A cross-species analysis method to analyze animal models’ similarity to human’s disease state

Shuhao Yu¹†, Lulu Zheng⁴, Yun Li², Chunyan Li³, Chenchen Ma², Yixue Li¹,², Xuan Li²*, Pei Hao³,⁵*

From The International Conference on Intelligent Biology and Medicine (ICIBM)
Nashville, TN, USA. 22-24 April 2012

Abstract

Background: Animal models are indispensable tools in studying the cause of human diseases and searching for the treatments. The scientific value of an animal model depends on the accurate mimicry of human diseases. The primary goal of the current study was to develop a cross-species method by using the animal models’ expression data to evaluate the similarity to human diseases’ and assess drug molecules’ efficiency in drug research. Therefore, we hoped to reveal that it is feasible and useful to compare gene expression profiles across species in the studies of pathology, toxicology, drug repositioning, and drug action mechanism.

Results: We developed a cross-species analysis method to analyze animal models’ similarity to human diseases and effectiveness in drug research by utilizing the existing animal gene expression data in the public database, and mined some meaningful information to help drug research, such as potential drug candidates, possible drug repositioning, side effects and analysis in pharmacology. New animal models could be evaluated by our method before they are used in drug discovery.

We applied the method to several cases of known animal model expression profiles and obtained some useful information to help drug research. We found that trichostatin A and some other HDACs could have very similar response across cell lines and species at gene expression level. Mouse hypoxia model could accurately mimic the human hypoxia, while mouse diabetes drug model might have some limitation. The transgenic mouse of Alzheimer was a useful model and we deeply analyzed the biological mechanisms of some drugs in this case. In addition, all the cases could provide some ideas for drug discovery and drug repositioning.

Conclusions: We developed a new cross-species gene expression module comparison method to use animal models’ expression data to analyse the effectiveness of animal models in drug research. Moreover, through data integration, our method could be applied for drug research, such as potential drug candidates, possible drug repositioning, side effects and information about pharmacology.
Background

The use of animal models is essential in the study of many human disorders, especially in the occasions when human patients are inaccessible, or ethical issue prevents using human subjects in such studies. Animal models can greatly reduce the costs of research and thus they are available and affordable to a broad scientific community. Animal models have been proved to be important in the areas of chronic wasting diseases, i.e. Alzheimer [1-3], cancers [4-6], and new drug development [7-11]. A study found that animal models could predict human toxicity in 71% of the cases [12].

However, despite the advantages in employing animal models to study various human diseases, it has still been a challenging task in drug research to test thousands of compounds in animal models for searching a promising candidates. Because important biological differences still exist between animal models and humans that could significantly impair drug discovery [11], although the models could usually recapitulate many of the key features in physiology. For example, mice do not own a true homologue of human interleukin-8 (IL-8), and presumably the function of this cytokine in mice is subsumed by other molecules. Thence, we cannot directly test IL-8 antagonists or agonists in murine systems [11]. In this regard, the scientific value of an animal model depends on how accurately it can mimic the human disease, and an assessment of the animal models’ similarity to human disease state is requisite.

As a dynamic and continuous variable, expression changes with the developmental and physiological states. Furthermore, it is known that a gene’s transcriptional response provides important clues to its function. Therefore, genes’ expression profiles across species can be compared to determine the conservation and divergence of transcription. Microarrays have collected the necessary data to evaluate the transcriptomic fidelity of an animal model in terms of the similarity of expression with the human tissues.

Strand and his colleagues have proved that regional gene expressions of brains between human and mouse were conserved [13]. Miller et al. also undertook a brain-specific comparison of human and mouse transcription profiles [3], and in agreement with Strand’s study, they found that both gene expression and the summation of gene co-expression relationships are generally well conserved. At the same time, they also identified some between-species differences that provided insight into human disease. However, whether orthologous gene pairs have the similar pattern of gene expression across species has been much discussed over the past two decades, but comparative analysis at the transcriptomic level has produced opposite conclusions [14,15]. Building on improved computational methods to correct such opposition, Chan et al. [16] compared multiple tissue-expression datasets across five vertebrate species: human, mouse, chicken, frog and pufferfish, and found the evidence of conserved expression in more than a third of unique orthologous genes. Consistent with Chan et al.’ discovery, Zheng-Bradley et al. confirmed the conservation of gene expression at a greater degree [14,15] by carrying out a large scale comparison of global gene expression patterns in human and mouse. They proved that the global tissue-specific expression patterns of orthologous genes are considerably conserved in mouse and human, and the expression of groups of orthologous genes in each tissue co-varied, in both the tissue-specific gene and the house-keeping gene of two species.

