

## An organogenesis network-based comparative transcriptome analysis for understanding early human development *in vivo* and *in vitro*

Fang *et al.*

RESEARCH ARTICLE

Open Access

# An organogenesis network-based comparative transcriptome analysis for understanding early human development *in vivo* and *in vitro*

Hai Fang<sup>1,2,3†</sup>, Wen Jin<sup>1†</sup>, Ying Yang<sup>2,4</sup>, Ying Jin<sup>2,4</sup>, Ji Zhang<sup>1,2\*</sup> and Kankan Wang<sup>1,2\*</sup>

## Abstract

**Background:** Integrated networks hold great promise in a variety of contexts. In a recent study, we have combined expression and interaction data to identify a putative network underlying early human organogenesis that contains two modules, the stemness-relevant module (hStemModule) and the differentiation-relevant module (hDiffModule). However, owing to its hypothetical nature, it remains unclear whether this network allows for comparative transcriptome analysis to advance our understanding of early human development, both *in vivo* and *in vitro*.

**Results:** Based on this integrated network, we here report comparisons with the context-dependent transcriptome data from a variety of sources. By viewing the network and its two modules as gene sets and conducting gene set enrichment analysis, we demonstrate the network's utility as a quantitative monitor of the stem potential *versus* the differentiation potential. During early human organogenesis, the hStemModule reflects the generality of a gradual loss of the stem potential. The hDiffModule indicates the stage-specific differentiation potential and is therefore not suitable for depicting an extended developmental window. Processing of cultured cells of different types further revealed that the hStemModule is a general indicator that distinguishes different cell types in terms of their stem potential. In contrast, the hDiffModule cannot distinguish between differentiated cells of different types but is able to predict differences in the differentiation potential of pluripotent cells of different origins. We also observed a significant positive correlation between each of these two modules and early embryoid bodies (EBs), which are used as *in vitro* differentiation models. Despite this, the network-oriented comparisons showed considerable differences between the developing embryos and the EBs that were cultured *in vitro* over time to try to mimic *in vivo* processes.

**Conclusions:** We strongly recommend the use of these two modules either when pluripotent cell types of different origins are involved or when the comparisons made are constrained to the *in vivo* embryos during early human organogenesis (and an equivalent *in vitro* differentiation models). Network-based comparative transcriptome analysis will contribute to an increase in knowledge about human embryogenesis, particularly when only transcriptome data are currently available. These advances will add an extra dimension to network applications.

**Keywords:** Integrated networks, Human organogenesis, Stemness-relevant module, Differentiation-relevant module, Gene set enrichment analysis, Transcriptome

\* Correspondence: jizhang@sibs.ac.cn; kankanwang@shsmu.edu.cn

† Contributed equally

<sup>1</sup>State Key Laboratory of Medical Genomics, Sino-French Research Center for Life Sciences and Genomics, Ruijin Hospital affiliated to Shanghai Jiao Tong University School of Medicine (SJTU-SM), Ruijin Rd. II, Shanghai 200025, China

Full list of author information is available at the end of the article

## Background

Molecular and genetic interaction networks have proven to be useful in a variety of contexts. They can potentially be used to predict gene functions [1], to predict perturbation phenotypes [2] and genetic modifier loci [3], to identify human disease genes and drug targets [4], to increase the statistical power in human genetics [5,6], and to study pathogen/virus-host crosstalk [7,8], to name just a few examples. Typically, they are constructed through the integration of multiple data sources such as expression data and interaction data [9-12]. The motivations for building such networks include the following: (i) from a biological perspective, genes are assumed to be interconnected into cohesive networks that control a certain biological process and (ii) from a methodological perspective, the integration of multiple layers of information is more likely to identify biologically relevant signals than analysis of either data source alone. Therefore, these integrated networks hold great promise for explaining the control mechanisms that underlie particular physiological and developmental processes.

In humans, embryogenesis is a complex process that consists of several sequential developmental events: fertilization, blastulation, gastrulation, and organogenesis [13]. Although several studies have attempted to understand the molecular networks that control early embryogenesis (the oocyte and preimplantation stages) [14-18], the extent to which these developmental events can be explained by their underlying networks is still unknown. The molecular profiling of human organogenesis is increasingly becoming the focus of considerable research [19-21]. Recently, we have reported the first comprehensive transcriptome analysis of early organogenesis, which ranged from Carnegie stages 9 (S9) to 14 (S14) [20]. Through the in-depth data mining [22-24] and comparisons with mouse embryos [25] and human embryonic stem cells (hESCs) [26-28], we have found sets of genes that are important for the initiation and maintenance of early human organogenesis. With further integration of interaction data [29-34], we have also shown that the coordination of early human organogenesis is probably under the control of a shared molecular network, or a human organogenesis network (hORGNNet; see Additional File 1). Preliminary analysis has revealed that this network contains a stemness-relevant module (hStemModule) and a differentiation-relevant module (hDiffModule). Given the hypothetical nature of this network [19,20], additional research is warranted to further explore its potentials for characterizing early human organogenesis. It also remains unclear whether this network can be extended to describe the other stages of human organogenesis. Because the network is inherently

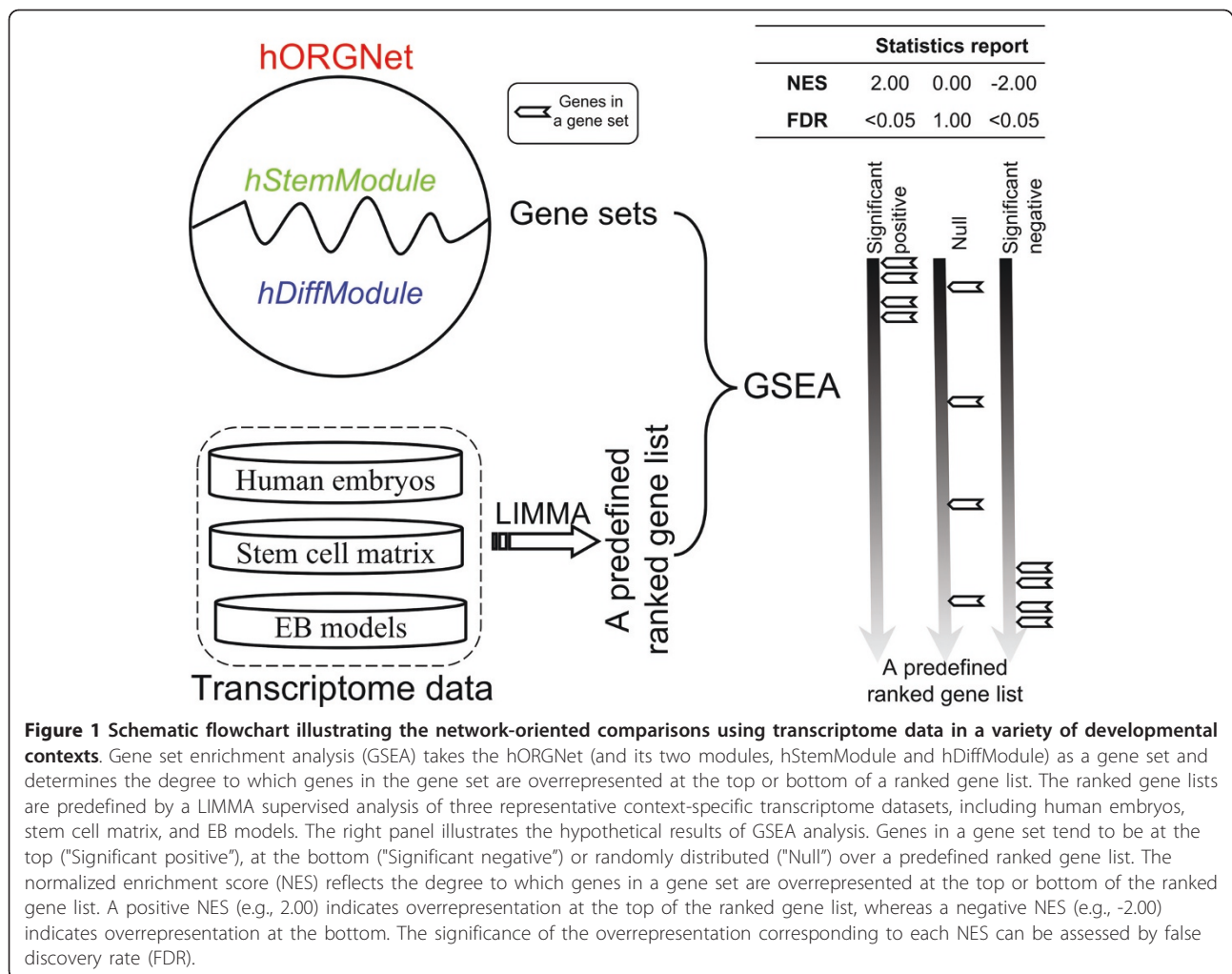
associated with two modules, there is a great need to clarify the circumstances in which it can be used as a reference for evaluating the stem potential *versus* the differentiation potential.

