DR. WEICHEN ZHOU (Orcid ID : 0000-0003-4755-1072)

Article type : Method

Title

A standardized fold change (SFC) method for microarray differential expression analysis used to reveal genes involved in acute rejection in murine allograft models

Authors

Weichen Zhou[1,3,†], Yi Wang[1,†], Masayuki Fujino[2,4], Leming Shi[1], Li Jin[1], Xiao-Kang Li[1,2,*], Jiucun Wang[1,*]

Addresses

[1]State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, Collaborative Innovation Center for Genetics and Development, School of Life Sciences and Institutes of Biomedical Sciences, Fudan University, Shanghai, China; [2]Division of Transplantation Immunology, National Research Institute for Child Health and Development, Tokyo, Japan; [3]Department of Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, USA; [4]AIDS Research Center, National Institute of Infectious Diseases, Tokyo, Japan

[†]Weichen Zhou and Yi Wang contributed equally to this work.

***Corresponding authors**

Xiao-Kang Li, M.D., Ph.D., Division of Transplantation Immunology, National Research Institute for Child Health and Development, 2-10-1 Okura, Setagaya-ku, Tokyo, 157-8535 Japan.

Tel: +81-3-3416-0181; Fax: +81-3-3417-2864; E-mail: ri-k@ncchd.go.jp

Jiucun Wang, Ph.D., School of Life Sciences, Fudan University, 2005 Songhu Road, Shanghai 200438, China.

Tel: +86-21-5163-0606; Fax: +86-21-5163-0607; E-mail: jcwang@fudan.edu.cn

Running title

SFC method for microarray analysis

Abbreviations

SFC: standardized fold change; Limma: linear models for microarray data; MAQC: micro-array quality control; POD: days post operation; qRT-PCR: quantitative reverse transcriptase-polymerase chain reaction; FPR: false positive rate; FNR: false negative rate

Keywords

Murine transplantation model, microarray analysis, standardized fold change, gene expressions

Conflict of interest statement

The authors have declared that no conflict of interest exists.

Abstract

Murine transplantation models are used extensively to research immunological rejection and tolerance. Here, we studied on both murine heart and liver allograft models using microarray technology. However, we had difficulties in identifying genes related to acute rejections expressed in both heart and liver transplantation models using two standard methodologies: Student's T-test and linear models for

microarray data (Limma). Here we describe a new method, standardized fold change (SFC), for differential analysis of microarray data. We estimated the performance of SFC, T-test and Limma by generating simulated microarray data 100 times. SFC performed better than T-test and showed a higher efficiency than Limma in sensitivity where a larger fold change of expression value exists. SFC had better reproducibility than Limma and T-test in real experimental data from Micro-Array Quality Control (MAQC) platform and expression data of mouse cardiac allograft. Eventually, a group of significant overlapping genes was detected by SFC in the expression data of mouse cardiac and hepatic allografts and further validated by the quantitative RT-PCR assay. The group included genes for important reactions of transplantation rejection and revealed functional changes of the immune system in both heart and liver of the mouse model. We suggest that SFC can be utilized to stably and effectively detect the differential gene expression and to explore microarray data in further studies.

Introduction

At the very stage of organ failure, the organ transplantation is the life-saving medical procedure, though still having some problems, e.g., transplant rejection, reception of life-long immunosuppressive drugs. Transplantation models without immunosuppression are important and the mechanisms of rejection and tolerance in these models are highly required to be disclosed.

The microarray is a well-established and widely used technology, providing a picture of gene expression or RNA profiling in different tissues [1]. To identify the differential expressions, Student's *T-test* and linear models for microarray data (*Limma*) are two popular choices [2-4]. *T-test* utilizes information of all the samples (or standard deviation) in one microarray probe and conducted independently among different probes [4]; while *Limma* uses the empirical Bayes approach to shrinkage of the estimated sample variances towards a pooled estimate. The information (means and standard deviations) from all probes in a replicate set of experiments are combined and used at one probe level to detect differential expressions in *Limma* [2].

In the present study, we estimated murine heart and liver allograft models and conducted microarray technology to reveal the significant genes that related to

transplant rejection. By using *T-test* and *Limma*, no significant intersecting genes were obtained in the models of murine heart and liver allografts. Therefore, we developed a new method named Standardized Fold Change (*SFC*) to detect differential expressions by borrowing information from neighbors of one probe with an adjustable bin size. To compare *SFC* with *T-test* and *Limma*, we generated a simulated data set to estimate the performance and conducted the real experimental datasets from Micro-Array Quality Control (MAQC) platform and the transplantation model to estimate the reproducibility. We concluded that *SFC* can be applied as a new and effective approach for differential expression detection and contribute more reliable results in the microarray studies. Eventually, we called a set of significant genes from expression data of murine heart and liver allograft model by *SFC* and validated them by qRT-PCR. Gene expression changes reveal functional reactions and pathway activities in the early stage of allograft in both heart and liver.