In view of the above information, we proposed a novel approach to assess whether the animal models recapitulate the essential features of human diseases for drug research. The approach was based on the gene expression data of the response of function-known drugs from Connectivity Map (cMap) [17]. cMap had collected many microarrays corresponding to treatment of 164 different small molecules in different human cell lines. By comparing the gene expression signatures of drugs, diseased samples, and mutants, cMap was able to connect compounds, diseases, and genes through gene expression profiles. Considering the similarity of orthologous gene expression profiles across species, we first matched human and other animal species’ genes using gene ortholog information in Roundup database [18], and then applied gene modularization technology to compare gene expression profiles, which was proposed by Li et al. [19]. We expected that this orthologous genes’ similarity could provide a way to explore the ability of animal models to mimic diseases of the human bodies. When the connection of function-known drugs and the disease was established, we were able to infer whether these drugs were the right reagents to the corresponding disease and thus conclude the similarity between animal models and human’s disease state. We also compared this gene modularization method with the distance method used by other researchers on cross-species analysis [14].

By applying the method to animal model expression profiles in several cases, lots of interesting information was obtained for drug research. We found that trichostatin A and some other HDACs could have very similar response across cell lines and species at gene expression level. Mouse hypoxia model could accurately mimic the human hypoxia, while mouse diabetes drug model might have much limitation in drug discovery. What’s more, the transgenic mouse of Alzheimer was also an
available model, and then we deeply analyzed the biological mechanisms of some drugs in this case. In addition, all the cases could provide some ideas for drug discovery and drug repositioning.

Results
Cross-species comparison of drug response at cell level
At first, we tested whether our cross-species method could find the similarity of drug responses across the species. From GEO, we downloaded 7 microarray data of mouse osteoblastic cells (MC3T3-E1) treated by Trichostatin A (TSA, an HDAC inhibitor), including three replicates of TSA treatment and four replicates of control (GEO: GDS3002). After performing one similarity search in the cMap database by our method, the top 10 chemicals with highest scores were presented in Table 1. The result of the distance comparison method was presented in Table 1.

The results of our method and distance comparison method were consistent. TSA itself appeared many times. For the rest, Vorinostat, and HC toxin, in spite of distant structures, were all HDAC inhibitors [20-22]. Data in “GO counts” column (Table 1a) showed that these chemicals were almost fully-positively correlated with the query, consistent with the fact that they performed the similar function.

Although all the cells in cMap database were from human tumor cell lines and the query data were obtained from mouse osteoblastic cells (MC3T3-E1), the result indicated that the expression similarity existed between different cells and species when treated with HDAC inhibitors.

In 2009, Dudley et al. [23] evaluated 429 experiments, representing 238 diseases and 122 tissues from 8435 microarrays, and found evidences of a general, pathophysiological concordance between microarray experiments measuring the same disease in different tissues. Our result showed that microarrays of cell response to drugs which altered the cellular expression pattern could also have similarity across cell lines or species. The consistent result of our method and distance comparison method also hinted that cross-species gene expression analysis was practicable in the field of drug research.

Exploring the effectiveness of mouse models of diseases and their relations with some drug molecules
Our approach could be used to determine whether the mouse model could be applied to preclinical drug screening and to identify potential novel drug or drug repositioning for certain diseases in the database. We tested three separate cases, hypoxia, Diabetes drug and Alzheimer by using gene expression profiles of mouse animal models.

Hypoxia
The response of mouse to hypoxia was derived from a study by Laifenfeld [24] in which mice received decreasing oxygen concentrations from 21% to 6% O2 for 30 minutes. Then, the mice remained at 6% O2 for another 120 minutes and the bone marrows were retrieved from the right humerus. We used 7 microarray assays (GEO: GSE17796) of bone marrow cells to run our test and the results were listed in Table 2a (our method) and Table 2b (distance comparison method).

In Table 2a, nine in ten chemicals were reported to be associated with hypoxia and seven of the nine agents showed fully-positive correlation with the query profiles. Resveratrol was reported to inhibit the accumulation of hypoxia-inducible factor-1alpha and VEGF expression in human tongue squamous cell carcinoma and hepatoma cells [25], which seemed to have a protective mechanism in hypoxia mice. Genistein postconditioning had a protective effect on hypoxia/reoxygenation-induced injury in human gastric epithelial cells [26]. Thioridazine was a member of the class of phenothiazines that act, in part, by inhibiting respiration and lead to hypoxia [27]. Deferoxamine, a chelating agent capable of binding free iron, acted to simulate hypoxia by altering the iron status of hydroxylases [28]. The calmodulin inhibitor, Trifluoperazine, could suppress the hypoxic hyperpolarization [29]. Ionomycin was used to raise the intracellular level of calcium and calpain activity in rat proximal tubules in order to simulate the effects of hypoxia [30]. Sirolimus was an mTOR inhibitor that leads to the inhibition of the Hypoxia-inducible factor activity [31,32].

