



Gene locations may contribute to predicting gene regulatory relationships*

Jun MENG^{†‡1}, Wen-yuan XU¹, Xiao CHEN¹, Tao LIN², Xiao-yu DENG¹

¹Department of System Science and Engineering, School of Electrical Engineering, Zhejiang University, Hangzhou 310027, China

²Laboratory of Machine Learning and Optimization, École Polytechnique Fédérale de Lausanne (EPFL),
Route Cantonale, 1015 Lausanne 999034, Switzerland

[†]E-mail: junmeng@zju.edu.cn

Received June 8, 2017; Revision accepted Nov. 19, 2017; Crosschecked Dec. 15, 2017

Abstract: We propose that locations of genes on chromosomes can contribute to the prediction of gene regulatory relationships. We constructed a time-based gene regulatory network of zebrafish cardiogenesis on the basis of a spatio-temporal neighborhood method. Through the network, specific regulatory pathways and order of gene expression during zebrafish cardiogenesis were obtained. By comparing the order with locations of these genes on chromosomes, we discovered that there exists a reversal phenomenon between the order and order of gene locations. The discovery provides an inherent rule to instruct exploration of gene regulatory relationships. Specifically, the discovery can help to predict if regulatory relationships between genes exist and contribute to evaluating the correctness of discovered gene regulatory relationships.

Key words: Prediction; Gene regulatory relationships; Order of gene location; Spatio-temporal neighborhood method; Zebrafish cardiogenesis
<https://doi.org/10.1631/jzus.B1700303>

CLC number: TP181

1 Introduction


To avoid uncertainty which may be generated from the construction of gene regulatory relationships, some inherent rules are necessary to instruct the construction. Probabilistic integrated gene networks, where linkages between genes indicate their likelihood of being involved in the same biological process, are based on functional genomics, proteomics, and comparative genomics (Lee et al., 2008). Common methods of exploration on gene regulatory relationships include high-throughput biological experiments

and computational methods (Zhong and Sternberg, 2006; Lee et al., 2007, 2008, 2010; Ma et al., 2007; Huttenhower et al., 2009; Bassel et al., 2011; Kourmpetis et al., 2011). Although functional genomics, proteomics, and comparative genomics contain much valuable data, they are incomplete and error-prone (Lee et al., 2008). In addition, exploration methods of regulatory relationships between genes also contain some blindness. This blindness together with the incompleteness of databases leads to the conclusion that a constructed gene regulatory network may have uncertainties. Therefore, we need some inherent features of gene regulatory relationships to instruct constructions.

Existing creatures are results of evolution; the embryogenesis of creatures should contain a certain amount of evolutionary information. During evolution, length and complexity of genes increase in the

[‡] Corresponding author

* Project supported by the National Natural Science Foundation of China (Nos. 60574079, 30872957, and 30572129)

 ORCID: Jun MENG, <https://orcid.org/0000-0002-7633-3624>

© Zhejiang University and Springer-Verlag GmbH Germany, part of Springer Nature 2018

space due to the addition of new genes, with conservation of “good” genes in the creatures under natural selection. Conserved genes in creatures are sequentially expressed during embryogenesis to development into mature individuals.

Take the evolution of heart and cardiogenesis of human as an example (Olson, 2006; Nemer, 2008). Beginning with cardiac muscle cells of Cnidaria to the pharynx of Nematodes, the pharynx preliminarily contracts like a heart. It is believed that the first heart-like organ appeared over 4 billion years ago, which resembled the simple tubular vessel-like organs of tunicates (Romer, 1967; Moorman and Christoffels, 2003; Bishopric, 2005; Simões-Costa et al., 2005; Lee et al., 2010). The first heart-like organ evolves from a single-layered tube with peristaltic contractility to a more efficient and powerful pump with thick muscular chambers dedicated to receiving and pumping blood (Moorman and Christoffels, 2003; Simões-Costa et al., 2005; Olson, 2006). To transit from an aquatic to a terrestrial environment, the hearts of fish contain a single atrial and a single ventricle (Olson, 2006), while the hearts of amphibians have two atria and a single ventricle, and amniotes have a four-chambered heart (Olson, 2006). Consistent with the evolution of the heart, formation of a human heart starts at early stages of embryogenesis with commitment of anterior lateral plate mesoderm cells to the cardiogenic lineage and their migration and organization into the cardiac crescent which later forms into a beating linear heart tube (Nemer, 2008). The linear heart tube contains a single atrial and a single ventricle, which undergoes chamber specification, septation, and trabeculation to form the definitive four-chambered heart (Nemer, 2008). During chamber specification, the atrial of the linear heart tube is specified before the ventricle. This example demonstrates that there probably exists some correlation between evolution and embryogenesis. Therefore, if we can find the correlation, we may discover some inherent features to instruct construction of gene regulatory relationships.

If the correlation between evolution and embryogenesis exists, there should be a correlation between spatial features and temporal features of genes. Length and complexity of genes increase in the space due to addition of new genes during evolution. The addition leads to formation of orders of gene locations

on chromosomes. Although chromosomal rearrangement occurs during evolution, orders of some genes on the partial chromosomes are conserved. The conserved gene locations have been discovered to determine or reflect a mechanism of genetic control (Anchel et al., 2016). During embryogenesis, genes are sequentially expressed by time, where there exists time order. Through comparison of orders of gene locations on chromosomes and the time order, we can deduce that if there exists some correlation between evolution and embryogenesis, and this helps us to determine that whether evolution can instruct the construction of regulatory relationships between genes. To simplify study, we only compare spatial features and temporal features of genes on the same chromosome, without consideration of chromosomal rearrangement during evolution.

2 Materials and methods

2.1 Prior knowledge of zebrafish cardiogenesis

The specific progress of zebrafish cardiogenesis is as follows (Stainier, 2001; Glickman and Yelon, 2002). Just before gastrulation, 5 h post fertilization (hpf), the heart progenitor cells are located throughout the ventral and lateral regions of the embryo. After involution, these cells converge towards the embryonic axis and reach their destination at the level of the future hindbrain by the five-somite stage (about 12 hpf). Three rows of cells are represented at this stage. The myocardial precursors lie mostly laterally and the endocardial precursors mostly medially (Alexander and Stainier, 1999). By the 13-somite stage (15.5 hpf), the myocardial precursors have segregated into preatrial and preventricular groups, although this segregation might well happen earlier (Yelon et al., 1999). Starting at 19 hpf, the myocardial precursors merge posteriorly to form a horseshoe-shaped structure (Yelon et al., 1999). By 19.5 hpf, as anterior cells migrate medially, the horseshoe transforms into a cone with the ventricular cells at its center and apex, and the atrial cells at its base. The endocardial cells line the inside of the cone. Next, the cone telescopes out to form a tube. The ventricular end of the heart tube assembles first, followed by the atrial end. By 24 hpf, the tube lies along the anteroposterior axis with the atrial end to the left of the midline. Subsequently, by

30 hpf, visibly distinct ventricular and atrial chambers form. By 36 hpf, the heart undergoes looping morphogenesis. By 48 hpf, functional valves are formed. Based on the time, we divide the process of cardiogenesis of zebrafish into developmental stages (Table 1).

After clarifying the developmental stages of the cardiogenesis of zebrafish, we have arranged functions of related genes (Table 2) and microRNAs (miRNAs) (Table 3), preparing for the construction of a gene regulatory network.

