Gene Regulation & Transcriptomics

Association of tissue lineage and gene expression: conservatively and differentially expressed genes define common and special functions of tissues

Yao Yu[1,2]*, Tao Xu[3]*, Xuan Li[1,4], Yongtao Yu[4], Yixue Li[1,3,4§], Pei Hao[1,4§]

[1] Key Lab of Systems Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

[2]Graduate School of the Chinese Academy of Sciences, Shanghai 200031, PR China

[3]College of life science and biotechnology, Shanghai Jiaotong University, Shanghai 200240, China

[4]Shanghai Center for Bioinformation Technology, 100 Qinzhou Road, Shanghai 200235, China

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXX

ABSTRACT

Embryogenesis is the process by which the embryo is formed and develops; and establishes developmental hierarchies of tissues. The tissue specific expression pattern of genes indicate important clue of tissue specific functions and cell differentiation features. We performed individual gene and gene set based analysis on multiple tissue expression data, in association with the classic topology of embryogenesis fate maps of human and mouse. On each sub-group from fate map, conservatively expressed, differentially expressed and correlated expressed genes (gene sets) were identified. Tissue distance was correlated with gene expression divergence. Tissues from the same segment on fate map share more similar expression pattern than those from different origins. This phenomenon is more significant in ectoderm tissues than in mesoderm and endoderm segments. Conservatively expressed genes (gene sets) define common functions in a tissue group and are related with tissue specific diseases, supported by results from Gene Ontology and KEGG pathway analysis. Gene divergence is higher in some human tissues than in the mouse homologous tissues. The association study of tissue lineage and gene expression was presented: the common function features of neighbor tissue group were defined by conservatively expressed genes; the differentially expressed genes contribute to the functional divergence. In addition a lower degree of divergence of certain tissue (group) might also be resulted from later cell differentiation and tissue maturity in embryogenesis. The difference of divergence in human and mouse homologous tissues reflects the species complexity, e.g. a higher neural development level and larger body size in human.

1 INTRODUCTION

To understand the expression divergence of multiple tissues is a fundamental issue for investigating the organism complexity (Gu and Su, 2007). Currently, the advance in microarray technology has provided a huge amount of quantitative data of tissue housekeeping and/or tissue specific gene expression in many species (Su, et al., 2004). Some studies have focused on the relationship among tissue specificity, expression conservation, expression level, and sequence conservation (Liao and Zhang, 2006; Liao and Zhang, 2006). Others paid attention to housekeeping, tissue-specific gene and tissue-specific transcriptional regulation for revealing molecular fundamental of tissue function and their evolutionary characters (Avila-Poveda, et al., 2009; Dezso, et al., 2008; Odom, et al., 2007). A gene set (Gene Ontology) based analysis (Su, et al., 2007) proposed a "tissue-driven" hypothesis which showed the relationship between the stabilizing constraints on tissue-specific gene expression and individual GO categories.

Embryogenesis is the process by which the embryo is formed and develops. It describes the developmental history from the single-celled zygote to the multi-celled adult (Salipante and Horwitz, 2006; Salipante, et al., 2008). The construction of *Caenorhabditis elegans* cell fate map (Sulston, et al., 1983) traces the sequence of cell division, migration, and apoptosis of each of the 671 cells. Fate maps of mammals (Eloy-Trinquet, et al., 2000) were proved necessary to establish developmental hierarchies of tissues. For example, during the early stages of embryonic development, the brain starts to form in three distinct segments: the prosencephalon, mesencephalon, and rhombencephalon. The rhombencephalon is the most caudal (toward the tail) segment of the embryonic brain; it is from this segment that the cerebellum develops (Muller and O'Rahilly, 1990). The gonads, which in males are the testes and in females are the ovaries, develop from Intermediate mesoderm - a type of mesoderm that is located between the paraxial mesoderm and the lateral plate (Avila-Poveda, et al., 2009; Ren, et al., 2008). The mammalian kidney also develops from intermediate mesoderm.

Combining gene expression profiling in multiple tissues and classic embryogenesis fate map, we investigated mammalian tissue development in molecular level. In the study we introduced indi-

*These authors contributed equally to this work.

§To whom correspondence should be addressed.

vidual gene and gene set based approaches (Li, et al., 2008) for evaluating expression similarity and divergence. To provide a valuable resource for the in-depth understanding of tissue development and tissue specific disease, we created a gene and gene set expression map along with embryogenesis. Here we present a study on the association of tissue lineage and gene expression, showing that 1) tissue developing divergence among different branches of embryogenesis tree is reflected by expression similarity of genes and gene sets; 2) due to the functional variation of tissues from different ancient segment, the degree of gene and gene set conservation varies among those developing branches in fate map. By comparing the results from human and mouse tissues, we further investigated the inter-species divergence of expression pattern in tissue development.