The remaining two of the nine agents showed negative correlation with the query profiles. Tretinoin (retinoic acid) stimulated erythropoietin gene transcription in

### Table 1 The top 10 instances obtained from gene modularization comparison method and distance comparison method in TSA case

| A. cMap ID | Molecule | GO counts | B. cMap ID | Molecule |
|------------|----------|-----------|------------|----------|
| 1050       | trichostatin A | 56+       | 448        | trichostatin A |
| 1072       | trichostatin A | 55+       | 909        | HC toxin |
| 448        | trichostatin A | 54+       | 1072       | trichostatin A |
| 1014       | trichostatin A | 53+       | 1000       | vorinostat |
| 909        | HC toxin     | 53+       | 1058       | vorinostat |
| 981        | trichostatin A | 52+       | 981        | trichostatin A |
| 1112       | trichostatin A | 50+       | 1050       | trichostatin A |
| 992        | trichostatin A | 50+       | 332        | trichostatin A |
| 331        | trichostatin A | 50+       | 331        | trichostatin A |
| 332        | trichostatin A | 50+       | 1112       | trichostatin A |

The top 10 instances sharing the largest number of significantly affected GOMs with TSA-treated cells are listed in A and the top 10 instances resulted from distance comparison method are listed in B.

For detailed parameter settings: BP is chosen as GO mode, and the P-value for cutting off insignificantly matched or reverse-matched GO modules is 0.05. ‘plus’ indicates the number of GO terms positively correlated; ‘minus’ indicates the number of GO terms negatively correlated.
embryonal carcinoma cells through the direct repeat of a steroid/thyroid hormone receptor response element half-site in the hypoxia-response enhancer element [33]. Clofibrate reduce hypoxia-inducible factor (HIF)-2alpha binding to the hypoxia-response element (HRE) [34]. In Table 2b, besides the molecules as mentioned above (deferoxamine, thioridazine and sirolimus), Haloperidol, Calmidazolium and Wortmannin were also reported to be associated with hypoxia [35-37].

Through these descriptions, we could see that the mouse model of the hypoxia was a good one to be used to observe the mechanism of hypoxia and help to discover drugs aiming to different targets or find side effects of some existing drugs in hypoxia. Moreover, our method could find some molecules (Tretinoin and Clofibrate) negatively correlated to hypoxia and they had a common feature: effect on hypoxia-response element. This result could not be obtained from the distance comparison method.

**Diabetes drug**

It had been reported that the mouse was not a reasonable animal model in the research of diabetes drug, because of its much lower AR expression level than that of human, which was probably insufficient to generate toxic by products [38].

We used our method to test if mouse models were suitable in diabetes drug study. We got microarray assays (GEO: GSE14888) of mouse 3T3-L1 adipocyte tissue cultures fed by metformin, then ran our method and the distance comparison method respectively, and presented the results in Table 3.

| A. cMap ID | Molecule       | GO counts | B. cMap ID | Molecule       |
|------------|----------------|-----------|------------|----------------|
| 595        | resveratrol    | 25+       | 573        | deferoxamine   |
| 1015       | genistein      | 22+       | 1068       | thioridazine   |
| 970        | 5230742        | 21+       | 418        | haloperidol    |
| 1068       | thioridazine   | 20+       | 921        | sirolimus      |
| 573        | deferoxamine   | 20+       | 1080       | sirolimus      |
| 1004       | trifluoperazine| 19+       | 906        | calmidazolium  |
| 991        | tretinoin      | 18-       | 904        | S109870        |
| 871        | 7-ionomycin    | 18+       | 1075       | fluphenazine   |
| 921        | sirolimus      | 17+       | 977        | wortmannin     |
| 444        | clofibrate     | 17-       | 417        | thioridazine   |

The results of our method and the distance comparison method were consistent. Among these molecules, only 15delta prostaglandin J2 had some treatment relations with diabetes. It is a ligand of the adipocyte determination factor PPAR gamma [39,40]. Nevertheless, it was very confusing that Rottlerin was positive with metformin in our result, because Rottlerin could inhibit insulin-stimulated glucose transport in 3T3-L1 adipocytes by uncoupling mitochondrial oxidative phosphorylation [41]. Besides, Zhang et al suggested new possible applications (drug repositioning) for Celastrol, such as diabetes management [42].