Table 1 Developmental stages of cardiogenesis of zebrafish

Time (hpf)	Developmental stages of cardiogenesis
0–5	Involution of heart progenitor cells
12–15.5	Heart progenitor cells differentiate into myocardial precursors
15.5–24	Proliferation of myocardial precursors and morphogenesis of heart
24–30	Morphological development of heart and development of ventricle
30–36	Development of ventricle and atrial
36–48	Looping morphogenesis
≥48	Formation of functional valves

Table 2 Arrangement of gene functions related to cardiogenesis of zebrafish

Gene	Functions of the gene
<i>Fgf8a</i>	<i>Fgf8</i> is expressed in cardiac precursors and later in the heart ventricle (Reifers et al., 2000). <i>Fgf8</i> is expressed in and required for development of the zebrafish heart precursors, particularly during initiation of cardiac gene expression (Reifers et al., 2000)
<i>Gata4</i>	There are interactions between <i>Gata4</i> and <i>Nkx2.5</i> . <i>Gata4</i> first functions in the early development of cardiac precursor cells (Serbedzija et al., 1998). <i>Gata4</i> is regulators of proliferation of myocardial precursors (Kikuchi et al., 2010). <i>Gata4</i> is an important regulator of cardiac hypertrophy, which functions together with <i>SRF</i> , <i>Nkx2.5</i> , <i>Tbx5</i> , and <i>Gata6</i> (Akazawa and Komuro, 2003)
<i>Tbx20</i>	<i>Tbx20</i> expression is enriched in the atrioventricular canal (AVC), the outflow tract and right ventricle but is reduced in the left ventricle (Plageman and Yutzey, 2004; Ataliotis et al., 2005). <i>Tbx1</i> interacts with the <i>Tbx20</i> during cardiac development, especially during the development of the outflow tract. <i>Tbx20</i> physically interacted with cardiac transcription factors <i>Nkx2-5</i> , <i>Gata4</i> , and <i>Gata5</i> , collaborating to synergistically activate cardiac gene expression (Kloosterman et al., 2006). <i>Tbx20</i> regulates development in myocardial and acts in synergy with other cardiac transcription factors of the homeodomain and zinc finger families, including <i>Nkx2.5</i> , <i>Gata 4</i> , and <i>Gata5</i> (Stennard et al., 2003)
<i>Nkx2.5</i>	<i>Nkx2.5</i> is expressed in early cardiac mesoderm and in heart muscle lineage throughout life. Targeted disruption is lethal to embryos and arrests cardiac development at the linear heart tube stage (Benson et al., 1999)
<i>Has2</i>	<i>Has2</i> is a crucial component of the cardiac jelly matrix, which requires <i>Bmp2</i> for myocardial expression (Ma et al., 2005)
<i>Tbx5a</i>	Increased expression of <i>Tbx5</i> leads to an increase in <i>Ndr4</i> expression, which promotes proliferation of cardiomyocytes and formation of atrioventricular canal (Qu et al., 2008; Ghosh et al., 2009). Maintenance of <i>Tbx5</i> expression and <i>cmlc2</i> -expressing are indispensable for myocardial differentiation (Yelon et al., 2000)
<i>Hand2</i>	<i>Hand2</i> is important for maintenance of <i>Tbx5</i> expression, <i>vmhc</i> expression, and heart tube formation (Yelon et al., 2000)
<i>Bmp2b</i>	<i>Bmp2b</i> expresses during the early stages of heart development. <i>Bmp2b</i> has important relations with cardiac differentiation in anterior medial mesoderm and expressions of some important genes related with cardiac development, such as <i>Nkx2.5</i> , <i>Gata4/5/6</i> (Schultheiss et al., 1997; Schlange et al., 2000)
<i>Tbx2a</i>	<i>Tbx2a</i> is essential for chamber differentiation, patterning the atrioventricular canal and morphogenesis of the outflow tract during heart development (Harrelson et al., 2004)
<i>Cx43</i>	Defects of <i>Cx43</i> lead to hindrance of cardiac looping morphogenesis (Liu et al., 2008)
<i>Tbx1</i>	With a 20% decrease of <i>Tbx1</i> , abnormal cardiac morphogenesis occurs due to hindrance of cardiac looping morphogenesis, including no-looping or wrong looping direction (Zhang et al., 2010). Decrease of <i>Tbx1</i> impairs myocardial function and leads to decrease of cardiac contraction and heart rate

Table 3 Arrangement of miRNA functions related to cardiogenesis of zebrafish

miRNA	Functions of the miRNA
<i>dre-miR-143</i>	<i>miR-143</i> is required for chamber morphogenesis through direct repression of <i>adducin3</i> . Knockdown of <i>miR-143</i> leads to ventricular collapse and decreased contractility (Moulton, 2010)
<i>dre-miR-138</i>	<i>miR-138</i> is expressed in ventricular cardiomyocyte and regulates formation of single-ventricle during cardiac development (Morton et al., 2008). Disruption of <i>miR-138</i> function leads to ventricular expansion of gene expression normally restricted to the atrio-ventricular valve region and, ultimately, to disrupted ventricular cardiomyocyte morphology and cardiac function (Morton et al., 2008)
<i>dre-miR-145</i>	Loss of <i>miR-145</i> results in defects of heart and gut. In early zebrafish development between 16 and 24 hpf, <i>miR-145</i> is expressed ubiquitously at very low levels with increasingly specific expression in the heart, ear, and pharyngeal arches between 24 and 72 hpf (Zeng et al., 2009)
<i>dre-miR-218</i>	<i>miR-218</i> regulates heart morphogenesis through <i>miR-218</i> during formation of ventricle (Chiavacci et al., 2012). Down-regulation of <i>miR-218</i> is able to rescue the heart defects generated by <i>Tbx5</i> over-expression supporting the notion that <i>miR-218</i> is a crucial mediator of <i>Tbx5</i> in heart development and suggesting its possible involvement in the onset of heart malformations (Chiavacci et al., 2012)
<i>dre-miR-499</i>	<i>miR-499</i> is expressed in the ventricle and atrium, the muscles of the head, and the somitic muscles (Kloosterman et al., 2006). The temporal expression of <i>miR-499</i> in situ is from 48 h to adult (Kloosterman et al., 2006)

2.2 Regulatory relationships between miRNAs and RNAs

We deduce that the function of mRNA against miRNA is that the transcription of mRNA promotes the generation of miRNA. Details of this deduction are shown as follows. miRNAs are produced from either their own genes or from introns. miRNA is divided into two categories according to sources, intergenic miRNA and intronic miRNA. Intergenic miRNA generates from non-protein coding genes. Thus miRNAs are initially transcribed as part of one arm of a RNA stem-loop (about 80 nucleotide (nt)) that in turn forms part of a several hundred nucleotides-long miRNA precursor termed a primary miRNA (Cai et al., 2004; Lee et al., 2004). Precursor miRNAs, about 70 nt hairpin structures, are in turn excised from a primary miRNA transcript by the enzyme Drosha. Finally, mature miRNAs are cleaved from pre-mRNAs by the enzyme Dicer (Esquela-Kerscher and Slack, 2006; Chen and Rajewsky, 2007; Okamura et al., 2007). The intronic miRNA is defined as small RNAs, derived from introns, processed by splicing and pre-miRNA degradations, and targets the intracellular mRNA for silencing of that gene expression (Ying and Lin, 2006). In consideration of the reliability of DNA data resources and clarity of DNA regionalism, we only take intronic miRNAs into account in this research.

A corollary of this deduction is that we can deduce whether a miRNA derives from an mRNA

through comparison of matching degree between the miRNA and introns of the mRNA. The probability of the miRNA deriving from the mRNA increases with the growth of matching times between the miRNA and introns of the mRNA. Hence, translation of mRNA, which means production of more introns, promotes generation of miRNA.