To do this, we started with our previously identified network (i.e., the hORGNNet and its two modules, hStemModule and hDiffModule) [20]. The network itself is associated with the intrinsic features of expression information from early human organogenesis and well-curated interaction information from existing human interactome resources. The genes in this network are collectively informative as a molecular signature of this developmental window, similar to the concept of using disease-perturbed networks as a basis for understanding disease initiation and progression [35]. With this network at hand, we applied gene set enrichment analysis (GSEA) to perform expression-based inspections of the hORGNNet and its two modules in different, yet representative developmental contexts, including human organogenesis, various human stem cell types, and a hESC-derived embryoid body (EB) model. These comparisons demonstrate the ability of this integrated network to improve our coarse-grained understanding of early human development, both *in vivo* and *in vitro*.

## Results

### The comparative transcriptome analysis pipeline using an integrated network during early human organogenesis (hORGNNet)

The procedures for network-orientated comparisons are illustrated in Figure 1. Briefly, transcriptome data from a variety of developmental contexts are available from public databases such as NCBI GEO [36]. In this study, we focused on three representative developmental contexts, including human embryos [20,21], the stem cell matrix (a transcriptome dataset of various human stem cell phenotypes [37]), and EB models [38,39]. By viewing the genes in the hORGNNet collectively as a signature (or gene set) of early human organogenesis, we were able to apply GSEA analysis [40] to explore the possibility of using the hORGNNet to re-interpret these context-specific transcriptome data. To do this, we first ranked the gene lists based on Linear Models for Microarray Data (LIMMA) supervised analysis of these context-specific transcriptome data [41]. Next, we performed GSEA analysis to determine the degree to which genes in the hORGNNet (and its two modules, hStemModule and hDiffModule) were overrepresented at the top or bottom of the ranked list of genes. We used this rank-based comparative approach because it has been proven to be highly reproducible and interpretable [42]. GSEA reports several useful statistics for interpreting the results, including a normalized enrichment score (NES)



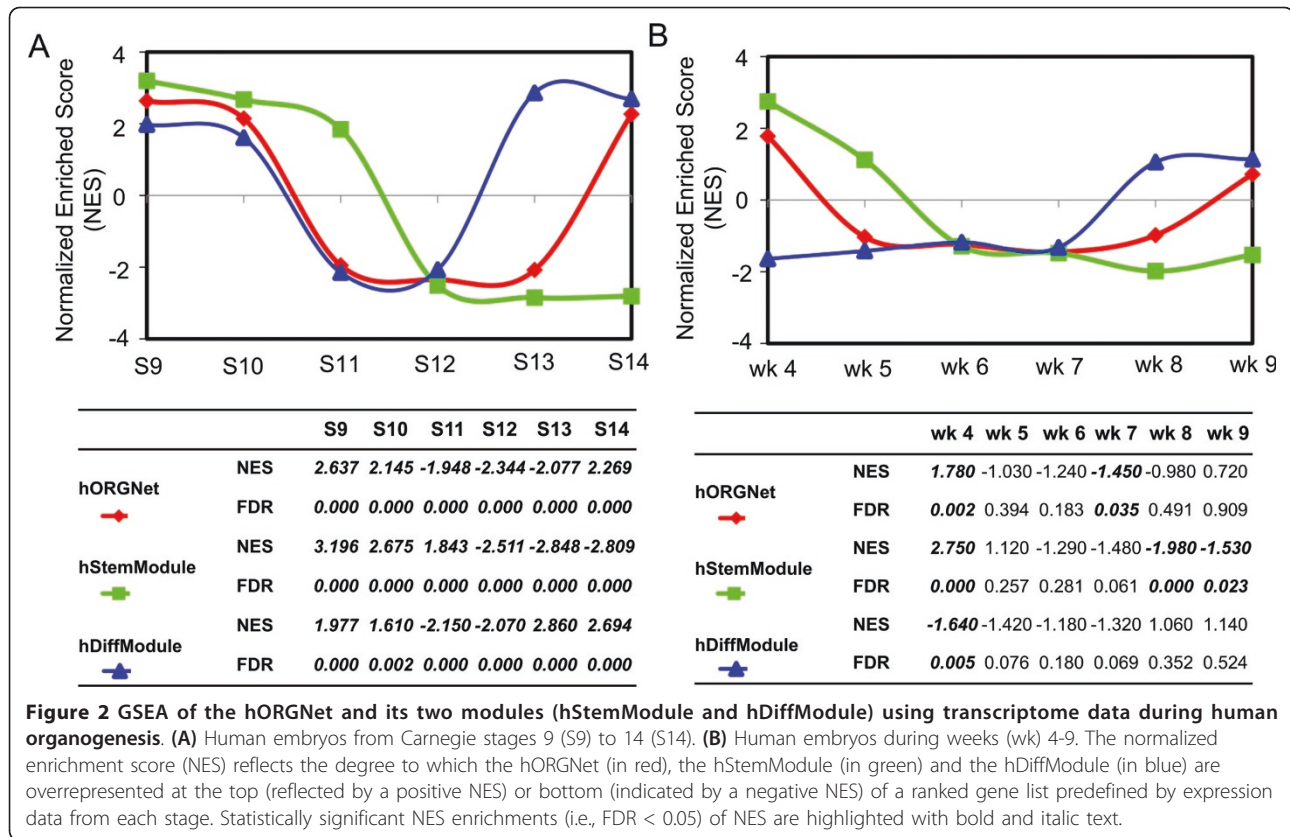
and a false discovery rate (FDR) [40]. The former indicates a positive or negative correlation, while the latter indicates the statistical significance. By analyzing transcriptome data from the stem cell matrix, EB models and human embryos, we found that the hORGNet and its two modules can advance our understanding of early human development, both *in vivo* and *in vitro*.