Other molecules had no report of any relations with diabetes. Therefore, it was suggested that the mouse and human had some differences in the effect of metformin. However, it was possible to make use of mouse model to do drug research related to 15delta prostaglandin J2, whose target was a nuclear receptor.

**Alzheimer**

Alzheimer disease (AD), the most common form of dementia, is incurable, degenerative and terminal. It has been advised that the mouse was not a good animal model for Alzheimer, because human and mouse’s brain transcriptome had a large divergence in Alzheimer disease pathways [3]. But if the mouse was transgenic, would it become a suitable model?

The animal model we used here was a transgenic (TG) mouse expressing human APP695 and bearing the double Swedish and Indiana amyloid precursor protein (APP) mutations [1]. Six microarray assays (Hippocampus cells from two normal and four APP transgenic mice, GEO: GSE14499) were analyzed using our method and the distance comparison method. Top ten hits were presented (Table 4).

As the table showed, no molecules were found by the distance comparison method to have a treatment on Alzheimer. In contrast, six of the top ten results detected by our method were negatively related to Alzheimer, promising possible therapeutic functions. Nordihydroguaiaretic acid could break down pre-formed Alzheimer’s β-amylod fibrils in vitro [1]. Tretinoin (Retinoid) was relevant to many pathophysiological features of AD, including amyloid plaques, inflammation, immunological changes, cell death and regeneration processes, altered neurotransmission, and age-related changes [43]. It made sense that Nordihydroguaiaretic acid and Tretinoin both had many negative correlation GO modules and could resist AD. Estradiol and alpha-estradiol also prevented AD-associated inflammation with an increasing PPAR gamma expression [44]. Monorden, also known as radicicol, was a natural product binding to Hsp90 (Heat Shock Protein 90) and altering its function, while Hsp90 acted as a regulator of pathogenic changes that led to the neurodegenerative
phenotype in AD [45]. LY-294002 held back the trafficking of APP and APP metabolites by inhibiting phosphatidylinositol 3-kinase (PI3K) [46].

Among the remaining molecules, Prazosin was a non-sedating generic medication used for hypertension and benign prostatic hypertrophy. It antagonizes NE effects at brain postsynaptic alpha-1 adreno receptors and new study said the prazosin improved patients’ behavioural symptoms such as agitation/aggression in AD [47]. Fulvestrant was an interesting drug, known to block estrogen receptors [48]. It could also dissociate HSP90 and trigger its intracellular degradation [49]. Considering the positive connection between fulvestrant and Alzheimer, we could infer that estrogen pathway was more important than HSP90 pathway in AD (details described in the Discussion section). The last molecule, ikarugamycin, had no report of any relation with AD, but we thought it might also have a potential side effect to induce AD because of the positive correlated modules in our result.

Because almost all molecules were related with AD in the result of our method, we thought that the transgenic AD model was a feasible model of AD in humans.

**Discussion**

Since the transgenic animal model of AD was feasible for drug discovery, we further performed an in-depth analysis of the results of the AD case, especially for the three candidates: fulvestrant, alpha-estradiol and monorden. Alpha-estradiol, the predominant sex hormone presented in females, and monorden, a kind of HSP90 inhibitor, were both negatively connected with AD, while fulvestrant, both estrogen blocker and HSP90 inhibitor, showed positive connection with AD (Figure 1).

Researchers had found that the molecular chaperone Hsp90 interacted with unliganded steroid hormone receptors (including estrogen receptor) and regulated their activity [50]. Thus, it is natural to infer that monorden might alleviate AD by acting Hsp90 receptor and regulating estrogen’s pathway. However, according to our result, the exact GO modules of alpha-estradiol was mainly associated with lipid process (GO: 0006610, GO: 0006629), while monorden showed strong connection with sterol process (GO: 0016126, GO: 0016125) (see Additional file 1). The two different processes indicated that Hsp90 should have a second pathway to connect with AD.