With respect to the effect of miRNA on mRNA, we mainly consider the effect of the regulation of miRNA on the translation of mRNA. The deduction is that miRNA will promote the translation of target mRNA. The specific reasoning process is as follows. It has been demonstrated that if there is complete complementation between the miRNA and target mRNA sequence, Ago2 can cleave the mRNA and lead to direct mRNA degradation. Yet, if there is not complete complementation the silencing is achieved by preventing translation (Lim et al., 2005). During the development of creatures, the level of receptor proteins remains relatively stable, and thus the number of receptor proteins expressed through mRNAs remains relatively stable. In the case that the number of receptor proteins expressed through mRNAs remains relatively stable, complete complementation between the miRNA and target mRNA sequence will lead to direct mRNA degradation, and therefore promote the translation of the target mRNA to keep the number of target mRNA relatively stable. In conclusion, if receptor proteins during development remain stable, complete complementation between

the miRNA and target mRNA sequence will promote the translations of mRNAs.

A corollary of the deduction is that we can deduce whether a miRNA promotes translation of an mRNA through comparison of matching degree between the miRNA and exons of the mRNA. The more times the miRNA matches with exons of the mRNA, which means more probability of complete complementation, the more mRNA will be degraded. The degradation will in turn promote translation of the mRNA. Hence, miRNA can promote the translation of mRNA, while the expression of protein is relatively stable.

In summary, some mRNAs produce mature miRNAs, which will lead to degradation of target mRNAs of miRNAs. Since receptor proteins during development remain stable, the degradation of target mRNAs promotes translations of the target mRNAs, and this will in turn promote productions of miRNAs. The conclusion is that introns of genes translate to form miRNAs, which in turn regress function of mRNA through complementary base pairing with exons of genes. This process includes mutual regulations of introns, exons, and miRNAs, based on which we can construct gene regulatory relationships through miRNAs. The following constructions of network will be based on this deduction.

2.3 Spatio-temporal neighborhood method

The nonlinear spatio-temporal neighborhood method is a data mining method which attempts to acquire casual links between dynamic data by clustering the data. The main notions and procedures of the method are presented as follows.

The spatio-temporal neighborhood method is a method dedicated to exploring relations between spatial features and temporal features during the process of development. Assume there are n points $e_1^t, e_2^t, e_3^t, \dots, e_n^t$ in space at time t and mapping parameter from e_i^t to e_j^t is λ_{ij}^t . Due to differences of parameters, these points form into clusters $C_1^t, C_2^t, \dots, C_m^t$, $C_k^t = \{c_{k1}^t e_1^t, c_{k2}^t e_2^t, \dots, c_{kn}^t e_n^t\}$ at time t , where $c_{kl}^t = \begin{cases} 0, & e_l^t \notin C_k^t \\ 1, & e_l^t \in C_k^t \end{cases}$. When time changes, clusters will change owing to changes of mapping parameters.

The temporal neighborhood that spatio-temporal neighborhood method takes research into includes t_1, t_2, \dots, t_p , while the spatial neighborhood includes clusters of all the points in the space at time t , such as $C_1^t, C_2^t, C_3^t, \dots, C_m^t$. By analyzing temporal changes of clusters in the spatial neighborhood, the evolutionary rule $F = \{f_1, f_2, \dots, f_l\}$ can be obtained based on changes of mapping parameters. The rule F between clusters at time t and $t+\Delta t$ is shown as follows:

$$\begin{aligned}
 & \begin{bmatrix} C_1^{(t+\Delta t)} \{c_{11}^{(t+\Delta t)} e_1^{(t+\Delta t)}, c_{12}^{(t+\Delta t)} e_2^{(t+\Delta t)}, \dots, c_{1n}^{(t+\Delta t)} e_n^{(t+\Delta t)}\} \\ C_2^{(t+\Delta t)} \{c_{21}^{(t+\Delta t)} e_1^{(t+\Delta t)}, c_{22}^{(t+\Delta t)} e_2^{(t+\Delta t)}, \dots, c_{2n}^{(t+\Delta t)} e_n^{(t+\Delta t)}\} \\ \vdots \\ C_m^{(t+\Delta t)} \{c_{m1}^{(t+\Delta t)} e_1^{(t+\Delta t)}, c_{m2}^{(t+\Delta t)} e_2^{(t+\Delta t)}, \dots, c_{mn}^{(t+\Delta t)} e_n^{(t+\Delta t)}\} \end{bmatrix} \\
 &= F \begin{bmatrix} C_1^t \{c_{11}^t e_1^t, c_{12}^t e_2^t, \dots, c_{1n}^t e_n^t\} \\ C_2^t \{c_{21}^t e_1^t, c_{22}^t e_2^t, \dots, c_{2n}^t e_n^t\} \\ \vdots \\ C_m^t \{c_{m1}^t e_1^t, c_{m2}^t e_2^t, \dots, c_{mn}^t e_n^t\} \end{bmatrix}.
 \end{aligned}$$

Through the rule F , we can obtain

$$\begin{bmatrix} c_i^{(t+\Delta t)} e_i^{(t+\Delta t)} \\ c_j^{(t+\Delta t)} e_j^{(t+\Delta t)} \end{bmatrix} = f_l \left(\begin{bmatrix} c_i^t e_i^t \\ c_j^t e_j^t \end{bmatrix} \right)$$

can contribute to predictions of $e_i^{t+\Delta t}$. Based on the predictions, clusters of $t+\Delta t$ can be predicted, where some important bifurcations can be discovered by comparing with clusters formed at time t .

Take an example to demonstrate procedures of spatio-temporal neighborhood method (Fig. 1). There are eight points $e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8$ in space. The eight points form into two clusters $C_1^{t_1}, C_2^{t_1}$ at time t_1 , while these points form into new clusters $C_1^{t_2}, C_2^{t_2}, C_3^{t_2}$ under rules f_1, f_2, f_3 at time t_2 , respectively. At time t_3 , new clusters $C_1^{t_3}, C_2^{t_3}, C_3^{t_3}$ are formed under rules f_1, f_2 , and f_3 , respectively. We can observe that there exists a strong correlation between clusters $C_1^{t_1}, C_1^{t_2}$, and $C_1^{t_3}$, which is shown as follows: (1) If $e \in C_1^{t_1}$ when $t=t_1$ and $e \in C_1^{t_2}$ when $t=t_2$, then $e \in C_1^{t_3}$ can be obtained when $t=t_3$; (2) If $e \in C_1^{t_1}$ when $t=t_1$ and $e \in C_2^{t_2}$ when $t=t_2$, then $e \in C_2^{t_3}$ can be obtained when $t=t_3$.