2 METHODS

2.1 Data description and preprocess

The human (HG-U133A and GNFIH) and mouse (GNF1M) Affymetrix microarray data (Affymetrix, Snata Clara, CA) (Su, et al., 2004) were retrieved from Gene Expression Omnibus (GEO) of National Center for Biotechnology Information (NCBI) (Edgar, et al., 2002) (Human: GDS594 and GDS596; Mouse: GDS 592). The data resource contains the expression data from 73 human tissues (GDS594/GDS596) and 69 mouse tissues (GDS592). Only common homology tissues from human and mouse were considered in the following analysis (24 orthologous adult tissue listed in Table S1). And for those biological replicates we assigned the average expression value as the signal. After mapping probes to the microarray chip platform annotation files, the final datasets include 14746 human genes and 13048 mouse genes. The homology gene pairs' information in human and mouse was gained from NCBI (http://www.nicbi.nlm.gov/HomoloGene). After extracting the unique hit homology pairs, we identified 5055 orthologous gene pairs have expression data in both human and mouse. We used packages from BioConductor (http://www.bioconductor.org/) for functional annotation of those genes. Package 'org.Hs.eg.db' and 'org.Mm.eg.db' (version 2.2.11) were used for Gene Ontology (GO) mapping of human and mouse genes respectively. Package 'KEGG.db' (version 2.2.11) was used for pathway mapping of genes. Only these GO modules with over five genes on the chip will be used in the following analysis.

2.2 Estimation of gene expression divergence in tissues

A p-rank method (Hao, et al., 2009) was performed to evaluate the gene expression distance between two tissues. A gene's expression data was first ranked among all genes in each tissue, and then divided by the number of genes *n*.

$$p-rank_i = \frac{Rank_i}{n}$$

Let $p_{i,t1}$ and $p_{i,t2}$ indicate the p-rank values of gene *i* in tissues *t1* and *t2*. We calculated the gene divergence E_i for every gene between each two tissues on fate map as:

$$E_i = \left| p_{i,t1} - p_{i,t2} \right|$$

Tissue conservatively expressed genes were defined as those genes with the least expression divergence between two tissue. We arbitrarily defined those genes whose divergence was in the lowest 5% of all gene divergence values to be conserved ones in each pair of tissues. The conserved genes in a group of tissues were the set of common conserved gene in all combinations of tissue pair in that group. In this study, we defined those genes which were expressed as the top 5% in all the tissues as ubiquitously expressed housekeeping genes.

2.3 Gene set based measurement for inter tissue expression similarity

Consider a set of n gene in one Gene Ontology (GO); three kinds of measurements were defined for characterizing the profiling of this GO among the tissues. If expression distance was significantly small for a given GO module in a pair of tissues, the expression level of this module was considered conservative in the tissue pair. If a significant *p* value was obtained in the Kolmogorov–Smirnov test (KS test) (Goodman, 1954) for the expression of a given GO module in a pair of tissues, the expression distribution of this module was referred to as a divergent pattern in these tissues. If Pearson correlation coefficient was significant for a given GO module in a pair of tissues, their expression relationship was referred to as a linear correlated pattern (Wang, et al., 2008).

2.3.1 Tissue expression distance

Let $S_{i,g,t}$ denote the (log2 transformed) expression levels of gene *i* of GO *g* in tissue *t*. The expression distance of GO *g* in tissue *t1* and tissue *t2* is calculated as:

$$D_{g,t-t'} = \sum_{i=1}^{n} (S_{i,g,t1} - S_{i,g,t2})^2 / n$$

, while *n* is the number of gene in GO *g*.

The *D* values in all tissue pairs in human and mouse follow a normal distribution (see in Supplementary). Based on the cumulative distribution function of the normal distribution, a *p* value was calculated to represent the significance of the expression similarity of GO *g* in tissue *t1* and tissue *t2*:

$$p_{g,t1-t2} = \Pr(X \le x) = \phi_{\mu,\sigma^2}(x) = \int_{-\infty}^{x} \varphi_{\mu,\sigma^2}(u)du$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-\frac{(u-\mu)^2}{2\sigma^2})du$$

$$x = D_{g,t1-t2}$$

, μ for the mean of *D* values and σ for the standard deviation of *D* values.

For a sub-group of tissues in fate map's branch, the intro group expression similarity of GO *g* is estimated using an integrated p value (Yu, et al., 2009). Since $D \sim N(\mu, \sigma^2)$, D_{sub} (the mean value of *D* values in a sub-group of tissues) should follow $D_{sub} \sim N(\mu, \sigma^2 / N)$, *N* for the sample size in this group. Thus the integrated p value could be calculated as:

$$p_{g,t-sub} = \Pr(X \le x) = \phi_{\mu,\sigma^2/N}(x), x = D_{g,t1-t2}$$

The GO module with a small p value or integrated p value was considered as conserved GO module in a pair or a group of tissues.