Compared with GO categories of alpha-estradiol and monorden, fulvestrant had two additional categories, “phosphoglyceride metabolic process” (GO: 0006650) and “phospholipid biosynthetic process” (GO: 0006854), listed. Based on these, we hypothesize that estrogen might act the phospholipid pathway designed to alleviate AD. Literature mining helped us reveal that estrogen

| Table 3 The top 10 instances obtained from gene modularization comparison method and distance comparison method in Diabetes case |
|---|---|---|
| A. cMap ID | Molecule | GO counts | B. cMap ID | Molecule |
| 1069 | 15-delta prostaglandin J2 | 63+ | 941 | rottlerin |
| 887 | celestrol | 54+ | 839 | rottlerin |
| 893 | pararosaniline | 54+ | 1069 | 15-delta prostaglandin J2 |
| 1011 | 15-delta prostaglandin J2 | 53+ | 978 | pyrvinium |
| 839 | S224221 | 53+ | 956 | S224221 |
| 956 | S224221 | 51+ | 887 | celestrol |
| 941 | rottlerin | 51+ | 893 | pararosaniline |
| 1140 | MG-132 | 46+ | 1140 | MG-132 |
| 1011 | thioridazine, [INN] | 45+ | 1011 | thioridazine, [INN] |
| 976 | S182598 | 44+ | |

The top 10 instances sharing the largest number of significantly affected GOMs with adipocyte tissue fed by metformin are listed in A and the top 10 instances resulted from distance comparison method are listed in B.

For detailed parameter settings: GO mode: BP; P-value: < 0.05.

| Table 4 The top 10 instances obtained from gene modularization comparison method and distance comparison method in Alzheimer case |
|---|---|---|
| A. cMap ID | Molecule | GO counts | B. cMap ID | Molecule |
| 415 | nordihydroguaiaretic acid | 25- | 367 | fulvestrant |
| 966 | tretinoin, [INN] | 22- | 447 | tretinoin |
| 513 | LY-294002 | 21- | 3 | metformin |
| 367 | fulvestrant, [INN] | 20+ | 582 | butein |
| 1048 | alpha-estradiol | 18- | 419 | chlorpromazine |
| 988 | estradiol, [INN] | 15- | 942 | prazosin |
| 953 | monorden | 16- | 970 | 5230742 |
| 502 | prazosin, [INN] | 17+ | 448 | trichostatin A |
| 361 | LY-294002 | 9- | 8+ | 254 | SC-58125 |
| 866 | ikarugamycin | 16+ | 460 | deferoxamine |

The top 10 instances sharing the largest number of significantly affected GOMs with Alzheimer hippocampus cells are listed in A and the top 10 instances resulted from distance comparison method are listed in B.

For detailed parameter settings: GO mode: BP; P-value: < 0.05.
reduced the risk of AD by anti-Aβ (β-amyloid) [51]. Estrogen (17β-E2) accelerated βAPP trafficking and precludes maximal Aβ generation within the TGN by modulating TGN phospholipid levels, particularly those of phosphatidylinositol. Therefore, fulvestrant, as estrogen blocker might activate phospholipid pathway to accelerate β-amyloid and aggravate AD. Based on our analysis, we propose that: 1) estrogen receptor pathway acts through phospholipid to alleviate AD; 2) Hsp90 inhibitors have another pathway to alleviate AD rather than estrogen pathway; 3) Despite that both estrogen receptor and Hsp90 inhibitors are promising drug targets for AD, estrogen receptor is a much better candidate.

The success of three cases between mouse models and human demonstrated that we can assess animal models’ similarity to human’s disease state. The main basis may be that orthologous genes were not only conserved at the sequence level and perform similar functions in different organisms, but also the corresponding gene expression patterns were conserved on a global level, especially between the human and mouse [15,16]. Our result that microarrays of cell response to molecules or drugs showed similarity across cell lines or tissues, to some extent, also explains why our approach was feasible to test mouse models. Nevertheless, as the diabetes drug case showed, it should be noticed that intrinsic differences always existed in normal and pathobiology states between species. [52]. Therefore, it was sometimes not appropriate for an animal model to mimic human diseases or drug response. Our method based on the analysis of the relationship of function-known drugs and human diseases utilizing microarray expression data performed well in both situations. Additionally, due to the introduction of GO annotations and the application of the statistical analysis, this cross-species method was able to provide bidirectional (both positive and negative) relationship between drugs and disease, and more clues about potential biological mechanisms. By contrast, the distance method seemed to be oversimplified and couldn’t provide such information.

In this paper, we presented data only from mouse animal models, for the mouse was the main model for common human diseases. At present, the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) enrolled 1295 datasets on homo sapiens and 1069 datasets on Mus musculus. Another
reason for using the mouse model was that the orthologous genes in mouse and human covered almost all genes in the cMap database.

Our cross-species analysis approach could also be extended to data from other cell lines, tissues, and human disease, which could be used to establish an animal model database instead of cMap. In addition, except for GO, other rules of gene partition such as KEGG were also good options. It was our primary goal to build extended references and additional gene modulation tools in the online service for biomedical research community.