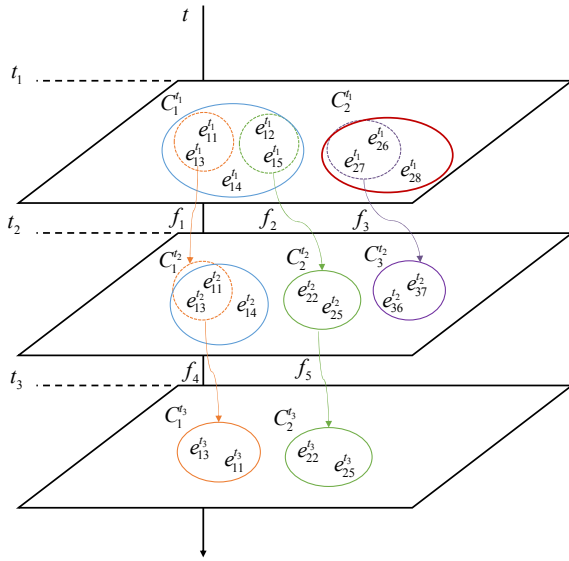


Fig. 1 An example to illustrate theoretical procedures of the spatio-temporal neighborhood method

e_{ij}^t represents a point in the space at the time of t , wherein t represents time, i represents clusters that the point belongs to and j represents label of the point. For example, e_{12}^t represents the point labeled two at the time of t_1 belongs to cluster one. C_j^t represents a cluster, wherein t represents time and j represents label of the cluster. For example, C_1^t represents the first cluster at the time of t_1 . f_j represents different rules for different j , which leads to formation of new clusters

In addition, we observe $e_1, e_2, e_3, e_4, e_5 \in C_1^{t_1}$ when $t=t_1$, while influenced by changes of rule f which result from changes of mapping parameters, $e_1, e_3, e_4 \in C_1^{t_2}$ and $e_2, e_5 \in C_2^{t_2}$ at time $t=t_2$, $e_1, e_3 \in C_1^{t_3}$ and $e_2, e_5 \in C_2^{t_3}$ at time $t=t_3$. Through changes of clusters that points belong to, we can obtain temporal bifurcations which contain significant information that represents the generation of new functions during evolution in the biology.

2.4 Multi-layer sub-network construction method

The multi-layer sub-network construction method obeys specific rules, which are:

(1) When we construct sub-network $n+1$ on the base of sub-network n at stage m , ignore regulatory relationships in the sub-network n , and only consider regulation of gene in the sub-network n on the added miRNAs of the sub-network $n+1$ and regulation of added genes in the sub-network $n+1$ on miRNAs of the sub-network n . In this case, we presume that the

sub-network n already exists at stage m and regulatory relationships have been formed at stage $m-1$.

(2) When we construct regulatory relations between stages m and $m+1$, ignore existed regulatory relationships between stages $m-1$ and m , and only take newly added regulatory relationships between stages m and $m+1$ into consideration.

(3) When a newly added miRNA at stage m regulates a gene of the sub-network n at stage $m+1$, we deduce that the miRNA promotes the formation of mRNA corresponding to the gene. In this case, the gene particularly refers to a newly added gene of the sub-network n with respect to the sub-network $n-1$.

(4) When a miRNA of sub-network $n+1$ is regulated by a gene of sub-network n , we deduce that the sub-network n has regulated the miRNA. In this case, the gene particularly refers to a newly added gene of the sub-network n with respect to the sub-network $n-1$.

(5) If a gene or a miRNA does not belong to sub-network n and has no relations with other genes or miRNAs, then we deduce that the gene or miRNA does not participate in the regulation at this time.

3 Results

To prove that orders of gene locations on chromosomes can instruct exploration on gene regulatory relationships, the time order that genes express in during cardiogenesis of zebrafish and location of genes on the same chromosome were compared. Also a reversal phenomenon was discovered to exist between the time order and order of gene locations on chromosomes (Fig. 2).

Based on spatio-temporal concentrates of miRNAs and expressing genes during zebrafish cardiogenesis (Schultheiss et al., 1997; Serbedzija et al., 1998; Alexander and Stainier, 1999; Benson et al., 1999; Yelon et al., 1999, 2000; Schlange et al., 2000; Reifers et al., 2000; Stainier, 2001; Glickman and Yelon, 2002; Akazawa and Komuro, 2003; Stennard et al., 2003; Harrelson et al., 2004; Plageman and Yutzey, 2004; Ataliotis et al., 2005; Ma et al., 2005; Kloosterman et al., 2006; Prall et al., 2007; Liu et al., 2008; Morton et al., 2008; Qu et al., 2008; Ghosh et al., 2009; Zeng et al., 2009; Kikuchi et al., 2010; Moulton, 2010; Zhang et al., 2010; Chiavacci et al.,

2012), time was introduced into the construction of the gene regulatory network (Fig. 3). miRNAs are endogenous about 23 nt RNAs that play important gene-regulatory roles in animals by pairing to the mRNAs of protein-coding genes to direct their post-transcriptional repression (Cai et al., 2004; Lee et al., 2004; Chen and Rajewsky, 2007; Chen et al., 2015). Through variations of miRNAs' spatio-temporal concentrates, the expression time of genes can be determined.

The construction of a time-based gene regulatory network is achieved through a spatio-temporal

neighborhood method. The spatio-temporal neighborhood method connects every sub-network that centers on a specific miRNA. From the constructed time-based gene regulatory network, the time order that genes express during the cardiogenesis of zebrafish can be obtained.

3.1 Determination of up-stream and down-stream mRNAs of miRNAs

Based on the deduction of regulatory relationships between miRNAs and mRNAs, we explore relationships between miRNA and mRNA by MATLAB.

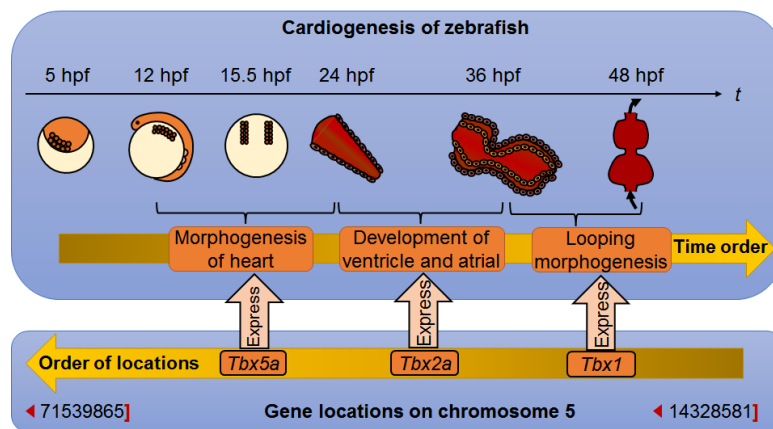


Fig. 2 A reversal phenomenon between time order of genes' expression during cardiogenesis of zebrafish and order of gene locations on chromosomes

Through a constructed time-based gene regulatory network, we found that *Tbx5a* first expresses during morphogenesis of heart (12–24 hpf), *Tbx2a* first expresses during development of ventricle and atrial (24–36 hpf), and *Tbx1* first expresses during looping morphogenesis (36–48 hpf). Therefore, the time order from beginning to end between these genes is *Tbx5a*, *Tbx2a*, *Tbx1*. These genes are all located on the chromosome 5 of zebrafish, and the order of locations from begin to end is *Tbx1*, *Tbx2a*, *Tbx5a*. Obviously, the order of locations of these genes is exactly the reverse of their time order

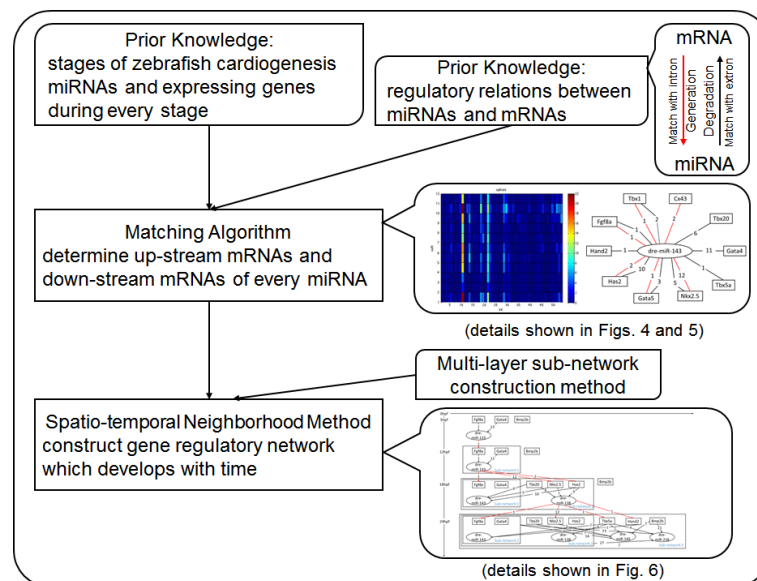


Fig. 3 Overview of the construction of gene regulatory network which develops with time

The deduction is that introns of mRNAs translate to generate miRNAs, while the generation regresses functions of mRNAs through complementary base pairing with exons of mRNAs (Cai et al., 2004; Lee et al., 2004; Lim et al., 2005; Esquela-Kerscher and Slack, 2006; Ying and Lin, 2006; Chen and Rajewsky, 2007; Okamura et al., 2007).