2.3.2 Tissue expression difference

To identify differentially expressed GO modules in a pair of tissues, a nonparametric KS test was performed. The p-rank values of the genes in certain GO module from a pair of tissues were used. A *p* value (denoted as $p_{g,t1-t2}$) less than 0.05 denoted that the difference of expression status of this GO module in two tissues should be considered as statistically significant.

For a group of tissues, the *p* values from the KS test in each pair of tissues from that group were integrated. The *p* values $p_{g,ti-tj}$ were transformed

to $Z_{g,ti\text{-}tj}$ with quantile function of normal distribution. Then Z score for a sub-group of tissues (Z_{sub}) was summarized from $Z_{g,ti\text{-}tj}$ with the function:

$$Z = \frac{sum(Z_{g,ti-tj})}{\sqrt{n}}$$

, n for the number of pair wise p values in this group

If these GO modules were not signature modules of a group of tissues, the $p_{g,ti\text{-}tj}$ would follow uniform distribution. If $p_{g,ti\text{-}tj}$ followed uniform distribution, $Z_{g,ti\text{-}tj}$ would follow norm distribution. As a result, Z score would also follow norm distribution. A significant small value of Z comparing to normal distribution corresponded to the significantly being perturbed of the GO modules under these tissues.

2.3.3 Tissue expression correlation

The expression correlation of GO g in a pair of tissues $t1$ and $t2$ was defined as the Pearson's correlation coefficient of the profiles of genes in GO g in tissue $t1$ and in tissue $t2$:

$$r_{g,t1-t2} = \frac{n\sum_{i=1}^{n}(S_{i,g,t1} \times S_{i,g,t2}) - \sum_{i=1}^{n} S_{i,g,t1} \times \sum_{i=1}^{n} S_{i,g,t2}}{\sqrt{n\sum_{i=1}^{n} S_{i,g,t1}^2 - (\sum_{i=1}^{n} S_{i,g,t1})^2} \times \sqrt{n\sum_{i=1}^{n} S_{i,g,t2}^2 - (\sum_{i=1}^{n} S_{i,g,t2})^2}}$$

, while n is the number of gene in GO g.

A high r indicates a high similarity of the expression profiles of the GO module in two tissues, thus this GO module was considered as correlated expressed GO module in tissues. The threshold for r between two tissues was set to be $r > 0.9$, while the threshold for r in a group of tissues was set to be $r_{mean} > 0.9$.

2.4 Gene enrichment analysis

For all the genes defined as tissue conserved ones, we made gene module enrichment using in GO modules and KEGG pathways. The enrichment significance of tissue conserved genes was calculated in a hyper genomic distribution model. Let k be the number of sequences corresponding to genes of interest in certain GO module (or KEGG pathway). The total number of sequences in genes of interest is n. Given the total size of genes in the libraries (m) and in all tissues (N), the probability of observing k or more sequences for gene g in liver can be calculated by the formula:

$$P_{liver}(g) = \sum_{x>k}^{n} \Pr(K = x)$$

$$\Pr(K = k) = f(k; N, m, n) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}$$

The Bonferroni correction was used to adjust the p-values for liver enrichment gene identification based on the hypergeometric distribution.

3 RESULTS

3.1 Estimation of gene expression divergence in tissues

Based on p-rank transferred value, individual gene expression divergence (see in Methods) in different tissue pairs was estimated (Table S2). Our results showed the expression divergence varies greatly in different tissue pairs. The divergence of the neighbor tissues on the fate map is less significant than randomly picked pairs. In addition, the degree of the difference in gene expression profile from two tissues is positively correlated with their distance on the fate map (Fig. 1). According to both t-test and KS test, the genes' expression distance within neighbor tissues were significantly smaller than that from relatively distant tissues on the fate map. This phenomenon is especially clear in the neural system from ectoderm and in tissues from endoderm, e.g. the distribution of gene expression distances between amygdale and other tissues (Fig, 1E). In the comparison pair of amygdale and its closest neighbor prefrontal cortex, more genes were identified as conservatively expressed ones than in other tissue comparison pairs. Furthermore, the lowest peak from the pairs of amygdale and other none neural tissue represented a higher expression divergence in those tissues. However, in the mesoderm tissues, the conservation of gene expression within sub-group was not that significant (supplementary figure 1). This result suggested that through the process of embryogenesis, tissues inherit some imprinting from their ancestor; however due to the function divergence and developing timing difference less was inherited in mesoderm tissues than in other segments. This phenomenon was observed in both human and mouse (see in Table S2 and S3).