Conclusions
In the present work, we introduced a new cross-species gene expression module comparison method to make the most of animal expression data and analyze the effectiveness of animal models in drug research. Through exploring the relations between drug molecules and mouse disease models, our method was able to assess whether the corresponding model recapitulates the essential features of the human disease. If so, this model may be suitable for drug molecules screening or even to test novel therapies systematically. Moreover, through data integration, our method could mine some meaningful information for drug research, such as potential drug candidates, possible drug repositioning, side effects and information about pharmacology.

Methods
Data source and preprocessing
Drug molecule response data was downloaded from Connectivity Map (cMap) (http://www.broad.mit.edu/cmap). cMap is a collection of gene-expression profiles of cultured human cells treated with bioactive small molecules or drug molecules. The data set was composed of mRNA expression data for 164 distinct small molecules and corresponding vehicle controls applied to human cell lines (564 gene expression profiles in total, representing 453 individual instances). All the data was generated by means of Affymetrix GeneChip microarrays. We normalized every instance by ranking the gene expressions and stored them in our own database for comparison.

The data of animal models were downloaded from GEO. In TSA case, there were 7 microarray data of mouse osteoblastic cells (MC3T3-E1) treated by Trichostatin A, including three replicates of TSA treatment and four replicates of control (GEO: GDS3002). In hypoxia case, we used 7 microarray assays (GEO: GSE117796) of bone marrow cells. The response of mouse to hypoxia was derived from a study by Laiwenfeld [24] in which mice received decreasing oxygen concentrations from 21% to 6% O2 for 30 minutes. Then, the mice remained at 6% O2 for another 120 minutes and the bone marrows were retrieved from the right humerus. In Diabetes drug case, we got microarray assays (GEO: GSE14888) of mouse 3T3-L1 adipocyte tissue cultures fed by metformin. In Alzheimer case, the animal model was transgenic (TG) mice expressing human APP695 and bearing the double Swedish and Indiana amyloid precursor protein (APP) mutations [1]. Six microarray assays (Hippocampus cells from two normal and four APP transgenic mice, GEO: GSE14499) were obtained.

Orthologous gene matching
Orthologous gene conversion relied on the Roundup database [18] (http://roundup.hms.harvard.edu), a large-scale database of orthologs. The orthologs were computed by the Reciprocal Smallest Distance (RSD) algorithm, which was developed by Wall et al. [53]. For human and mouse, about 13264 genes were selected by RSD algorithm. These genes covered almost all genes in the small molecule database of cMap.

Gene modularization comparison method
The processes of our method are depicted in Figure 2. After ortholog matching on the gene expression data of animal model, 1.5-fold change was used as default threshold for differential expression (2-fold change, 1.3-fold change or t-test could also be used), and then hyper geometric test was performed in every Gene Ontology Module (GOM). We chose Gene Ontology Module as our modularization reference, because it was the most widely used in exploring biological features of genes with respect to their molecular functions, biological processes as well as cellular components.

GOMs were selected, when p-values from hyper geometric test were smaller than 0.05. Based on each selected GOMs, the expression pattern similarity between the animal model data and the chemicals data in the cMap database was calculated.

The algorithm was derived from Kolmogorov-Smirnov statistics (KS-test), which was called connectivity score in Lamb et al.’s work [17]. But Lamb et al. applied the algorithm on the whole profile and we applied it in every GOM. The KS score indicated the similarity of two samples. For each GOM, it showed genes that had the same or reverse pattern of expression between the query and reference chemicals. If the KS score was positive in a certain GOM, the query and reference chemicals would have similar pattern of expression in this GOM, and vice versa. P-value was also calculated to indicate significance of the comparison [19]. Similarly, only GOMs with p-value < 0.05 would be selected.

The result of performing one similarity search was a table, whose each column represented a chemical in
reference library and each row represented a GOM. The value in each grid was the KS score or p-value (the p-value of comparison, not the p-value of hyper geometric test) of the query and reference chemicals in certain GOM. The top 10 reference chemicals which had the most similar GOM numbers were selected for each analysis.

Distance comparison method
As a control to our method, we also used distance method to perform a cross-species analysis. The distance method has been used by other researchers in the cross-species analysis, where euclidean distances were computed to cluster the similar samples [13]. But in this study we applied absolute distances to show the similarity between the gene expression data from animal model and human, in the case that all the gene expression data in the cMap database was given ranking values.

First, orthologous genes matching and differential expression analysis were done on the gene expression data of animal models. Then the differential expressed genes were ranked, similar to the corresponding genes of each instance in the cMap. Absolute distances were calculated between the animal model and each instance by

\[ |x_1 - y_1| + \ldots + |x_k - y_k| \]

where \( k \) means the number of genes and \( x \) and \( y \) are animal and instances samples, respectively. The top 10 instances which had the smallest distance values were selected.