Using a matching algorithm which takes only complete complementation as a match, we obtain relationships between miRNAs and introns of mRNAs (Fig. 4a), which determines the up-stream of miRNAs. Based on the same method, we acquire relationships between miRNAs and exons of mRNAs (Fig. 4b), which determines the down-stream of miRNAs. To clarify the up-stream and down-stream of miRNAs, we have visualized matching results of Fig. 4a and Fig. 4b in terms of miRNAs and mRNAs (Fig. 5).

3.2 Construction of a time-based gene regulatory network

Based on prior knowledge of zebrafish cardiogenesis and determination of up-stream and down-stream of miRNAs, we adopt a spatio-temporal neighborhood method to construct a time-based gene regulatory network.

The constructed time-based gene regulatory network (Fig. 6) demonstrates procedures of the

morphological evolution of a zebrafish's heart, which have guiding significance on the research of morphological evolution of zebrafish's heart.

Although we know that *Bmp2b* and *Gata4* probably occur earliest at stage 5 hpf, the constructed time-based gene regulatory network shows that *Bmp2b* participates in the regulation at time 24 hpf instead of 5 hpf. Although the inconsistency reduces completeness of the network, it cannot impact on the correctness of the regulatory network. Overall, the constructed time-based gene regulatory network is consistent with current work.

Through the time-based gene regulatory network (Fig. 6), we have obtained the time order of main functioning genes during cardiogenesis. The time order from beginning to the end is *Fgf8a*, *Gata4*, *Tbx20*, *Nkx2.5*, *Has2*, *Tbx5a*, *Hand2*, *Bmp2b*, *Tbx2a*, *Cx43*, *Tbx1*.

3.3 Comparison of spatial feature and temporal feature of genes

In order to compare the time order with the order of gene locations on chromosomes, we have arranged locations of these genes according to the National Center for Biotechnology Information (NCBI) database (Table 4). The arrangement shows that *Tbx5a*, *Tbx2a*, and *Tbx1* are all on the same chromosome 5 and the location order is *Tbx1*, *Tbx2a*, and *Tbx5a*. Since the time order we have from the time-based

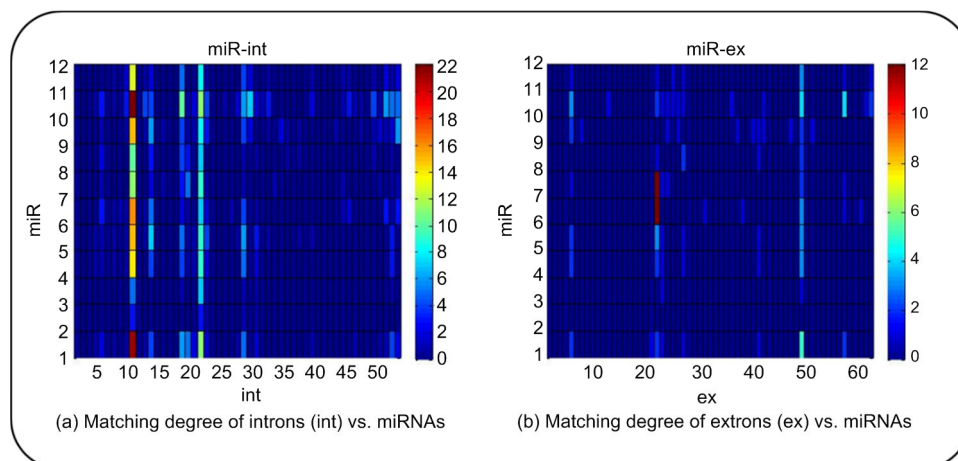


Fig. 4 Determination of up-stream and down-stream mRNAs of miRNAs based on matching degree

(a) Y-axis represents miRNAs of related genes, X-axis refers to introns of related genes and colors represent times that a miRNA matches with introns of an mRNA. Based on prior knowledge of regulatory relationships between miRNAs and mRNAs, we recognize an mRNA whose introns match with a miRNA at least one time as up-stream of the miRNA. (b) Y-axis represents miRNAs of related genes, X-axis represents exons of related genes and colors represent times that a miRNA matches with exons of an mRNA. According to prior knowledge of regulatory relationships between miRNAs and mRNAs, we obtain an mRNA whose exons match with a miRNA at least one time as down-stream of the miRNA

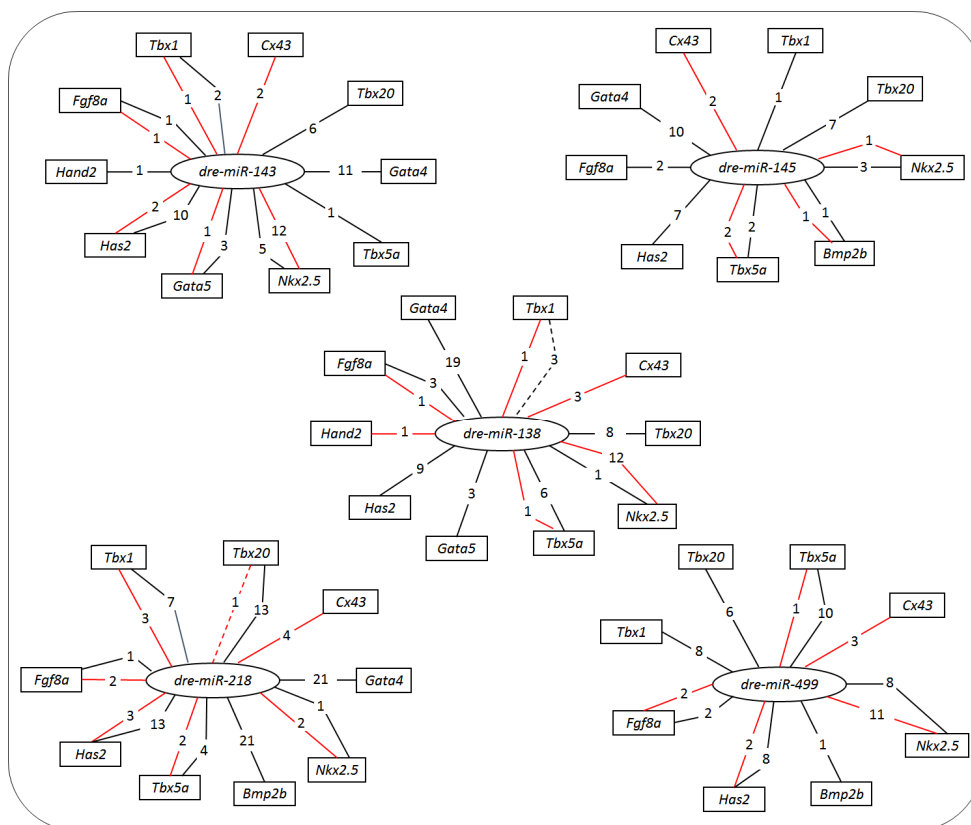


Fig. 5 Visualization of the matching results of miRNAs and genes in center of miRNAs

Black lines represent up-stream of miRNAs and red lines represent down-stream of miRNAs. The number on black lines represents times that a miRNA matches with introns of an mRNA, while the number on red lines refers to times that a miRNA matches with exons of an mRNA (Note: for interpretation of the references to color in this figure legend, the reader is referred to the web version of this article)

gene regulatory network is *Tbx5a*, *Tbx2a*, and *Tbx1*, a reverse is found between location order and time order. The reverse is also found between *Tbx20* and *Has2*, as well as *Gata4*, *Bmp2b*, and *Cx43*.