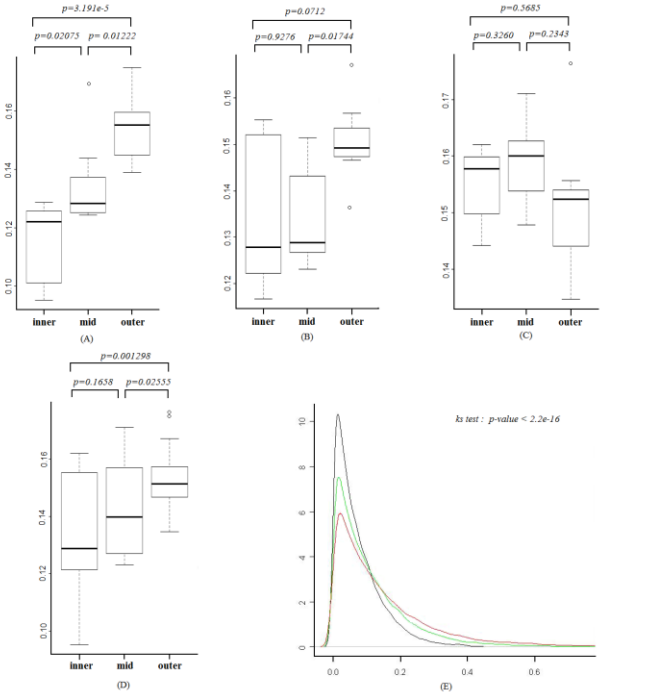

Fig. 1. The distribution of gene expression distance in tissue pairs: (A) for ectoderm tissues, (B) for endoderm tissues, (C) for mesoderm tissues, and (D) for all tissues considered in this study. 'Inner' represents the group of neighbor tissues, defined as tissues belong to the same super-node (red rectangle in Fig. 2). 'Mid' represents the group of tissue pairs from the segments of ectoderm, endoderm and mesoderm, while not in the same 'inner' group. 'Outer' represents the group of tissue pairs from different segments. Student's t-tests were performed on each data pairs of these groups from (A) to (D). (E) The gene expression divergence between amygdale and other tissues. The black curve represents the density of divergence between amygdale and prefrontal cortex. The green curve represents the density of the mean divergence of amygdale and other neural tissues. Red curve represents the mean divergence density between amygdale and other somatic tissues. KS test was performed to investigate the significance of the difference of these curves.

In function enrichment analysis, we found that conservatively expressed genes in the tissue group from the same ancient segments in embryogenesis define some common functions in the tissue group. For example, calcium signaling pathway is vital for transduction of nerve stimulus, which is a basic function for the nerve cells. In human, gene *CALM2*, which is well known as the important phosphorylase kinase in calcium signaling pathway (Mikkers, et al., 2002), was conservatively expressed through all tissues derived from neural tube. Interestingly, in mouse, kinase-coding gene *Camk2g*, which has a similar function, was conservatively expressed in these tissues, too. Such phenomenon was quite explicit when we examined the tissue group of high order on the fate map, such as the group of tissues developed from mesoderm, ectoderm and endoderm, etc. Another example was the relationship between tissue-conserved genes and diseases. Those genes conservatively expressed in the neural tube tissues were significantly enriched in several eurologic diseases related pathway, such as Parkinson's disease and Alzheimer's disease. In the tissue group of pancreas, liver, lung and trachea, the conserved genes were enriched in such pathway as Heparan sulfate biosynthesis, alpha-Linolenic acid metabolism, closely related to the functions these tissues played in human body.

We further analyzed the tissue-conserved genes highly expressed among all tissues, which were hypothesized to be housekeeping genes. In the previous studies (Su, et al., 2004), the housekeeping genes volume changed dramatically as the criteria varied. About 1% of all genes were detected as housekeeping ones by Su (Su, et al., 2004). While in the study of She et al (She, et al., 2009), there were 1523 genes of the same dataset universally expressed. In this study, we used a criterion according to the suggestion of Tu et al (Tu, et al., 2006): the expression intensities of those genes were greater than 550 standard units, and should be observed across all tissues. We identified 392 human (about 2% of all human genes in the dataset) and 443 mouse (about 3% of all mouse data) ubiquitously-expressed genes in our study (Table S4 and S5). These genes appear to play some basic roles in primary cellular machineries, such as the ubiquitin-mediated proteolysis and the ribosome pathway.

3.2 Tissues expression distance in view of Gene Ontology

We calculated the expression distance of GO categories in each pair of tissues. Conservatively expressed gene set was defined as the one with similar expression level in a pair (group) of tissues. According to our methods, conserved GO modules were identified in those sub-groups of tissues on the fate map. Fig. 2 illustrated the number of conserved GO modules on each node of fate map. Generally, the number of conserved GO modules in sub-groups of tissues was decreasing along with the process of development from zygote to adult tissues. There is no conserved GO module detected in mesendoderm and zygote. This is consistent with the fact that tissue developing process is reflected by expression divergence growing in molecular level. In different developing branches, the growing of gene set divergence has large variance. For instance, 201 GO modules had been identified as conservative modules in brain, while 20 GO modules had been identified as conservative ones in primitive gut. These conservative GO modules in brain focus on such functions as the regulation of G-protein coupled receptor protein signaling pathway (GO:0045744), as well as some cell cycling related housekeeping functions (Table S6). The conservative GO modules in gonad (testis and ovary) contains: growth hormone receptor signaling pathway (GO:0060396) and response to growth hormone stimulus (GO:0060416). All these conservative GO modules in different node on fate map have close relationship with the common functions of the sub-group of adult tissues.