#### The two modules of the hORGNet capture the expression patterns of early human organogenesis

Previously, we have constructed a hORGNet based on both expression and interaction information [20]. Preliminary analysis indicates that the hORGNet is probably inherited with the Yin-Yang crosstalk of a stemness-relevant module (hStemModule) and another differentiation-relevant module (hDiffModule). Prior to the applications to other developmental processes, we first asked whether the hORGNet and its two modules were associated with the gradual loss of the stem potential and the increased diversity of the differentiation potential during

development. To address this question, we conducted GSEA analysis of the hORGNet and its two modules using transcriptome data of human embryos from Carnegie stages 9 (S9) to 14 (S14) [20]. The GSEA results showed that the hStemModule enrichments monotonically shifted from the most positive at S9 (NES = 3.196; FDR = 0) to the most negative at S14 (NES = -2.809; FDR = 0), whereas the hDiffModule showed more dynamic changes during early human organogenesis (Figure 2A). Recently, another study has reported a transcriptome analysis of human embryos during weeks (wk) 4-9 [21]. As shown in Figure 2B, the GSEA results showed that the hStemModule enrichments decreased gradually from a significant positive correlation at wk 4 (NES = 2.750; FDR = 0) to no significant correlation during wk 5-7 (FDR > 0.05) to a significant negative correlation at wk 8 (NES = -1.980; FDR = 0) and at wk 9 (NES = -1.530; FDR = 0.023). A significant negative correlation between the hDiffModule and wk 4 human embryos was also observed (NES = -1.640; FDR = 0.005). Beyond wk 4 (i.e., out of the





developmental window S9-S14), however, we found there was no significant correlation between the hDiffModule and human embryos (see the bottom panel in Figure 2B). Taken together, analysis of this developmental window (S9-S14) of early human organogenesis and the beyond suggests that the hStemModule may in general reflect the gradual loss of the stem potential, while the hDiffModule reflects the dynamic changes in the differentiation potential that are required for proper differentiation at each stage of this developmental window.

**The hStemModule is a general indicator that distinguishes different cell types in terms of the stem potential**

Stem cell matrix is a database for transcriptome data from various cultured cells including pluripotent, multipotent and differentiated cell types [37] (for details see Methods, and shown in Additional File 2). We first used the stem cell matrix to test whether the hStemModule is indicative of the stem potential in these varied cell types. As shown in Table 1 (also see Additional File 3),

**Table 1 GSEA of the hStemModule for the stem cell matrix.**

Cell types <sup>1</sup>	NES <sup>2</sup>	FDR <sup>3</sup>	Correlation <sup>3</sup>
ePSC	2.132	0.000	Positive**
tPSC	1.899	0.000	Positive**
iPSC	2.658	0.000	Positive**
ePSC_NSC	-1.909	0.000	Negative**
tPSC_Nlin	-0.820	0.888	Null
fNSPC	-2.478	0.000	Negative**
HANSE	-1.843	0.000	Negative**
BM_MSC	-1.754	0.000	Negative**
HUVECS	-2.216	0.000	Negative**
ePSC_EB	2.020	0.000	Positive**

<sup>1</sup>ePSC = embryonic pluripotent stem cells, tPSC = teratocarcinoma pluripotent stem cells, iPSC = induced pluripotent stem cells, ePSC\_NSC = embryonic pluripotent stem cell-derived neural stem cells, tPSC\_Nlin = teratocarcinoma pluripotent stem cells differentiated into dopaminergic neural lineage, fNSPC = fetal neural stem cell or primary fetal neural precursor cells, HANSE = adult surgery neural precursors, BM\_MSC = bone marrow mesenchymal stem cells, HUVECS = umbilical vein endothelial cells, ePSC\_EB = embryonic pluripotent stem cell-derived embryoid bodies; <sup>2</sup> (+) NES for positive correlation, (-) NES for negative correlation; <sup>3</sup>Significance for correlation: \*FDR < 0.05, \*\*FDR < 0.01, and 'Null' for no significant correlation.

the application of GSEA to the stem cell matrix showed that the hStemModule was significantly and positively correlated with pluripotent cells including embryonic pluripotent stem cells (ePSC; NES = 2.132; FDR = 0), teratocarcinoma pluripotent stem cells (tPSC; NES = 1.899; FDR = 0) and induced pluripotent stem cells (iPSC; NES = 2.658; FDR = 0). In contrast, no correlation or negative correlation was observed with other multipotent and differentiated cell types (Table 1). Furthermore, the iPSCs were more likely to be associated with the hStemModule than either the ePSCs or the tPSCs (Table 2 and Additional File 4; see Discussion). More surprisingly, the hStemModule was also positively correlated with embryonic pluripotent stem cell-derived embryoid bodies (ePSC\_EB; NES = 2.020; FDR = 0) (Table 1). To exclude the possibility of artifacts associated with the above observation, we chose another set of transcriptome data from an early stage EB (3.5 days) that was derived from two hESC lines (H1 and H9) [38] as an independent validation. The GSEA results again indicated that there was a significant positive correlation between the hStemModule and early EBs: NES = 1.686 and FDR = 0 for H1-derived EBs and NES = 1.667 and FDR = 0 for H9-derived EBs (Table 3 and Additional File 5). These results clearly demonstrate the discriminative power of the stemness-relevant module in distinguishing cultured cell types of various stem potentials.

**The hDiffModule is seemingly able to predict differences in the differentiation potential among pluripotent cells of different origins, but not among differentiated cells of different types**

Next, we used the stem cell matrix to examine whether the hDiffModule could be used to evaluate the differentiation potential among different cell types. Our previous work [20] showed that the hDiffModule is largely composed of differentiation-associated genes that are regulated during early human organogenesis. Because those genes are under-expressed in hESCs (i.e., are part of the consensus differentiation gene list defined in [27]), the hDiffModule is expected to negatively correlate with hESCs. Indeed, we observed a significant negative

**Table 2 GSEA of the hStemModule based on pair-wise comparisons among pluripotent cell types of different origins.**

Pair-wise comparisons <sup>1</sup>	NES <sup>2</sup>	FDR <sup>3</sup>	Correlation <sup>3</sup>
iPSC vs. ePSC	1.752	0.000	Positive**
iPSC vs. tPSC	1.519	0.003	Positive**
tPSC vs. ePSC	-0.518	1.000	Null

<sup>1</sup>ePSC = embryonic pluripotent stem cells, tPSC = teratocarcinoma pluripotent stem cells, iPSC = induced pluripotent stem cells; <sup>2</sup> (+) NES for positive correlation, (-) NES for negative correlation; <sup>3</sup>Significance for correlation: \*\*FDR < 0.01, and 'Null' for no significant association.

**Table 3 GSEA of the hStemModule for ESC (H1 and H9)-derived EBs.**

Cell types <sup>1</sup>	NES <sup>2</sup>	FDR <sup>3</sup>	Correlation <sup>3</sup>
H1_ESC	1.389	0.010	Positive*
H1_EB	1.686	0.000	Positive**
H9_ESC	1.966	0.000	Positive**
H9_EB	1.667	0.000	Positive**

<sup>1</sup>H1\_ESC = embryonic stem cell line H1, H1\_EB = H1-derived embryoid bodies, H9\_ESC = embryonic stem cell line H9, H9\_EB = H9-derived embryoid bodies; <sup>2</sup> (+) NES for positive correlation, (-) NES for negative correlation; <sup>3</sup>Significance for correlation: \*FDR < 0.05, \*\*FDR < 0.01.

correlation between this module and both ePSCs (NES = -2.234; FDR = 0) and embryonal carcinomas, or tPSCs (NES = -1.490; FDR = 0), but did not observe a correlation between this module and most of differentiated cell types (Table 4; see Discussion). We unexpectedly found that the DiffModule was positively correlated with the iPSCs, the pluripotent cells of non-embryonic origins (NES = 1.373; FDR = 0.029). Consistent with this result, we also found a significant positive correlation between the hDiffModule and iPSCs vs. ePSCs (NES = 2.434; FDR = 0) and iPSCs vs. tPSCs (NES = 1.847; FDR = 0) (Table 5; see Discussion). Similar to the hStemModule, the hDiffModule was also positively correlated with ePSC\_EB (NES = 2.793; FDR = 0) (Table 4); this observation was repeated with a separate dataset (Table 6). Notably, *in vitro* EB differentiation models consistently showed a positive correlation with the hDiffModule, the hStemModule, and the hORGNet made up by these two modules (Additional Files 3 and 5). This suggests the possibility of further characterizing relationships between the developing embryo and the *in vitro* differentiation models that are intended to mimic *in vivo* events.