4 Discussion

To avoid the uncertainty which can be generated from the construction of gene regulatory relationships, some inherent rules are necessary to instruct the construction. By comparing the time order of the main functioning genes during cardiogenesis of a zebrafish and their locations on chromosomes, we find that there exists a reverse between time order that genes express in and the order of gene locations on the same chromosome. For example, genes *Tbx1*, *Tbx2a*, and *Tbx5a* express in the order of *Tbx5a*, *Tbx2a*, and *Tbx1*, while locations of these genes on chromosome is in the order of *Tbx1*, *Tbx2a*, and *Tbx5a*.

Based on this reversal phenomenon, orders of gene locations can help to predict gene regulatory relationships. For example, if gene A's location on a chromosome is before gene B's location on the same chromosome, then during embryogenesis, gene A will probably express after gene B.

Through the prediction, the number of genes to be selected can be much reduced when researchers try to determine the following expressing gene in a gene regulatory network. For example, if gene C's express time has been determined in a gene regulatory network, then genes expressing after gene C's are mostly located before gene C.

Furthermore, the prediction can contribute to evaluating the correctness of discovered regulations between genes. For example, if gene D's location on a chromosome is before gene E's location on the same chromosome, and gene D expresses before gene E in a gene regulatory network, then work is needed to verify the correctness of the gene regulatory network.

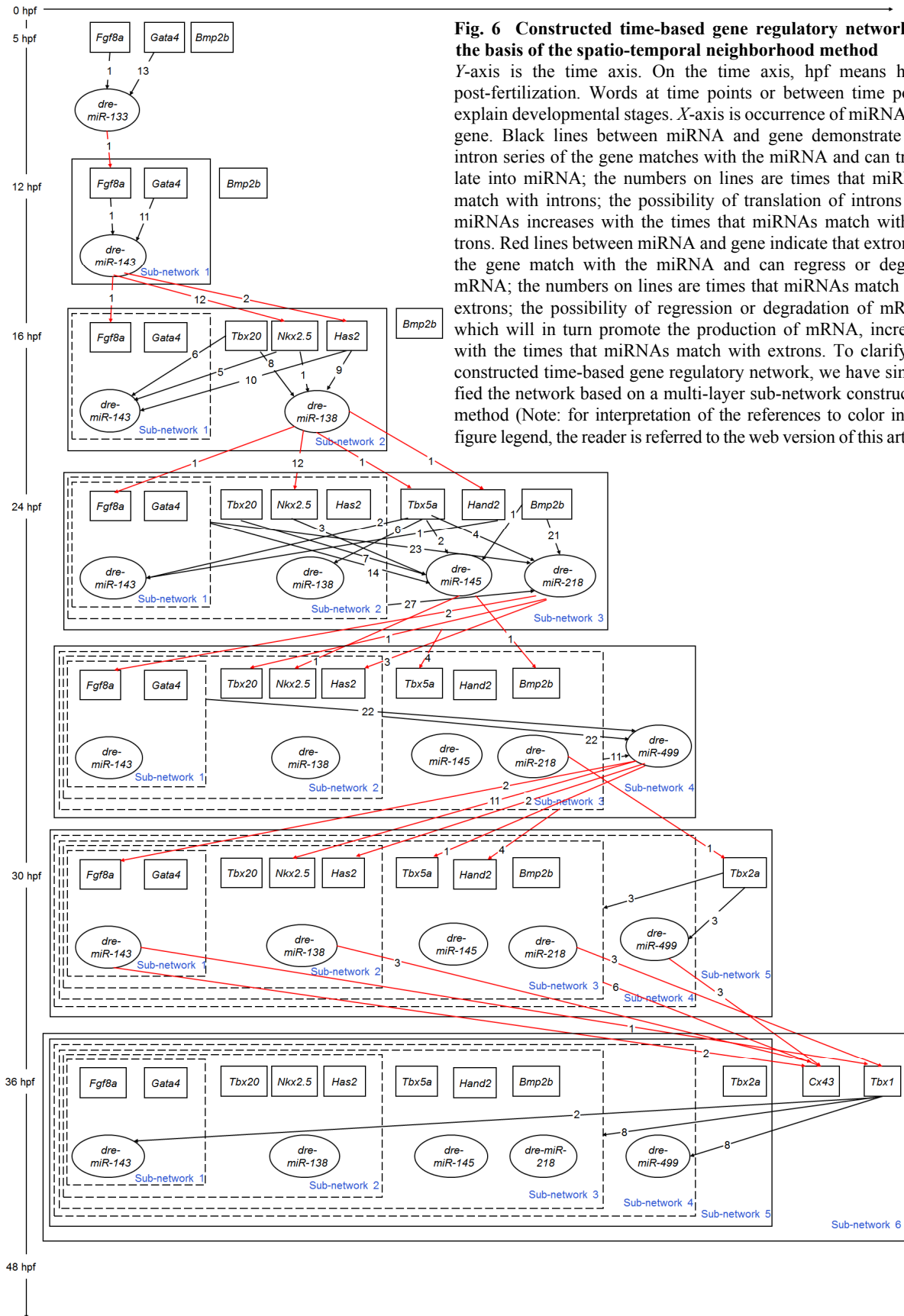


Fig. 6 Constructed time-based gene regulatory network on the basis of the spatio-temporal neighborhood method

Y-axis is the time axis. On the time axis, hpf means hours post-fertilization. Words at time points or between time points explain developmental stages. X-axis is occurrence of miRNA and gene. Black lines between miRNA and gene demonstrate that intron series of the gene matches with the miRNA and can translate into miRNA; the numbers on lines are times that miRNAs match with introns; the possibility of translation of introns into miRNAs increases with the times that miRNAs match with introns. Red lines between miRNA and gene indicate that exons of the gene match with the miRNA and can regress or degrade mRNA; the numbers on lines are times that miRNAs match with exons; the possibility of regression or degradation of mRNA, which will in turn promote the production of mRNA, increases with the times that miRNAs match with exons. To clarify the constructed time-based gene regulatory network, we have simplified the network based on a multi-layer sub-network construction method (Note: for interpretation of the references to color in this figure legend, the reader is referred to the web version of this article)

Table 4 Arrangement of gene locations on chromosomes

Related chromosome where gene is located	Gene	Location of gene on the chromosome
<i>Danio rerio</i> strain Tuebingen chromosome 13, Zv9 (NC_007124.6)	<i>Fgf8a</i>	28365067–28371535
<i>Danio rerio</i> strain Tuebingen chromosome 16, Zv9 (NC_007127.6)	<i>Tbx20</i>	42605255–42619962
	<i>Has2</i>	15481299–15497339
<i>Danio rerio</i> strain Tuebingen chromosome 5, Zv9 (NC_007116.6)	<i>Tbx5a</i>	71474708–71505188
	<i>Tbx2a</i>	55834959–55843728
	<i>Tbx1</i>	14701844–14716157
<i>Danio rerio</i> strain Tuebingen chromosome 20, Zv9 (NC_007131.6)	<i>Gata4</i>	53146771–53162461
	<i>Bmp2b</i>	45989447–45995178
	<i>Cx43</i>	40818053–40823343
<i>Danio rerio</i> strain Tuebingen chromosome 14, Zv9 (NC_007125.6)	<i>Nkx2.5</i>	23943996–23946789
<i>Danio rerio</i> strain Tuebingen chromosome 1, Zv9 (NC_007112.6)	<i>Hand2</i>	37473969–37475834

Although the reverse between time order that genes express in and order of gene locations on the same chromosome has been found during the cardiogenesis of zebrafish, there still needs to be more research to establish whether this reversal also exists in other species. In another paper, we have studied the correlation between time order and order of gene locations in the development of *Arabidopsis thaliana* and we have a conclusion consistent with this paper (data not shown).

