal elements around the CFTR locus. *Am J Hum Genet*, 98(1):185–201.

Snell, G. D., editor (1956). *Biology of the Laboratory Mouse*. Dover, New York, 2nd edition.

Song, J., Zhong, H., and Wang, H. (2020). 'Ckmeans.1d.dp': Optimal, fast, and reproducible univariate clustering. R package version 4.3.3.

Song, M., Lewis, C. K., Lance, E. R., Chesler, E. J., Yordanova, R. K., Langston, M. A., Lodowski, K. H., and Bergeson, S. E. (2009). Reconstructing generalized logical networks of transcriptional regulation in mouse brain from temporal gene expression data. *EURASIP J Bioinform Syst Biol*, 2009:545176.

Song, M., Zhang, Y., Katzaroff, A. J., Edgar, B. A., and Buttitta, L. (2014). Hunting complex differential gene interaction patterns across molecular contexts. *Nucleic Acids Res*, 42(7):e57.

Song, M. and Zhong, H. (2020). Efficient weighted univariate clustering maps outstanding dysregulated genomic zones in human cancers. *Bioinformatics*, 36(20):5027–5036.

Steinhart, Z. and Angers, S. (2018). Wnt signaling in development and tissue homeostasis. *Development*, 145(11):dev146589.

Stelzer, G., Harel, A., Dalah, A., Rosen, N., Shmoish, M., Iny-Stein, T., Sirota, A., Madi, A., Safran, M., and Lancet, D. (2008). Genecards: One stop site for human gene research. *FISEB (ILANIT)*.

Stoeckli, E. T. (2018). Understanding axon guidance: are we nearly there yet? *Development*, 145(10).

Suurväli, J., Boudinot, P., Kanellopoulos, J., and Boudinot, S. R. (2017). P2X4: A fast and sensitive purinergic receptor. *Biomed J*, 40(5):245–256.

Thiagarajan, R. D., Morey, R., and Laurent, L. C. (2014). The epigenome in pluripotency and differentiation. *Epigenomics*, 6(1):121–137.

Van Roey, K., Dinkel, H., Weatheritt, R. J., Gibson, T. J., and Davey, N. E. (2013). The switches.ELM resource: a compendium of conditional regulatory interaction interfaces. *Sci Signal*, 6(269):rs7.

Vlachos, I. S., Zagganas, K., Paraskevopoulou, M. D., Georgakilas, G., Karagkouni, D., Vergoulis, T., Dalamagas, T., and Hatzigeorgiou, A. G. (2015). DIANA-miRPath v3.0: deciphering microRNA function with experimental support. *Nucleic Acids Res*, 43(W1):W460–W466.

Wang, H. and Song, M. (2011). Ckmeans.1d.dp: Optimal k-means clustering in one dimension by dynamic programming. *R J*, 3(2):29–33.

Wang, L., Zhou, K., Fu, Z., Yu, D., Huang, H., Zang, X., and Mo, X. (2017). Brain development and Akt signaling: the crossroads of signaling pathway and neurodevelopmental diseases. *J Mol Neurosci*, 61(3):379–384.

Watanabe, A., Yamada, Y., and Yamanaka, S. (2013). Epigenetic regulation in pluripotent stem cells: a key to breaking the epigenetic barrier. *Philos Trans R Soc Lond B Biol Sci*, 368(1609):20120292.

Xu, J., Bernstein, A. M., Wong, A., Lu, X.-H., Khoja, S., Yang, X. W., Davies, D. L., Micevych, P., Sofroniew, M. V., and Khakh, B. S. (2016). P2x4 receptor reporter mice: sparse brain expression and feeding-related presynaptic facilitation in the arcuate nucleus. *J Neurosci*, 36(34):8902–8920.

Xu, T., Ou-Yang, L., Yan, H., and Zhang, X.-F. (2019). Time-v/arying differential network analysis for revealing network rewiring over cancer progression. *IEEE/ACM Trans Comput Biol Bioinform*. doi: 10.1109/TCBB.2019.2949039. [Epub ahead of print].

Yang, Y., Yang, Y.-C. T., Yuan, J., Lu, Z. J., and Li, J. J. (2017). Large-scale mapping of mammalian transcriptomes identifies conserved genes associated with different cell states. *Nucleic Acids Res*, 45(4):1657–1672.

Zar, J. H. (2009). *Biostatistical Analysis*. Prentice Hall, New Jersey, 5th edition.

Zhang, W. and Liu, H. T. (2002). Mapk signal pathways in the regulation of cell proliferation in mammalian cells. *Cell Research*, 12(1):9–18.

Zhang, Y., Liu, Z. L., and Song, M. (2015). ChiNet uncovers rewired transcription subnetworks in tolerant yeast for advanced biofuels conversion. *Nucleic Acids Res*, 43(9):4393–4407.

Zhang, Y., Zhong, H., Nguyen, H., Sharma, R., Kumar, S., and Song, J. (2020). *'FunChisq': Model-Free Functional Chi-Squared and Exact Tests*. R package version 2.5.1.