**The hORGNet-based characterization of relationships between early human organogenesis *in vivo* and hESC-derived EBs *in vitro***

To further explore the usefulness of the hORGNet (and its two modules) in characterizing relationships between

**Table 4 GSEA of the hDiffModule for the stem cell matrix**

Cell types <sup>1</sup>	NES <sup>2</sup>	FDR <sup>3</sup>	Correlation <sup>3</sup>
ePSC	-2.234	0.000	Negative**
tPSC	-1.490	0.009	Negative**
iPSC	1.373	0.029	Positive*
ePSC_NSC	1.232	0.092	Null
tPSC_Nlin	1.194	0.255	Null
fNSPC	-1.768	0.000	Negative**
HANSE	-1.170	0.111	Null
BM_MSC	1.204	0.130	Null
HUVECS	-1.307	0.041	Negative*
ePSC_EB	2.793	0.000	Positive**

<sup>1,2,3</sup>The same as in Table 1.

**Table 5 GSEA of the hDiffModule based on pair-wise comparisons among pluripotent cell types of different origins.**

Pair-wise comparisons <sup>1</sup>	NES <sup>2</sup>	FDR <sup>3</sup>	Correlation <sup>3</sup>
iPSC vs. ePSC	2.434	0.000	Positive**
iPSC vs. tPSC	1.847	0.000	Positive**
tPSC vs. ePSC	1.085	0.556	Null

<sup>1,2,3</sup>The same as in Table 2.

early human organogenesis *in vivo* (S19-S14) and the EB models *in vitro*, we used a time-course of transcriptome data from SHhES1-derived EBs at days 8, 13 and 18 [39] to perform GSEA of the hORGNNet and its two modules (Figure 3A). First, we found that the 8-day EB was comparable to S11 (Additional File 6); both were positively correlated with the hStemModule and negatively correlated with the hDiffModule. This is consistent with the timing of the *in vitro* model, which mimics complex *in vivo* events. Second, a positive correlation with the hDiffModule was observed for the 13- and 18-day EBs, which probably reflects the sustained differentiation *in vitro* (the bottom in Figure 3A). The hStemModule experiences a shift from the positive correlation seen in the 8-day EB (NES = 1.648; FDR = 0) to the negative correlation seen in the 18-day EB (NES = -1.443; FDR = 0.010), indicating the loss of the stem potential (the middle in Figure 3A). Third, the GSEA results with respect to the hORGNNet showed a tendency towards an increased correlation between the hORGNNet and the *in vitro* EB model, partially supporting the idea of sustained differentiation in this *in vitro* EB model (the top in Figure 3A). Therefore, the GSEA analyses suggest a resemblance between the 8-day EB and S11 with regard to both the stem and differentiation potentials, and also suggest that the sustained differentiation *in vitro* in 13- and 18-day EBs could explain their lack of correspondence to any embryonic stage after S11 (see Discussion for details). To vividly display these relationships between the early human organogenesis *in vivo* and this EB model *in vitro*, we performed a principle component analysis (PCA) on the expression matrix of the member genes in the hORGNNet during early human organogenesis and during EB differentiation. As illustrated in Figure 3B, two distinct trajectories were revealed, one

**Table 6 GSEA of the hDiffModule for ESC (H1 and H9)-derived EBs.**

Cell types <sup>1</sup>	NES <sup>2</sup>	FDR <sup>3</sup>	Correlation <sup>3</sup>
H1_ESC	-2.412	0.000	Negative**
H1_EB	1.608	0.000	Positive**
H9_ESC	-2.072	0.000	Negative**
H9_EB	1.833	0.000	Positive**

<sup>1,2,3</sup>The same as in Table 3.

representing the developmental trajectory *in vivo* during early human organogenesis and the other representing the sustained differentiation *in vitro* in the SHhES1-derived EB model. The positions along each of the trajectories probably reflect the developmental nature of the embryos and the sustained, differentiating nature of the EB model, respectively. These two different trajectories clearly show considerable differences between the developing embryos *in vivo* and the EB cultures over time *in vitro*.

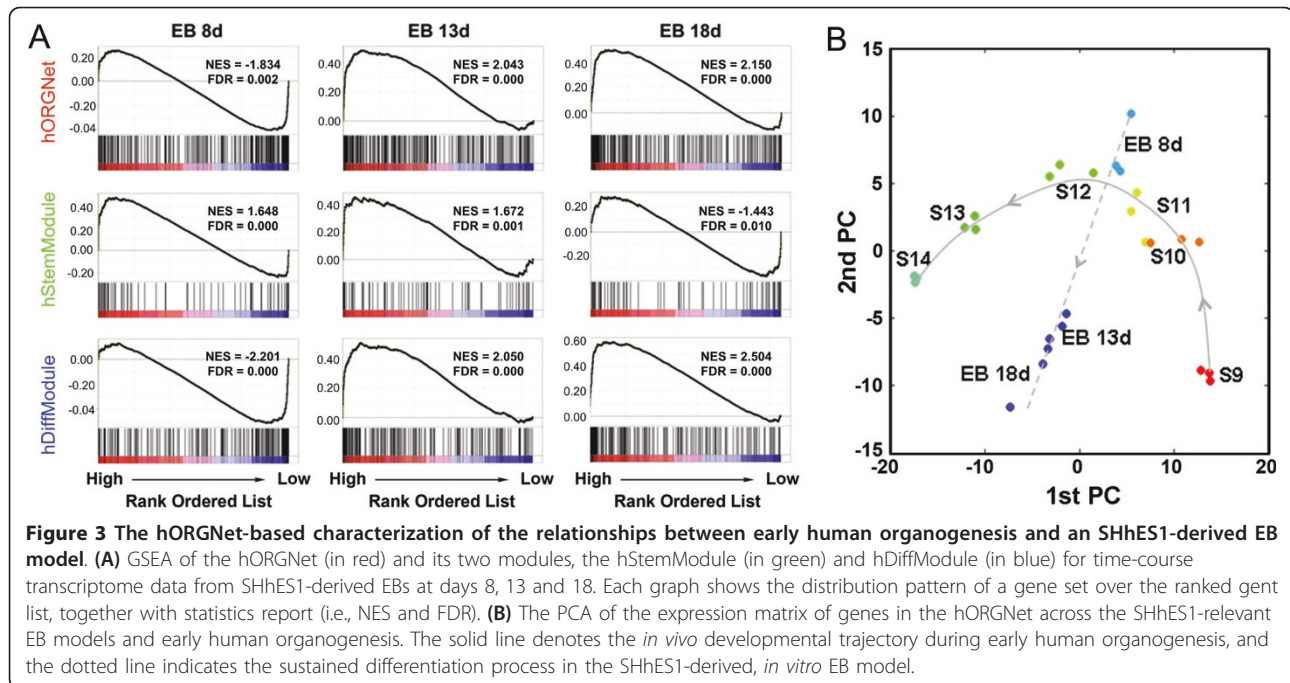
## Discussion

**The different behaviors of the two modules are consistent with the very nature of human embryogenesis**  
 During embryogenesis in humans, early embryonic cells progressively confine their lineage commitment by changing their developmental potential, i.e., their ability to develop into multiple distinct cell types [13]. Cell fate specification during development was first described by C. Waddington as the epigenetic landscape [43], which has recently gained popularity with the advent of cell reprogramming [44-47]. In addition to the Waddington landscape, embryonic cells being committed to descendants can also be viewed as a result of a Yin-Yang-like crosstalk between two key aspects of the developmental potential: the stem potential and the differentiation potential. The former is assumed to maintain the stemness properties, while the latter is crucial for specifying the proper differentiation. Together, they act together in harmony to ensure the successful implementation of embryogenesis. The two modules (i.e., the hStemModule and the hDiffModule) in the hORGNNet might meet the needs of these two respective roles. The GSEA analysis in Figure 2 shows that the expression patterns of genes in these two modules correlate well with our current knowledge about the development potential of early embryonic cells: a gradual loss of stemness and a concomitant diversity of cell types. Therefore, it is logical to speculate that the monotonous behavior of the hStemModule is the necessary outcome of the gradual loss of the stemness during the embryogenesis. The dynamic changes of the hDiffModule, however, could be required for proper differentiation in a stage-specific and context-dependent manner. The different behaviors of these two modules are the biological basis of their utility, as demonstrated in this study, for distinguishing various cell types (Tables 1, 2, 3, 4, 5 and 6) and characterizing the relationships between embryogenesis *in vivo* and differentiation models *in vitro* (Figure 3).