The reverse we have obtained between time order that genes express in and order of gene locations on the same chromosome provides a rule for construction of a gene regulatory network. Based on the rule, the uncertainty generated from construction of gene regulatory relationships will be decreased, and this will increase the correctness of gene regulatory network and save much research efforts. Most importantly, the paper has not only provided an easier way to construct a gene regulatory network with gene locations on the chromosome determining the basic structure of a network, but has also shown a dynamic relationship between genes since miRNAs can determine the expression of genes and the time at which they express.

Acknowledgements

We would like to thank miRBase database for providing all the miRNA data until June 2013 through the file mature.fa and NCBI database for gene locations on chromosomes. We are particularly grateful to Shu ZHENG in the Second Affiliated Hospital, School of Medicine, Zhejiang University (Hangzhou, China) for valuable advice on miRNAs and embryogenesis, Zhao-lin WANG for instructing the writing of this manuscript.

Compliance with ethics guidelines

Jun MENG, Wen-yuan XU, Xiao CHEN, Tao LIN, and Xiao-yu DENG declare that they have no conflict of interest.

All institutional and national guidelines for the care and use of laboratory animals were followed.

References

- Akazawa H, Komuro I, 2003. Roles of cardiac transcription factors in cardiac hypertrophy. *Circ Res*, 92(10):1079-1088.
<https://doi.org/10.1161/01.RES.0000072977.86706.23>
- Alexander J, Stainier DYR, 1999. A molecular pathway leading to endoderm formation in zebrafish. *Curr Biol*, 9(20):1147-1157.
[https://doi.org/10.1016/S0960-9822\(00\)80016-0](https://doi.org/10.1016/S0960-9822(00)80016-0)
- Anchel D, Ching RW, Cotton R, et al., 2016. A novel single cell method to identify the genetic composition at a single nuclear body. *Sci Rep*, 6:29191.
<https://doi.org/10.1038/srep29191>
- Ataliotis P, Ivins S, Mohun TJ, et al., 2005. XTbx1 is a transcriptional activator involved in head and pharyngeal arch development in *Xenopus laevis*. *Dev Dyn*, 232(4):979-991.
<https://doi.org/10.1002/dvdy.20276>
- Bassel GW, Glaab E, Marquez J, et al., 2011. Functional network construction in *Arabidopsis* using rule-based machine learning on large-scale data sets. *Plant Cell*, 23(9):3101-3116.
<https://doi.org/10.1105/tpc.111.088153>
- Benson DW, Silberbach GM, Kavanaughmchugh A, et al., 1999. Mutations in the cardiac transcription factor NKX2.5 affect diverse cardiac developmental pathways. *J Clin Invest*, 104(11):1567-1573.
<https://doi.org/10.1172/JCI8154>
- Bishopric NH, 2005. Evolution of the heart from bacteria to man. *Ann NY Acad Sci*, 1047(1):13-29.
<https://doi.org/10.1196/annals.1341.002>
- Cai X, Hagedorn CH, Cullen BR, 2004. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA*, 10(12):1957-1966.

- <https://doi.org/10.1261/rna.7135204>
- Chiavacci E, Dolfi L, Verduci L, et al., 2012. *MicroRNA 218* mediates the effects of *Tbx5a* over-expression on zebrafish heart development. *PLoS ONE*, 7(11):e50536. <https://doi.org/10.1371/journal.pone.0050536>
- Chen K, Rajewsky N, 2007. The evolution of gene regulation by transcription factors and microRNAs. *Nat Rev Genet*, 8(2):93-103. <https://doi.org/10.1038/nrg1990>
- Chen W, Huang J, Yu X, et al., 2015. Identification of microRNAs and their target genes in Alport syndrome using deep sequencing of iPSCs samples. *J Zhejiang Univ-Sci B (Biomed & Biotechnol)*, 16(3):235-250. <https://doi.org/10.1631/jzus.B1400272>
- Esquela-Kerscher A, Slack FJ, 2006. Oncomirs-microRNAs with a role in cancer. *Nat Rev Cancer*, 6(4):259-269. <https://doi.org/10.1038/nrc1840>
- Ghosh TK, Song FF, Packham EA, et al., 2009. Physical interaction between *TBX5* and *MEF2C* is required for early heart development. *Mol Cell Biol*, 29(8):2205-2218. <https://doi.org/10.1128/MCB.01923-08>
- Glickman NS, Yelon D, 2002. Cardiac development in zebrafish: coordination of form and function. *Semin Cell Dev Biol*, 13(6):507-513. <https://doi.org/10.1016/S1084952102001040>
- Harrelson Z, Kelly RG, Goldin SN, et al., 2004. *Tbx2* is essential for patterning the atrioventricular canal and for morphogenesis of the outflow tract during heart development. *Development*, 131(20):5041-5052. <https://doi.org/10.1242/dev.01378>
- Huttenhower C, Haley EM, Hibbs MA, et al., 2009. Exploring the human genome with functional maps. *Genome Res*, 19(6):1093-1106. <https://doi.org/10.1101/gr.082214.108>
- Kikuchi K, Holdway JE, Werdich AA, et al., 2010. Primary contribution to zebrafish heart regeneration by *gata4*⁺ cardiomyocytes. *Nature*, 464(7288):601-605. <https://doi.org/10.1038/nature08804>
- Kloosterman WP, Steiner FA, Berezikov E, et al., 2006. Cloning and expression of new microRNAs from zebrafish. *Nucleic Acids Res*, 34(9):2558-2569. <https://doi.org/10.1093/nar/gkl278>
- Kourmpetis YA, van Dijk AD, van Ham RC, et al., 2011. Genome-wide computational function prediction of *Arabidopsis* proteins by integration of multiple data sources. *Plant Physiol*, 155(1):271-281. <https://doi.org/10.1104/pp.110.162164>
- Lee I, Li Z, Marcotte EM, 2007. An improved, bias-reduced probabilistic functional gene network of baker's yeast, *Saccharomyces cerevisiae*. *PLoS ONE*, 2(10):e988. <https://doi.org/10.1371/journal.pone.0000988>
- Lee I, Lehner B, Crombie C, et al., 2008. A single gene network accurately predicts phenotypic effects of gene perturbation in *Caenorhabditis elegans*. *Nat Genet*, 40(2):181-188. <https://doi.org/10.1038/ng.2007.70>
- Lee I, Ambaru B, Thakkar P, et al., 2010. Rational association of genes with traits using a genome-scale gene network for *Arabidopsis thaliana*. *Nat Biotechnol*, 28(2):149-156. <https://doi.org/10.1038/nbt.1603>
- Lee Y, Kim M, Han J, et al., 2004. MicroRNA genes are transcribed by RNA polymerase II. *EMBO J*, 23(20):4051-4060. <https://doi.org/10.1038/sj.emboj.7600385>
- Lim LP, Lau NC, Garrett-Engle P, et al., 2005. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*, 433(7027):769-773. <https://doi.org/10.1038/nature03315>
- Liu D, Wang YX, HU JY, et al., 2008. The effects of connexin43 down regulation on the development of the embryonic heart and vasculature in zebrafish. *Prog Biochem Biophys*, 35(7):766-771 (in Chinese).
- Ma L, Ma L, Schwartz RJ, et al., 2005. *Bmp2* is essential for cardiac cushion epithelial-mesenchymal transition and myocardial patterning. *Development*, 132(24):5601-5611. <https://doi.org/10.1242/dev.02156>
- Ma S, Gong Q, Bohnert HJ, 2007. An *Arabidopsis* gene network based on the graphical Gaussian model. *Genome Res*, 17(11):1614-1625. <https://doi.org/10.1101/gr.6911207>
- Moorman AFM, Christoffels VM, 2003. Cardiac chamber formation: development, genes, and evolution. *Physiol Rev*, 83(4):1223-1267. <https://doi.org/10.1152/physrev.00006.2003>
- Morton SU, Scherz PJ, Cordes KR, et al., 2008. microRNA-138 modulates cardiac patterning during embryonic development. *Proc Natl Acad Sci USA*, 105(46):17830-17835. <https://doi.org/10.1073/pnas.0804673105>
- Moulton J, 2010. The miR-143-*adducin3* pathway is essential for cardiac chamber morphogenesis. *Development*, 137(11):1887-1896. <https://doi.org/10.1242/dev.050526>
- Nemer M, 2008. Genetic insights into normal and abnormal heart development. *Cardiovasc Pathol*, 17(1):48-54. <https://doi.org/10.1016/j.carpath.2007.06.005>
- Okamura K, Hagen JW, Duan H, et al., 2007. The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. *Cell*, 130(1):89-100. <https://doi.org/10.1016/j.cell.2007.06.028>
- Olson EN, 2006. Gene regulatory networks in the evolution and development of the heart. *Science*, 313(5795):1922-1927. <https://doi.org/10.1126/science.1132292>
- Plageman TF, Yutzey KE, 2004. Differential expression and function of *Tbx5* and *Tbx20* in cardiac development. *J Biol Chem*, 279(18):19026-19034. <https://doi.org/10.1074/jbc.M314041200>
- Prall OWJ, Menon MK, Solloway MJ, et al., 2007. An *Nkx2-5/Bmp2/Smad1* negative feedback loop controls heart progenitor specification and proliferation. *Cell*, 128(5):947-959. <https://doi.org/10.1016/j.cell.2007.01.042>