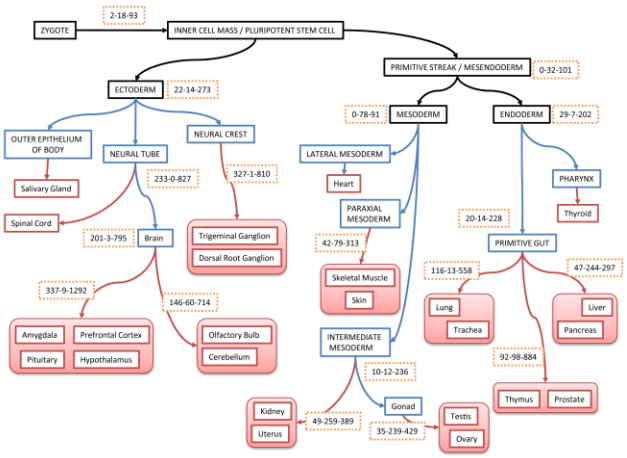

Fig. 2. The embryogenesis fate map of 24 adult tissues considered in this study and gene set based tissue divergence analysis results. The topology of embryo fate map was constructed mainly based on *"Developmental Biology"* (Gilbert, 1994).The red nodes on the tree represent the adult tissues used in microarray experiment. The blue nodes represent the middle stages in developing process. The black nodes represent the primitive organizes of embryo. The numbers in the frame with dotted edges represent the numbers of identified GO modules according to three kinds of measurement: conservative gene sets according to expression distance, differential expressed gene sets according to KS test and correlated gene sets according to Pearson's correlation.

KS tests were performed to detect significantly differentially expressed GO modules, of which the distribution of the gene expression levels varies greatly in different tissues. Tissues from mesoderm and endoderm segments contain more differentially expressed GO modules found by pair wise comparisons of these tissues (Fig. 2). The GO modules for synaptic transmission (GO:0007268, GO:0019226) and transmission of nerve impulse (GO:0045202) were identified as the most three significantly differentially expressed ones in brain, while in the sub-group of brain (prefrontal cortex, amygdale, pituitary, hypothalamus) the expression profiles of these modules remain stable (Table S7). The cerebellum plays an important role in the integration of sensory perception, coordination and motor control. In order to coordinate motor control and response to stimulation, genes in these modules are actively expressed in this region to link the cerebellum with the cerebral motor cortex (which sends information to the muscles causing them to move) and the spinocerebellar tract (which provides proprioceptive feedback on the position of the body in space) (Fine, et al., 2002). The significantly differentially expressed GO modules in ectoderm contain function modules for immune system (GO:0002376, GO:0006955), response to defense (GO:0006952),

Table 1. Gene ontology based tissue expression divergence.

TissueNode	Human (4207 GO modules)			Mouse (2321 GO modules)			Annotation
	GO Distance	KS.test	Correlation	GO Distance	KS.test	Correlation	
Zygote	2	18	93	2	3	94	pluripotent stem cell
Mesendoderm	0	32	101	1	12	102	endoderm & mesoderm
Endoderm	29	7	202	1	29	109	endoderm tissues
PrimitiveGut	20	14	228	3	26	122	primitive gut 1 & 2 & 3
PrimitiveGut3	92	98	884	64	78	535	thymus & prostate
PrimitiveGut2	47	244	297	7	124	115	pancreas & liver
PrimitiveGut1	116	13	558	180	4	1399	trachea & lung
Mesoderm	0	78	91	20	15	222	mesoderm tissues
ParaxialMesoderm	42	79	313	108	18	722	skin & skeletal muscle
IntermediateMesoderm	10	120	236	30	47	255	intermediate
KU	49	259	389	91	30	600	uterus & kidney
Gonad	35	239	429	33	228	235	testis & ovary
Ectoderm	22	14	273	111	0	703	ectoderm tissues
NeuralCrest	327	1	810	437	0	2106	neural crest
NeuralTube	233	0	827	241	0	1587	neural tube
Brain	201	3	795	216	0	1524	brain 1 & 2
Brain2	146	60	714	368	0	1728	cerebellum & olfactory bulb
Brain1	337	9	1292	224	0	1491	prefrontalcortex & amygdala & pituitary & hypothalar

The thresholds for these measurements are: GO Distance: $p < 0.05$ (or integrated $p < 0.05$); KS test: $p < 0.05$ (or integrated $p < 0.05$); Correlation: $r > 0.9$.

response to stimulus (GO:0050896) and etc. The differentially expressed GO modules play the roles in organizing function divergence of tissues.