## The two modules differ in their power to distinguish cultured cells of different types

Choosing a single, yet comprehensive transcriptome dataset (called the stem cell matrix [37]) of various





cultured cells as an input may reduce the potential technical biases that could otherwise be introduced when using different types of detection methods from different labs. Processing the stem cell matrix revealed the differences in the discriminative power of the two modules. The hStemModule consistently distinguished cell types of various stem potentials (Table 1). In contrast, the hDiffModule appeared to be unable to distinguish differentiated cells of different types (Table 4), although it provided useful clues for pluripotent cells of different origins (for a detailed discussion, see the next subsection). Their differences in discriminative power can be partially explained by the expression patterns of their member genes in hESCs compared to many differentiated cell types [27]. In our previous work [20], we have showed that the hStemModule is enriched for stemness-relevant genes that are consistently over-expressed in hESCs, while the hDiffModule contains differentiation-relevant genes that are consistently under-expressed in hESCs. In other words, we know for sure that the genes in the hStemModule should be consistently expressed in hESCs. However, the genes in the hDiffModule may be expressed in one or more types of differentiated cells that we cannot identify with certainty, and the hDiffModule itself is therefore not informative regarding differentiated cells of different types.

#### Differences and similarities of the two modules in distinguishing pluripotent cells of different origins

Pluripotent cells were first isolated from embryonic sources, such as ePSCs from the inner cell mass of the

blastocyst [48] and tPSCs from embryonal carcinoma [49]. As a new source of pluripotent cells, iPSCs are generated from non-pluripotent cells (typically somatic cells) that are genetically reprogrammed to an ePSC-like state [50-53]. Initially, iPSCs were thought to be quite similar to their embryonic counterparts, but recent studies have suggested substantial differences between them at both the gene expression [54] and the epigenetic levels [55]. In this study, we showed that the hStemModule positively correlates with all types of pluripotent cells (Table 1), whereas the hDiffModule is negatively correlated with the pluripotent cells of embryonic origins other than iPSCs (Table 4). This difference may be meaningful. In terms of the stem potential, all pluripotent cells should share the characteristics of pluripotency. However, with regard to the differentiation potential, the pluripotent cells of embryonic origins completely repress the expression of differentiation-associated genes, while iPSCs derived from the differentiated cells may inevitably retain an imprint from their origins despite being reprogrammed to a fully pluripotent state. Apart from this difference, we also observed the similarities between the two modules when comparing pluripotent cells of different origins. Compared with the pluripotent cells of embryonic origins (ePSCs and tPSCs), iPSCs showed a significant positive correlation with each of the two modules (Tables 2 and 4). Although the exact implication remains unclear, this may reflect the unique nature of iPSCs; their stemness- and differentiation-contexts may be more similar to those of the hStemModule and the hDiffModule than



those of pluripotent cells of the embryonic origins. Future studies will clarify these observations in a wet experimental setting.

#### Recommended circumstances for using the two modules

In addition to the above situation involving pluripotent cells of different origins, we suggest that the following *a priori* knowledge will be indispensable for using two modules as a monitor of the stem potential *versus* differentiation potential, especially for the hDiffModule (and the hORGNet). Owing to the stage-specific and context-dependent nature of the hDiffModule (Figure 2), it is only valid when the comparisons are constrained to the S9-S14 developmental window or the equivalent *in vitro* differentiation processes (such as the EB models in Figure 3) that recapitulate the *in vivo* cues of this developmental window. The stage-specific nature of the hDiffModule (and thus the hORGNet) does not necessarily mean that it is unsuitable for characterizing the S9-S14 or the equivalent models. Genes in the hDiffModule show both reduced and increased expression patterns from S9 to S14 (see in our previous study [20]), and the positive or negative correlation from the GSEA analysis implies the extent of expression changes required for proper differentiation at each stage. The stage-specific expression profiles of the hDiffModule-containing genes are indicative of each stage, and their correlations (no matter being positive or negative) all have statistical significance as shown in Figure 2A. Together with the hStemModule, the hDiffModule gives the hORGNet as a specific signature for each stage during early human organogenesis from S9 to S14. Additionally, it raises the possibility of using the hORGNet to understand relationships between early human organogenesis *in vivo* and EB models *in vitro*.

#### Implications of the differences between the *in vivo* and *in vitro* developmental trajectories captured by the hORGNet

The observations of (i) the resemblance between 8-day EB and S11 and (ii) the divergence of 13- and 18-day EBs away from the subsequent developmental stages (i. e., S12-S14), have several implications. First, hESCs differentiated *in vitro* into EBs can mimic events that occur *in vivo* both before and after the embryonic implantation [56,57], even extending to S11 at least in terms of the stem and differentiation potential. Second, the prolonged differentiation in culture raises concerns over the limitations of *in vitro* EB models. The EB at days 13 and 18 tended to be positively correlated with the hDiffModule, which is probably due to the sustained expression of the same subsets of differentiation-associated genes that make up the hDiffModule. However, in *in vivo* embryos, different subsets of differentiation-

associated genes from the hDiffModule were expressed at each of the different stages, even though the overall correlations between the hDiffModule and stages S13/S14 also remained significantly positive. This selective expression of hDiffModule genes in the developing embryos *in vivo* and the sustained expression of hDiffModule genes in EBs cultured *in vitro* could explain the two different trajectories as revealed in Figure 3B. Finally, the two overlapping, yet different trajectories observed in this study will warrant the use of this integrated network and its two modules in future studies on human embryogenesis, both *in vivo* and *in vitro*. This network will be particularly useful for those studies that focus on evaluating the stem potential *versus* the differentiation potential.

#### Conclusions

Using a previously proposed integrated network (hORGNet) and its two modules, the stemness-relevant module (hStemModule) and the differentiation-relevant module (hDiffModule), we illustrate its utility by analyzing transcriptome data from a wide variety of developmental contexts (Figure 1). This analysis provides new insights into early human development, both *in vivo* and *in vitro* (Figures 2 and 3; Tables 1, 2, 3, 4, 5 and 6). These advances add an additional dimension to network applications. We strongly recommend the use of this network and its two modules for the circumstances (i) when pluripotent cell types of different origins are involved and (ii) when the comparisons are constrained to the *in vivo* embryos during early human organogenesis or to the equivalent *in vitro* differentiation processes. As the transcriptome data coverage for human embryos improves, we anticipate that even more precise relationships will be revealed using similar network-based comparative transcriptome analyses.

#### Methods

##### A putative molecular interaction network during early human organogenesis

In our recent work [20], we performed a transcriptome analysis of human embryos from Carnegie stages 9 (S9) to 14 (S14), which covers the first third of organogenesis. Further integration of this expression data with interaction information allowed us to identify a putative molecular interaction network that coordinates early human organogenesis (termed hORGNet). A preliminary analysis revealed that the hORGNet is composed of two relatively distinct modules, a stemness-relevant module (hStemModule) and a differentiation-relevant module (hDiffModule). Here, we further evaluate the utility of this hypothetical network and its two modules for characterizing the stem potential *versus* the differentiation potential in various developmental contexts (see below).