Acknowledgements
Funding: This work is supported by the National Basic Research Program of China (973 Program) (2012CB517905, 2012CB316501), National Key Basic Research Program (Grant No. 90913009). Research supported by the listed fundings are open for public access. We Gratefully Acknowledges the Support of SA-SIBS Scholarship Program. This article has been published as part of BMC Systems Biology Volume 6 Supplement 3, 2012: Proceedings of The International Conference on Intelligent Biology and Medicine (ICIBM) - Systems Biology. The full contents of the supplement are available online at http://www.biomedcentral.com/bmcisyb/supplements/6/S3.

Author details
1College of Life Science and Biotechnology, Shanghai Jiaotong University, 800 Dongchuan Road, Shanghai 200240, P.R.China. 2Key Lab of Systems Biology/Key Laboratory of Synthetic Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 Yueyang Road, Shanghai 200031, P.R.China. 3Institut Pasteur of Shanghai, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 Yueyang Road, Shanghai 200031, P.R.China. 4College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, P.R.China. 5Shanghai Center for Bioinformation Technology, Shanghai, P.R.China.

Authors’ contributions
SY performed design. SY and LZ statistical analyses for the study and drafted the manuscript. YL, CL and CM participated in the design of the study and provided guidance. YXL, PH and XL conceived the study, and finalized the organization and contents of the manuscript. All authors approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.
Published: 17 December 2012

References
1. Nagahara AH, Merrill DA, Coppola G, Tsukada S, Schroeder BE, Shaked GM, Wang L, Bleisch A, Kim A, Conner JM, et al. Neuroprotective effects of brain-derived neurotrophic factor in rodent and primate models of Alzheimer's disease. Nat Med 2009, 15(3):331-337.
2. Chin J Selecting a mouse model of Alzheimer's disease. Methods Mol Biol 2011, 670:169-189.
3. Miller JA, Horvath S, Genschwind DH: Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. Proc Natl Acad Sci USA 2010, 107(28):12698-12703.
4. Hallak M, Hancock KD, Coffin CM, Lessnick SL, Capecchi MR. A conditional mouse model of synovial sarcoma: insights into a myogenic origin. Cancer Cell 2007, 11(4):375-388.
5. Hung KE, Maricevich MA, Richard LG, Chen WY, Richardson MP, Kunin A, Hung-Son H, Sun RL, Yamashita Y, et al. Environmental effects on murine vascular smooth muscle cell growth. Int J Dev Biol 2007, 51(12):2327-2335.
6. Singh M, Johnson L: Using Genetically Engineered Mouse Models of Alzheimer's Disease for Drug Discovery and Development. Current Drug Metabolism 2009, 10(2)116-124.
7. Zon LI, Peterson RT. In vivo drug discovery in the zebrafish. Nat Rev Drug Discov 2005, 4(1)35-44.
8. Olson H, Betton G, Robinson D, Thomas K, Monro A, Kolaja G, Lilly P, Sanders J, Sipes G, Brackett W, et al. Concordance of the toxicity of pharmaceuticals in humans and in animals. Regul Toxicopharmacol 2000, 32(1):56-67.
9. Strand AD, Aragaki AK, Baquet ZC, Hodges A, Cunningham P, Holmans P, Jones KR, Jones L, Koopertberg C, Olson JM. Conservation of regional gene expression in mouse and human brain. PLoS Genet 2007, 3(4)el9.
10. Dowell RD. The similarity of gene expression between human and mouse tissues. Genome Biol 2011, 12(1)101.
11. Sharpless NE, Depinho RA. The mighty mouse: genetically engineered mouse models in cancer drug development. Nat Rev Drug Discov 2006, 5(9)741-754.
12. Olson H, Betton G, Robinson D, Thomas K, Monro A, Kolaja G, Lilly P, Sanders J, Sipes G, Brackett W, et al. Concordance of the toxicity of pharmaceuticals in humans and in animals. Regul Toxicopharmacol 2000, 32(1):56-67.
13. Strand AD, Aragaki AK, Baquet ZC, Hodges A, Cunningham P, Holmans P, Jones KR, Jones L, Koopertberg C, Olson JM. Conservation of regional gene expression in mouse and human brain. PLoS Genet 2007, 3(4)el9.
14. Dowell RD. The similarity of gene expression between human and mouse tissues. Genome Biol 2011, 12(1)101.
15. Zheng-Bradley X, Rung J, Parkinson H, Brazma A. Large scale comparison of global gene expression patterns in human and mouse. Genome Biol 2010, 11(1)R214.
16. Chan ET, Quon GT, Chua G, Babak T, Trochesset M, Zirmgbl RA, Aubin J, Raciliffe NJ, Wildle A, Brodno M, et al. Conservation of core gene expression in vertebrate tissues. J Biol 2009, 8(3)33.
17. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross K, et al. The Connectivity Map: using gene expression signatures to connect small molecules, genes, and disease. Science 2010, 32(1)5795:1929-1935.
18. Deluca TF, Wu IH, Pu J, Monaghan T, Pethkin L, Singh S, Wall DP. Roundup: a multi-genome repository of orthologs and evolutionary distances. Bioinformatics 2006, 22(10)2044-2046.
19. Li Y, Hao P, Zheng S, Tu K, Fan H, Zhu R, Ding G, Dong C, Wang C, Li X, et al. Gene expression module-based chemical function similarity search. Nucleic Acids Res 2008, 36(20)e137.
20. Dowdell KC, Pesinick L, Hoffmann Y, Steadman K, Remaley AT, Cohen JI, Strauss SE, Rao VK. Valproic acid (VPA), a histone deacetylase (HDAC) inhibitor, diminishes lymphoproliferation in the Fas -deficient murine model of autoimmune lymphoproliferative syndrome (ALPS). Exp Hematol 2009, 37(4)487-494.
21. Hrzenjak A, Mointal F, Kremser ML, Strohmeier B, Petru E, Zlatoluk K, Denk H. Histone deacetylase inhibitor vorinostat suppresses the growth of uterine sarcomas in vitro and in vivo. Mol Cancer 2010, 9:49.
22. Badyaro Y, Drosch G, Graessel S, Trojer P, Walton JD. Characterization of inhibitor-resistant histone deacetylase activity in plant-pathogenic fungi. Eukaryot Cell 2002, 1(1)538-547.
inflammation correlating with an increase PPAR gamma expression in cultured astrocytes. Brain Res 2010, 1312:138-144.