According to Pearson's correlation coefficient, correlated expressed GO modules were identified. The expression of genes in these modules simultaneously up-regulated or down-regulated (or remain same) in different tissues (Table S8). Neural related tissues contain more correlated GO modules than the tissues developed from mesoderm and endoderm. There were 827 and 810 highly correlated (correlation over 0.9) GO modules in neural tube and neural crest, respectively. However, the correlated expressed modules in intermediate mesoderm and primitive gut were 236 and 228, respectively. These results suggested that the gene expression in non-neural tissues is more divergent than in neural related tissues. Most of the correlated expression genes of neural tissues were enriched into such pathways like signal peptide processing (GO:0006465), G-protein signaling pathway (GO:0007189, GO:0010578 and GO:0010579), BMP signaling pathway regulation (GO:0030510). Comparatively the expression of these GO modules in endoderm and mesoderm tissues showed no significant correlation.

3.3 Tissues expression in human and mouse

Gene set (GO module) based methods for estimating the expression divergence in tissues of fate map were used on the mouse data (Table S9-S11). Compared with the result from the human data, the trends of tissue divergence are similar in the two species (Table 1). Neural related tissues developed from ectoderm were the most conservative tissue groups on the fate map. The GO modules expression distances in this group of tissues in mouse were significantly shorter than those from endoderm and mesoderm tissues. The KS tests showed that no significant differentially expressed GO modules could be identified in ectoderm tissues. The most different tissue sub-group on fate map in expression level is gonad (consist of testis and ovary, with 228 GO modules differentially expressed test by KS test), which is consistent with the result in human (239 differentially expressed GO modules). There were more correlated expressed GO modules in ectoderm tissues than in other tissues on the fate map.

The comparison of the numbers of conservatively and differentially expressed GO modules in human and mouse tissue was illustrated in Fig. 3. Compared with the common size adjusted data, the mouse tissues contain more conserved GO modules and less differentially expressed ones than the corresponding human tissues. Also there exists significant difference in gene set expression of individual tissue. Uterus and kidney (noted as KU) in mouse have more similar expression profiles than their homologous tissues in human. According to expression distance in KU there were more conserved GO modules in mouse (91 GO modules) than that in human (49 ones), while according to KS test there were less differential expressed modules in mouse (30 ones) than that in human (259 ones). In addition, there are much more highly correlated expressed GO modules in tissues in mouse than in human, especially in tissues from ectoderm. Regardless of the fact that there are fewer mouse genes than human genes on chip, which might have some bias for the analysis, our result suggested a less divergence in mouse tissues in molecular level than in human tissues. This phenomenon is more pronounced in the nervous system and the paraxial mesoderm. According to Alexander E., et al (Vinogradov and Anatskaya, 2007), the prominent human–mouse divergence in neural system might reflect both the increase in organism complexity and the transition to the qualitatively higher complexity level: social organization. Besides, the higher cell differentiation level in paraxial mesoderm (such as skeletal muscle) might due to the larger body size.

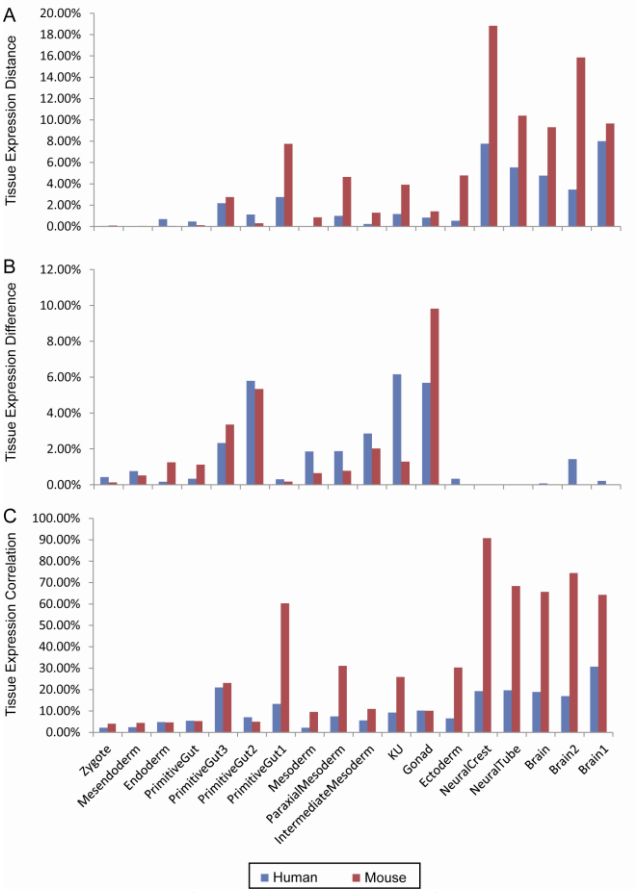

Fig. 3. Bar plots of common size statement of tissue conservative GO modules (A), tissue differentially expressed GO modules (B), and highly correlated expressed GO modules (C) in human and mouse. The names of tissue nodes on the x axis are consistent with those used in Table 1.