### Sources of transcriptome data from a variety of developmental contexts

The stem cell matrix [37] was obtained from NCBI GEO (GSE11508). It contains transcriptome data from the cultured stem cells in the context of a wide variety of pluripotent, multipotent and differentiated cell types. Based on the published cluster results of core dataset samples (further restricted by sample information, such as source tissue, cell type, differentiation state and lineage of the cells), 136 out of the 219 samples were extracted and annotated as belonging to one of 10 clusters. Samples within each cluster displayed similar expression profiles as revealed by a component plane presentation integrated self-organizing map (CPP-SOM) [58,59]. These clusters, each associated with biological- and profile-similar characteristics, included embryonic pluripotent stem cells (ePSC), induced pluripotent stem cells (iPSC), teratocarcinoma pluripotent stem cells (tPSC), embryonic pluripotent stem cell-derived neural stem cells (ePSC\_NSC), teratocarcinoma pluripotent stem cells differentiated into dopaminergic neural lineage (tPSC\_Nlin), fetal neural stem cell or primary fetal neural precursor cells (fNSPC), adult surgery neural precursors (HANSE), bone marrow mesenchymal stem cells (BM\_MSC), umbilical vein endothelial cells (HUVECS) and embryonic pluripotent stem cell-derived embryoid bodies (ePSC\_EBs).

Transcriptome data for early stage EBs (3.5 days) derived from two human ESC lines (H1 and H9) were obtained from a published study [38], and a time course transcriptome dataset from the SHhES1-derived EBs at days 8, 13 and 18 was obtained from a previously published report [39]. Two genome-wide expression datasets for human embryos at six successive time periods (days 20-32) [20] and at six interval-longer time points (weeks 4-9) [21] were obtained from NCBI GEO using the accession numbers GSE1887 and GSE15744, respectively.

### Gene set enrichment analysis (GSEA) of the hORGNet and its two modules

GSEA [60] is a computational method for determining whether an *a priori* defined set of genes (e.g., those genes in the hORGNet) shows statistically significant, concordant differences between two biological states (e.g., one embryonic stage compared to the average of all human embryo stages). We used GSEAPreranked to determine the degree to which genes in the hORGNet (and its two modules, hStemModule and hDiffModule) were overrepresented at the top or bottom of a predefined list of ranked genes. The ranked lists of genes were predefined according to transcriptome data sources as mentioned above in the previous subsection. These gene lists were ranked by means of LIMMA supervised

analysis [41], which uses linear models and empirical Bayes methods to assess differential expression. GSEA calculates an enrichment score (ES) to reflect the enrichment of a gene set at the top (a positive ES) or bottom (a negative ES) of a ranked list of genes. Accounting for differences in the gene set size, GSEA also reports a normalized enrichment score (NES) for comparing results across different gene sets. The significance of the enrichment associated with each NES can be assessed by estimating the false discovery rate (FDR [61]). The detailed explanations for these GSEA statistics can be found in the original paper [40].

For the stem cell matrix, LIMMA supervised analysis was used to determine the ranked gene lists between the assigned cluster and the remaining clusters (or between the assigned cluster and another cluster), followed by GSEA of the hStemModule and the hDiffModule. Similar analysis was also applied to transcriptome data of H1/H9-derived EBs. Regarding human embryos at stages S9-S14, SHhES1-derived EB at days 8, 13 and 18, and human embryos at weeks 4-9, LIMMA was applied to predefine the ranked gene lists between each time point against the average of all time points. GSEA results (i.e., NES and FDR) are detailed in Additional Files 3, 4, 5 and 6. An FDR of 0.05 or lower was accepted as indicating statistical significance for NES (positive or negative).

### Additional material

**Additional file 1: The content of hORGNet.** This table lists genes contained in the hORGNet and its two modules, a stemness-relevant module (hStemModule) and a differentiation-relevant module (hDiffModule).

**Additional file 2: CPP-SOM of the stem cell matrix.** Out of the 219 samples in the stem cell matrix, 136 were extracted according to published cluster results and sample information (e.g., source tissue, cell type, differentiation state and lineage of the cells). They were grouped into 10 clusters, each associated with biological- and profile-similar characteristics. The transcriptome profiles are visualized by Component plane presentation integrated self-organizing map (CPP-SOM). Each presentation illustrates a sample-specific transcriptome map, in which all of the up-regulated (represented by neurons in red), down-regulated (represented by neurons in blue) and moderately regulated (represented by neurons in yellow and green) genes are well delineated. All the presentations are linked by positions. The colours bar stands for expression values (log ratio with base 2), with brighter colours denoting the higher values.

**Additional file 3: GSEA using the stem cell matrix.** GSEA of the hORGNet and its two modules (hStemModule and hDiffModule) using transcriptome data from the stem cell matrix.

**Additional file 4: GSEA for comparing pluripotent cells of different origins.** GSEA of the hORGNet and its two modules (hStemModule and hDiffModule) was performed for pair-wise comparisons between embryonic pluripotent stem cells (ePSC), teratocarcinoma pluripotent stem cells (tPSC), and induced pluripotent stem cells (iPSC). Notably, when compared to ePSCs and tPSCs, iPSCs are more likely to be associated with the hStemModule (in terms of the stemness potential) and the hDiffModule (in terms of the differentiation potential), respectively.

**Additional file 5: GSEA using transcriptome data of early stage EBs.**

GSEA of the hORGN and its two modules (hStemModule and hDiffModule) was performed using transcriptome data of early stage EBs (3.5 days) derived from two human ESC lines (H1 and H9). Notably, GSEA results indicated significant positive correlations between the hStemModule/hDiffModule and early EBs.

**Additional file 6: Comparisons of GSEA results.** GSEA results with *in vivo* early human organogenesis (S11-S14) (A) and *in vitro* EB model (8 d, 13 d and 18 d) (B) were compared. Based on NES profiles, 8-day EB is matched to the S11 (framed in pink), which is consistent with the timing of this *in vitro* model that mimics complex *in vivo* events. The expression-based positive correlation between the hDiffModule and 13-day (and 18-day) EB probably reflects the *in vitro* sustained differentiation of the *in vivo* S11, which is further inferred from the tendency toward increased correlation between the hORGN and the *in vitro* EB model.

### Abbreviations

hORGN: human organogenesis network; hStemModule: stemness-relevant module; hDiffModule: differentiation-relevant module; hESCs: human embryonic stem cells; EBs: embryoid bodies; GSEA: gene set enrichment analysis; NES: normalized enrichment score; FDR: false discovery rate; LIMMA: linear models for microarray data; PCA: principle component analysis; CPP-SOM: component plane presentation integrated self-organizing map; ePSC: embryonic pluripotent stem cells; iPSC: induced pluripotent stem cells; tPSC: teratocarcinoma pluripotent stem cells; ePSC\_NSC: embryonic pluripotent stem cell-derived neural stem cells; tPSC\_Nlin: teratocarcinoma pluripotent stem cells differentiated into dopaminergic neural lineage; fNSPC: fetal neural stem cell or primary fetal neural precursor cells; HANSE: adult surgery neural precursors; BM\_MSC: bone marrow mesenchymal stem cells; HUVECS: umbilical vein endothelial cells; ePSC\_EB: embryonic pluripotent stem cell-derived embryoid bodies.