45. Luo W, Rodina A, Chiosis G: Heat shock protein 90: translation from cancer to Alzheimer’s disease treatment? BMC Neurosci 2008, 9(Suppl 2):S7.

46. Shineman DW, Dain AS, Kim ML, Lee YM: Constitutively active Akt inhibits trafficking of amyloid precursor protein and amyloid precursor protein metabolites through feedback inhibition of phosphoinositide 3-kinase. Biochemistry 2009, 48(17):3787-3794.

47. Wang L, Shifer HB, Rehde K, Hart KL, Hoff DJ, McFall YH, Raskind MA, Peskind EP: Prazosin for the treatment of behavioral symptoms in patients with Alzheimer disease with agitation and aggression. Am J Geriatr Psychiatry 2009, 17(9):744-751.

48. Han G, Ma H, Chintala R, Fulton DJ, Barman SA, White RE: Essential role of the 90-kilodalton heat shock protein in mediating nongenomic estrogen signaling in coronary artery smooth muscle. J Pharmacol Exp Ther 2009, 329(3):850-855.

49. Iorio F, Rosotti R, Scarchetti E, Belcastro V, Mithbaokar P, Ferrero R, Murino L, Tagliaferri R, Brunetti-Pierri N, Isacchi A, et al: Discovery of drug mode of action and drug repositioning from transcriptional responses. Proc Natl Acad Sci USA 2010, 107(33):14621-14626.

50. Fliss AE, Benazeo S, Rao J, Caplan AJ: Control of estrogen receptor ligand binding by Hsp90. J Steroid Biochem Mol Biol 2000, 72(5):229-230.

51. Greenfield JP, Leung LW, Cai D, Kaasik K, Gross RS, Rodriguez Poulan F, Greengard P, Xu H: Estrogen lowers Alzheimer beta-amyloid generation by stimulating trans-Golgi network vesicle biogenesis. J Biol Chem 2002, 277(14):12128-12136.

52. Schofield PN, Sundberg JP, Hoehndorf R, Gkoutos GV: New approaches to the representation and analysis of phenotype knowledge in human diseases and their animal models. Brief Funct Genomics 2011, 10(5):258-265.

53. Wall DP, Fraser HB, Hirsch AE: Detecting putative orthologs. Bioinformatics 2003, 19(13):1710-1711.

doi:10.1186/1752-0509-6-S3-S18

Cite this article as: Yu et al.: A cross-species analysis method to analyze animal models’ similarity to human’s disease state. BMC Systems Biology 2012, 6(Suppl 3):S18.