4 DISCUSSION

In this study, we mapped human and mouse tissues onto the embryogenesis fate map and further analyzed the expression conservation of tissues with individual gene and gene set (Gene Ontology) based methods. Our results shown that, tissues from the same segment on fate map share more similar expression patterns than those from different origins of embryogenesis. Tissues have inherited some imprinting from their ancester in developing process, which was observed in molecular level in our study. These imprinting which might include methylation, phosphorylation and other epigenetic characters resulted into the conservation of gene express in neighbor tissues and divergence in distinct segments. In addition, the expression divergence of gene and gene set varied greatly in different sub-group of neighbor tissues on the fate map. Neural tissues developed from ectoderm were the most conserved subgroup. The expression level of genes in these tissues is more conservative to each other: more conserved individual genes, more conserved GO modules, less differentially expressed GO modules and more significantly correlated GO modules, compared with tissues developed from mesendoderm (Su, et al., 2007).

In most of the tissues, tissue-specific function is reflected in the stabily expressed genes and GO categories. For example, genes from ubiquitin mediated proteolysis and ribosome pathway show universal conservation in all tissues, and especially their expression levels in sub-group of tissues remain stable. Different tissue groups have their identical expression conserved gene group. *CALM* of human genome or *Camk2g* of mouse genome are conservatively expressed in brain and other tissues developed from neural tube, while neither these genes nor the calcium signaling pathway is detected in the conservation and enrichment analysis of mesendoderm tissues. The construction of tissue specificity is more likely to be governed by those gene sets with specific molecular functions.

The timeline of tissue development and maturity might also be account for this phenomenon. By week 20 of gestation, fetal kidneys and urinary tract start producing urine, releasing it into the amniotic fluid, and the fetal digestive system produces meconium, a black, tar-like substance that are its first few bowel movements. However the maturity of neural system is later than that time. In Rhombic lip, a primary region thought to give rise to the neurons that make up the cerebellum, neurons move to the external granular layer by embryonic week 27. The later differentiation of neural tissues might result in conservative expression status in molecular level. A comparatively larger timeline of developing for mesoderm tissues might be one of the reason why less conservation of gene expression was identified, besides their great divergence of tissue function.

Through the comparison of the tissue gene and gene set expression in human and mouse, the similar trends of function-driven tissue development were identified with certain variance in individual tissues. The group of ectoderm tissues is more conservative in expression level than the group of mesoderm or endoderm tissues. The difference of human and mouse tissue includes: the decreasing divergence of neural tissue group in mouse, denoted as more conservative GO modules in mouse ectoderm tissues; other difference in the internal organs, such as the more conservative of paraxial mesoderm in mouse than that in human. As discussed previously, this might be resulted from high requirement of neural system activity in human social life and extremely large difference in body size of human and mouse.

ACKNOWLEDGEMENTS

Funding: This work is supported by National High-Tech R&D Program of China (863) (grant 2007AA02Z330, 2007AA02Z332), National Key Basic Research Program (Grant No. 30800210) and Shanghai Committee of Science and Technology (grant 07JC14048, 08JC1416600).

Conflict of Interest: none declared.

REFERENCES

Avila-Poveda, O.H., Colin-Flores, R.F. and Rosas, C. (2009) Gonad development during the early life of Octopus maya (Mollusca: Cephalopoda), *Biol Bull*, **216**, 94-102.

Dezso, Z., Nikolsky, Y., Sviridov, E., Shi, W., Serebriyskaya, T., Dosymbekov, D., Bugrim, A., Rakhmatulin, E., Brennan, R.J., Guryanov, A., Li, K., Blake, J., Samaha, R.R. and Nikolskaya, T. (2008) A comprehensive functional analysis of tissue specificity of human gene expression, *BMC Biol*, **6**, 49.

Edgar, R., Domrachev, M. and Lash, A.E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository, *Nucleic Acids Res*, **30**, 207-210.

Eloy-Trinquet, S., Mathis, L. and Nicolas, J.F. (2000) Retrospective tracing of the developmental lineage of the mouse myotome, *Curr Top Dev Biol*, **47**, 33-80.

Fine, E.J., Ionita, C.C. and Lohr, L. (2002) The history of the development of the cerebellar examination, *Semin Neurol*, **22**, 375-384.