- Qu X, Jia H, Garrity DM, et al., 2008. NdrG4 is required for normal myocyte proliferation during early cardiac development in zebrafish. *Dev Biol*, 317(2):486-496. <https://doi.org/10.1016/j.ydbio.2008.02.044>
- Reifers F, Walsh EC, Léger S, et al., 2000. Induction and differentiation of the zebrafish heart requires fibroblast growth factor 8 (fgf8/acerebellar). *Development*, 127(2):225-235.
- Romer AS, 1967. Major steps in vertebrate evolution. *Science*, 158(3809):1629-1637. <https://doi.org/10.1126/science.158.3809.1629>
- Schlange T, Andrée B, Arnold HH, et al., 2000. BMP2 is required for early heart development during a distinct time period. *Mech Dev*, 91(1-2):259-270. [https://doi.org/10.1016/S0925-4773\(99\)00311-1](https://doi.org/10.1016/S0925-4773(99)00311-1)
- Schultheiss TM, Burch JB, Lassar AB, 1997. A role for bone morphogenetic proteins in the induction of cardiac myogenesis. *Genes Dev*, 11(4):451-462. <https://doi.org/10.1101/gad.11.4.451>
- Serbedzija GN, Chen JN, Fishman MC, 1998. Regulation in the heart field of zebrafish. *Development*, 125(6):1095-1101.
- Simões-Costa MS, Vasconcelos M, Sampaio AC, et al., 2005. The evolutionary origin of cardiac chambers. *Dev Biol*, 277(1):1-15. <https://doi.org/10.1016/j.ydbio.2004.09.026>
- Stainier DYR, 2001. Zebrafish genetics and vertebrate heart formation. *Nat Rev Genet*, 2(1):39-48. <https://doi.org/10.1038/35047564>
- Stennard FA, Costa MW, Elliott DA, et al., 2003. Cardiac T-box factor Tbx20 directly interacts with Nkx2-5, GATA4, and GATA5 in regulation of gene expression in the developing heart. *Dev Biol*, 262(2):206-224. [https://doi.org/10.1016/S0012-1606\(03\)00385-3](https://doi.org/10.1016/S0012-1606(03)00385-3)
- Yelon D, Horne SA, Stainier DYR, 1999. Restricted expression of cardiac myosin genes reveals regulated aspects of heart tube assembly in zebrafish. *Dev Biol*, 214(1):23-37. <https://doi.org/10.1006/dbio.1999.9406>
- Yelon D, Ticho B, Halpern ME, et al., 2000. The bHLH transcription factor Hand2 plays parallel roles in zebrafish heart and pectoral fin development. *Development*, 127(12):2573-2582.
- Ying SY, Lin SL, 2006. Current perspectives in intronic microRNAs (miRNAs). *J Biomed Sci*, 13(1):5-15. <https://doi.org/10.1007/s11373-005-9036-8>
- Zeng L, Carter AD, Childs SJ, 2009. miR-145 directs intestinal maturation in zebrafish. *Proc Natl Acad Sci USA*, 106(42):17793-17798. <https://doi.org/10.1073/pnas.0903693106>
- Zhang LF, Gui YH, Wang YX, et al., 2010. Effect of Tbx1 knock down on cardiac performance in zebrafish. *Chin Med J*, 123(9):1182-1189. <https://doi.org/10.3760/cma.j.issn.0366-6999.2010.09.015>
- Zhong W, Sternberg PW, 2006. Genome-wide prediction of *C. elegans* genetic interactions. *Science*, 311(3766):1481-1484. <https://doi.org/10.1126/science.1123287>

中文概要

题目: 基因位置可以用于预测基因调控关系

目的: 证明基因位置顺序与基因调控顺序的关系, 说明利用基因位置顺序信息可以作为预测基因调控关系的先验知识。

创新点: 通过研究与斑马鱼心脏发育阶段有关的基因和 miRNA, 利用时空邻域法, 建立了一个基于时间的基因调控网络。通过这个基因调控网络, 我们获得了一些特殊的基因调控通路和基因之间调控的时间顺序。通过研究斑马鱼心脏发育阶段有关的基因在染色体上的位置, 将基因位置顺序与基因调控时间顺序比较, 我们发现了这两个顺序具有逆转现象。这个逆转现象的存在, 可以帮助我们利用基因位置顺序来预测基因调控关系和调控的时间顺序。

方法: 采用时空邻域法研究了斑马鱼心脏发育阶段的基因和 miRNA, 建立基因调控网络 (图 5), 并通过进一步比较, 得到了基因位置顺序和基因调控时间顺序的联系。

结论: 基因位置顺序信息可以用于预测基因调控关系。

关键词: 预测; 基因调控关系; 基因位置顺序信息; 时空邻域法; 斑马鱼心脏发育期