### Acknowledgements

This work was supported in part by the Knowledge Innovation Program of Chinese Academy of Sciences (KSCX1-YW-22-01), Ministry of Science and Technology of China Grants (2009CB825607 and 2011CB910202), National Natural Science Foundation Grants (30730033 and 90919059), Shanghai Postdoctoral Scientific Program (09R21414900), China Postdoctoral Science Foundation (20090450573) and European Community Grants of FP7 (TB-VIR network, 200973). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Author details

<sup>1</sup>State Key Laboratory of Medical Genomics, Sino-French Research Center for Life Sciences and Genomics, Ruijin Hospital affiliated to Shanghai Jiao Tong University School of Medicine (SJTU-SM), Ruijin Rd. II, Shanghai 200025, China. <sup>2</sup>Key Laboratory of Stem Cell Biology, Institute of Health Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences and SJTU-SM, 225 South Chongqing Road, Shanghai 200025, China. <sup>3</sup>Department of Computer Science, University of Bristol, Woodland Road, Bristol BS8 1UB, UK. <sup>4</sup>Shanghai Stem Cell Institute, 225 South Chongqing Road, Shanghai 200025, China.

### Authors' contributions

HF conceived and designed the study, carried out the data analysis and interpretation, and drafted and revised the manuscript. WJ contributed to the data analysis and revised the manuscript. YY contributed to materials and participated in the design of the study. YJ contributed to materials and helped the data interpretation. JZ conceived the study, participated in its coordination and helped to draft the manuscript. KKW designed and coordinated the study, interpreted the results, and drafted and revised the manuscript. All authors read and approved the final manuscript.

### Competing interests

The authors have declared that no competing interests exist.

Received: 4 April 2011 Accepted: 6 July 2011 Published: 6 July 2011

### References

1. Sharan R, Ulitsky I, Shamir R: **Network-based prediction of protein function.** *Mol Syst Biol* 2007, **3**:88.
2. Lee I, Lehner B, Crombie C, Wong W, Fraser AG, Marcotte EM: **A single gene network accurately predicts phenotypic effects of gene perturbation in *Caenorhabditis elegans*.** *Nat Genet* 2008, **40**(2):181-188.
3. Lee I, Lehner B, Vavouri T, Shin J, Fraser AG, Marcotte EM: **Predicting genetic modifier loci using functional gene networks.** *Genome Res* 2010, **20**(8):1143-1153.
4. Barabasi AL, Gulbahce N, Loscalzo J: **Network medicine: a network-based approach to human disease.** *Nat Rev Genet* 2011, **12**(1):56-68.
5. Pattin KA, Moore JH: **Role for protein-protein interaction databases in human genetics.** *Expert Rev Proteomics* 2009, **6**(6):647-659.
6. Emily M, Mailund T, Hein J, Schauser L, Schierup MH: **Using biological networks to search for interacting loci in genome-wide association studies.** *Eur J Hum Genet* 2009, **17**(10):1231-1240.
7. Kumar D, Nath L, Kamal MA, Varshney A, Jain A, Singh S, Rao KV: **Genome-wide analysis of the host intracellular network that regulates survival of *Mycobacterium tuberculosis*.** *Cell* 2010, **140**(5):731-743.
8. Navratil V, de Chasse B, Combe CR, Lotteau V: **When the human viral infectome and disease networks collide: towards a systems biology platform for the aetiology of human diseases.** *BMC Syst Biol* 2011, **5**:13.
9. Ideker T, Ozier O, Schwikowski B, Siegel AF: **Discovering regulatory and signalling circuits in molecular interaction networks.** *Bioinformatics* 2002, **18**(Suppl 1):S233-240.
10. Ulitsky I, Shamir R: **Identification of functional modules using network topology and high-throughput data.** *BMC Syst Biol* 2007, **1**:8.
11. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, Hanspers K, Isserlin R, Kelley R, Killcoyne S, Lotia S, Maere S, Morris J, Ono K, Pavlovic V, Pico AR, Vailaya A, Wang PL, Adler A, Conklin BR, Hood L, Kuiper M, Sander C, Schmulevich I, Schwikowski B, Warner GJ, et al: **Integration of biological networks and gene expression data using Cytoscape.** *Nat Protoc* 2007, **2**(10):2366-2382.
12. Lin CC, Hsiang JT, Wu CY, Oyang YJ, Juan HF, Huang HC: **Dynamic functional modules in co-expressed protein interaction networks of dilated cardiomyopathy.** *BMC Syst Biol* 2010, **4**:138.
13. Carlson BM: **Human embryology and developmental biology.** St. Louis: Mosby; 3 2004.
14. Dobson AT, Raja R, Abeyta MJ, Taylor T, Shen S, Haqq C, Pera RA: **The unique transcriptome through day 3 of human preimplantation development.** *Hum Mol Genet* 2004, **13**(14):1461-1470.
15. Li SS, Liu YH, Tseng CN, Singh S: **Analysis of gene expression in single human oocytes and preimplantation embryos.** *Biochem Biophys Res Commun* 2006, **340**(1):48-53.
16. Wells D, Patrizio P: **Gene expression profiling of human oocytes at different maturational stages and after *in vitro* maturation.** *Am J Obstet Gynecol* 2008, **198**(4):e451-459, discussion 455 e459-411.
17. Zhang P, Zucchelli M, Bruce S, Hambiliki F, Stavreus-Evers A, Levkov L, Skottman H, Kerkela E, Kere J, Hovatta O: **Transcriptome profiling of human pre-implantation development.** *PLoS One* 2009, **4**(11):e7844.
18. Assou S, Boumela I, Haouzi D, Anahory T, Dechaud H, De Vos J, Hamamah S: **Dynamic changes in gene expression during human early embryo development: from fundamental aspects to clinical applications.** *Hum Reprod Update* 2011, **17**(2):272-290.
19. Geng X, Oliver G: **Elucidating the molecular characteristics of organogenesis in human embryos.** *Genome Biol* 2010, **11**(8):130.
20. Fang H, Yang Y, Li C, Fu S, Yang Z, Jin G, Wang K, Zhang J, Jin Y: **Transcriptome analysis of early organogenesis in human embryos.** *Dev Cell* 2010, **19**(1):174-184.
21. Yi H, Xue L, Guo MX, Ma J, Zeng Y, Wang W, Cai JY, Hu HM, Shu HB, Shi YB, Li WX: **Gene expression atlas for human embryogenesis.** *FASEB J* 2010, **24**(9):3341-3350.
22. Wang K, Fang H, Xiao D, Zhu X, He M, Pan X, Shi J, Zhang H, Xia X, Du Y, Zhang J: **Converting redox signaling to apoptotic activities by stress-responsive regulators HSF1 and NRF2 in fenretinide treated cancer cells.** *PLoS One* 2009, **4**(10):e7538.
23. Fang H, Wang K, Zhang J: **Transcriptome and proteome analyses of drug interactions with natural products.** *Curr Drug Metab* 2008, **9**(10):1038-1048.