Gilbert, S.F. (1994) *Developmental Biology*.

Goodman, L.A. (1954) Kolmogorov-Smirnov tests for psychological research, *Psychol Bull*, **51**, 160-168.

Gu, X. and Su, Z. (2007) Tissue-driven hypothesis of genomic evolution and sequence-expression correlations, *Proc Natl Acad Sci U S A*, **104**, 2779-2784.

Hao, P., Zheng, S., Ping, J., Tu, K., Gieger, C., Wang-Sattler, R., Zhong, Y. and Li, Y. (2009) Human gene expression sensitivity according to large scale meta-analysis, *BMC Bioinformatics*, **10 Suppl 1**, S56.

Li, Y., Hao, P., Zheng, S., Tu, K., Fan, H., Zhu, R., Ding, G., Dong, C., Wang, C., Li, X., Thiesen, H.J., Chen, Y.E., Jiang, H. and Liu, L. (2008) Gene expression module-based chemical function similarity search, *Nucleic Acids Res*, **36**, e137.

Liao, B.Y. and Zhang, J. (2006) Evolutionary conservation of expression profiles between human and mouse orthologous genes, *Mol Biol Evol*, **23**, 530-540.

Liao, B.Y. and Zhang, J. (2006) Low rates of expression profile divergence in highly expressed genes and tissue-specific genes during mammalian evolution, *Mol Biol Evol*, **23**, 1119-1128.

Mikkers, H., Allen, J., Knipscheer, P., Romeijn, L., Hart, A., Vink, E. and Berns, A. (2002) High-throughput retroviral tagging to identify components of specific signaling pathways in cancer, *Nat Genet*, **32**, 153-159.

Muller, F. and O'Rahilly, R. (1990) The human brain at stages 21-23, with particular reference to the cerebral cortical plate and to the development of the cerebellum, *Anat Embryol (Berl)*, **182**, 375-400.

Odom, D.T., Dowell, R.D., Jacobsen, E.S., Gordon, W., Danford, T.W., MacIsaac, K.D., Rolfe, P.A., Conboy, C.M., Gifford, D.K. and Fraenkel, E. (2007) Tissue-specific transcriptional regulation has diverged significantly between human and mouse, *Nat Genet*, **39**, 730-732.

Ren, D., Xing, Y., Lin, M., Wu, Y., Li, K., Li, W., Yang, S., Guo, T., Ren, J., Ma, J., Lan, L. and Huang, L. (2008) Evaluations of Boar Gonad Development, Spermatogenesis with regard to Semen Characteristics, Libido and Serum Testosterone Levels based on large White Duroc x Chinese Erhualian Crossbred Boars, *Reprod Domest Anim*.

Salipante, S.J. and Horwitz, M.S. (2006) Phylogenetic fate mapping, *Proc Natl Acad Sci U S A*, **103**, 5448-5453.

Salipante, S.J., Thompson, J.M. and Horwitz, M.S. (2008) Phylogenetic fate mapping: theoretical and experimental studies applied to the development of mouse fibroblasts, *Genetics*, **178**, 967-977.

She, X., Rohl, C.A., Castle, J.C., Kulkarni, A.V., Johnson, J.M. and Chen, R. (2009) Definition, conservation and epigenetics of housekeeping and tissue-enriched genes, *BMC Genomics*, **10**, 269.

Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., Cooke, M.P., Walker, J.R. and Hogenesch, J.B. (2004) A gene atlas of the mouse and human protein-encoding transcriptomes, *Proc Natl Acad Sci U S A*, **101**, 6062-6067.

Su, Z., Huang, Y. and Gu, X. (2007) Tissue-driven hypothesis with Gene Ontology (GO) analysis, *Ann Biomed Eng*, **35**, 1088-1094.

Sulston, J.E., Schierenberg, E., White, J.G. and Thomson, J.N. (1983) The embryonic cell lineage of the nematode Caenorhabditis elegans, *Dev Biol*, **100**, 64-119.

Tu, Z., Wang, L., Xu, M., Zhou, X., Chen, T. and Sun, F. (2006) Further understanding human disease genes by comparing with housekeeping genes and other genes, *BMC Genomics*, **7**, 31.

Vinogradov, A.E. and Anatskaya, O.V. (2007) Organismal complexity, cell differentiation and gene expression: human over mouse, *Nucleic Acids Res*, **35**, 6350-6356.

Wang, H., Wang, Q., Li, X., Shen, B., Ding, M. and Shen, Z. (2008) Towards patterns tree of gene coexpression in eukaryotic species, *Bioinformatics*, **24**, 1367-1373.

Yu, Y., Tu, K., Zheng, S., Li, Y., Ding, G., Ping, J. and Hao, P. (2009) GEOGLE: context mining tool for the correlation between gene expression and the phenotypic distinction, *BMC Bioinformatics*, **10**, 264.