24. Fang H, Du Y, Xia L, Li J, Zhang J, Wang K: **A Topology-Preserving Selection and Clustering Approach to Multidimensional Biological Data.** *OMICS* 2011.
25. Mitiku N, Baker JC: **Genomic analysis of gastrulation and organogenesis in the mouse.** *Dev Cell* 2007, **13**(6):897-907.
26. Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, Gifford DK, Melton DA, Jaenisch R, Young RA: **Core transcriptional regulatory circuitry in human embryonic stem cells.** *Cell* 2005, **122**(6):947-956.
27. Assou S, Le Carrouer T, Tondeur S, Strom S, Gabelle A, Marty S, Nadal L, Pantesco V, Reme T, Hugnot JP, Gasca S, Hovatta O, Hamamah S, Klein B, De Vos J: **A meta-analysis of human embryonic stem cells transcriptome integrated into a web-based expression atlas.** *Stem Cells* 2007, **25**(4):961-973.
28. Pan G, Tian S, Nie J, Yang C, Ruotti V, Wei H, Jonsdottir GA, Stewart R, Thomson JA: **Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells.** *Cell Stem Cell* 2007, **1**(3):299-312.
29. Bader GD, Donaldson I, Wolting C, Ouellette BF, Pawson T, Hogue CW: **BIND-The Biomolecular Interaction Network Database.** *Nucleic Acids Res* 2001, **29**(1):242-245.
30. Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D: **The Database of Interacting Proteins: 2004 update.** *Nucleic Acids Res* 2004, **32** Database: D449-451.
31. Kerrien S, Alam-Faruque Y, Aranda B, Bancarz I, Bridge A, Derow C, Dimmer E, Feuermann M, Friedrichsen A, Huntley R, Kohler C, Khadake J, Leroy C, Liban A, Lieftink C, Montecchi-Palazzi L, Orchard S, Risse J, Robbe K, Roehert B, Thorneycroft D, Zhang Y, Apweiler R, Hermjakob H: **IntAct-open source resource for molecular interaction data.** *Nucleic Acids Res* 2007, **35** Database: D561-565.
32. Mishra GR, Suresh M, Kumaran K, Kannabiran N, Suresh S, Bala P, Shivakumar K, Anuradha N, Reddy R, Raghavan TM, Menon S, Hanumanthu G, Gupta M, Upendran S, Gupta S, Mahesh M, Jacob B, Mathew P, Chatterjee P, Arun KS, Sharma S, Chandrika KN, Deshpande N, Palvankar K, Raghavnath R, Krishnakanth R, Karathia H, Rekha B, Nayak R, Vishnupriya G, et al: **Human protein reference database-2006 update.** *Nucleic Acids Res* 2006, **34** Database: D411-414.
33. Joshi-Tope G, Gillespie M, Vastrik I, D'Eustachio P, Schmidt E, de Bono B, Jassal B, Gopinath GR, Wu GR, Matthews L, Lewis S, Birney E, Stein L: **Reactome: a knowledgebase of biological pathways.** *Nucleic Acids Res* 2005, **33** Database: D428-432.
34. von Mering C, Jensen LJ, Kuhn M, Chaffron S, Doerks T, Kruger B, Snel B, Bork P: **STRING 7-recent developments in the integration and prediction of protein interactions.** *Nucleic Acids Res* 2007, **35** Database: D358-362.
35. Hood L, Heath JR, Phelps ME, Lin B: **Systems biology and new technologies enable predictive and preventative medicine.** *Science* 2004, **306**(5696):640-643.
36. Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, Kim IF, Soboleva A, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Muetterter RN, Edgar R: **NCBI GEO: archive for high-throughput functional genomic data.** *Nucleic Acids Res* 2009, **37** Database: D885-890.
37. Muller FJ, Laurent LC, Kostka D, Ulitsky I, Williams R, Lu C, Park IH, Rao MS, Shamir R, Schwartz PH, Schmidt NO, Loring JF: **Regulatory networks define phenotypic classes of human stem cell lines.** *Nature* 2008, **455**(7211):401-405.
38. Lu SJ, Hipp JA, Feng Q, Hipp JD, Lanza R, Atala A: **GeneChip analysis of human embryonic stem cell differentiation into hemangioblasts: an in silico dissection of mixed phenotypes.** *Genome Biol* 2007, **8**(11):R240.
39. Sun BW, Yang AC, Feng Y, Sun YJ, Zhu Y, Zhang Y, Jiang H, Li CL, Gao FR, Zhang ZH, Wang WC, Kong XY, Jin G, Fu SJ, Jin Y: **Temporal and parental-specific expression of imprinted genes in a newly derived Chinese human embryonic stem cell line and embryoid bodies.** *Hum Mol Genet* 2006, **15**(1):65-75.
40. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** *Proc Natl Acad Sci USA* 2005, **102**(43):15545-15550.
41. Smyth GK: **Linear models and empirical bayes methods for assessing differential expression in microarray experiments.** *Stat Appl Genet Mol Biol* 2004, **3**:Article 3.
42. Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP: **GSEA-P: a desktop application for Gene Set Enrichment Analysis.** *Bioinformatics* 2007, **23**(23):3251-3253.
43. Waddington CH: **The strategy of the genes: a discussion of some aspects of theoretical biology.** Allen, Unwin; 1957.
44. Graf T, Enver T: **Forcing cells to change lineages.** *Nature* 2009, **462**(7273):587-594.
45. Macarthur BD, Ma'ayan A, Lemischka IR: **Systems biology of stem cell fate and cellular reprogramming.** *Nat Rev Mol Cell Biol* 2009, **10**(10):672-681.
46. Bhattacharya S, Zhang Q, Andersen ME: **A deterministic map of Waddington's epigenetic landscape for cell fate specification.** *BMC Syst Biol* 2011, **5**:85.
47. Wang J, Zhang K, Xu L, Wang E: **Quantifying the Waddington landscape and biological paths for development and differentiation.** *Proc Natl Acad Sci USA* 2011, **108**(20):8257-8262.
48. Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM: **Embryonic stem cell lines derived from human blastocysts.** *Science* 1998, **282**(5391):1145-1147.
49. Pera MF, Cooper S, Mills J, Parrington JM: **Isolation and characterization of a multipotent clone of human embryonal carcinoma cells.** *Differentiation* 1989, **42**(1):10-23.
50. Takahashi K, Yamanaka S: **Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.** *Cell* 2006, **126**(4):663-676.
51. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S: **Induction of pluripotent stem cells from adult human fibroblasts by defined factors.** *Cell* 2007, **131**(5):861-872.
52. Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R, Thomson JA: **Induced pluripotent stem cell lines derived from human somatic cells.** *Science* 2007, **318**(5858):1917-1920.
53. Park IH, Zhao R, West JA, Yabuuchi A, Huo H, Ince TA, Lerou PH, Lensch MW, Daley GQ: **Reprogramming of human somatic cells to pluripotency with defined factors.** *Nature* 2008, **451**(7175):141-146.
54. Chin MH, Mason MJ, Xie W, Volinia S, Singer M, Peterson C, Ambartsumyan G, Aimiwu O, Richter L, Zhang J, Khvorostov I, Ott V, Grunstein M, Lavon N, Benvenisty N, Croce CM, Clark AT, Baxter T, Pyle AD, Teitell MA, Pelegriani M, Plath K, Lowry WE: **Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures.** *Cell Stem Cell* 2009, **5**(1):111-123.
55. Doi A, Park IH, Wen B, Murakami P, Arjee MJ, Irizarry R, Herb B, Ladd-Acosta C, Rho J, Loewer S, Miller J, Schlaeger T, Daley GQ, Feinberg AP: **Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts.** *Nat Genet* 2009, **41**(12):1350-1353.
56. Poirier F, Chan CT, Timmons PM, Robertson EJ, Evans MJ, Rigby PW: **The murine H19 gene is activated during embryonic stem cell differentiation in vitro and at the time of implantation in the developing embryo.** *Development* 1991, **113**(4):1105-1114.
57. Doetschman TC, Eistetter H, Katz M, Schmidt W, Kemler R: **The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium.** *J Embryol Exp Morphol* 1985, **87**:27-45.
58. Xiao L, Wang K, Teng Y, Zhang J: **Component plane presentation integrated self-organizing map for microarray data analysis.** *FEBS Lett* 2003, **538**(1-3):117-124.
59. Fang H: **Topology-Preserving Selection and Clustering (TPSC).** [http://www.cs.bris.ac.uk/~hfang/TPSC/index.html].
60. **Gene Set Enrichment Analysis (GSEA).** [http://www.broadinstitute.org/gsea/index.jsp].
61. Benjamini Y, Hochberg Y: **Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing.** *Journal of the Royal Statistical Society Series B-Methodological* 1995, **57**(1):289-300.

doi:10.1186/1752-0509-5-108

**Cite this article as:** Fang et al.: An organogenesis network-based comparative transcriptome analysis for understanding early human development *in vivo* and *in vitro*. *BMC Systems Biology* 2011 **5**:108.