Results

The SFC method

We observed the distribution of the mean value and variance of one probe signal is non-linear (**Supplementary Figure 1).** The information of neighboring probes can usually be borrowed to improve the statistical power [2]. *SFC* was introduced to estimate variance for each probe: rather than obtaining from all samples; it takes information from neighbors of that probe with an adjustable bin size *b*. As we set up the default value of *b* as 1000, the variances of cases and controls in one probe can be obtained by calculating the median of those probes, separately. Eventually, following by the Formula (1), we can obtain the statistic *SFC* for every probe, and *p*-value can be further estimated based on these.

SFC had a better sensitivity and specificity based on simulation data

We investigated the false positive rate (FPR) and the false negative rate (FNR) of these three methods under the null hypothesis and alternative hypothesis. As indicated in Formula (2), signals of the null hypothesis were generated by a simple formula $y=x$

with a Gaussian noise added. The basic formulas are adjustable with the parameters *k*. The signals of alternative hypothesis were described by Formula (3), with variable values of θ and the portion of real positive calls. We calculated the FPR and FNR for every different θ and portion of real positive calls with a 0.05 significant threshold and 100-time simulation (**Table 1)**.

Under the null hypothesis, the rates of three methods are all near the significant threshold between 5% and 6% **(Figure 1A)**. Under the alternative hypothesis, *SFC* had a better performance of FPR than the other two methods generally (**Figure 1B**). With an increasing θ and portion of real positive calls, the FPR of *SFC* showed a decreasing bias, whereas *Limma* and *T-test* showed a positive bias with these parameters (**Table. 1**). To the FNR, as the θ and portion of real positive calls increase, *Limma* showed a faster decline than *T-test*, while *SFC* had a lower FNR than *Limma* and performed better with larger θ and portion of real positive calls. Interestingly, *SFC* shows a relatively small number of calls (from 4.9% to 10.5%, **Table. 1**), while *Limma* and *T-test* calls a larger set in this situation. To sum, comparing with *Limma* and *T-test* at the significant threshold of 0.05, *SFC* had a better sensitivity and specificity, especially with a larger value of differential expression fold change (θ = 50%).

Reproducibility of SFC is better than Limma and T-test based on MAQC and mouse transplantation data

Reproducibility is an indispensable estimator for the experiments and algorithms [5, 6]. We choose both MAQC dataset and mouse cardiac transplantation data to assess the reproducibility of *SFC*, *Limma* and *T-test*.

We calculated the reproducibility of top 100 and top 1000 genes for MAQC by using the three methods. For the inter-platform, heat-map shows that *SFC* performed a better reproducibility than *Limma* and *T-test* among six platforms when detecting both top 100 and top 1000; while for intra-platform reproducibility, all three methods did not perform well in detecting either top 100 or 1000 significant genes (**Figure 2A&B)**. Same operations were conducted in the mouse cardiac transplantation data, where

SFC also showed a better performance than the others (**Figure 2C)**. Therefore, according to better performances of reproducibility in both MAQC data and mouse transplantation data, *SFC* is more stable than *Limma* or *T-test*.

Intersected significances from mouse transplantation data were found by SFC and validated by qRT-PCR

We further utilized the three methods to analyze the mouse organ transplantation data and validated the results. After experimental process generating CEL files from mouse tissues, we conducted these methods on the expression data of POD5 of cardiac transplantation and the POD5 and POD8 of hepatic transplantation.

According to *SFC*, one hundred and seventy-eight significant genes were differentially expressed in the cardiac allografts compared with isografts, including 158 over-expressed genes and 20 under-expressed genes (**Figure 3**). There were also 362 genes (263 over-expression and 99 under-expression) have significantly different expressions in the hepatic POD5 allografts compared with isografts, and 389 genes (258 over-expression and 131 under-expression) have significantly different expressions in the hepatic POD8 allografts compared with isografts, respectively. Based on these, an intersection of these three groups including 52 important genes was obtained, in which they are all over-expressed for cardiac transplantation and 51 over-expressed and one under-expressed for hepatic transplantation (**Figure 3)**. At the same time, the calling sets of significant genes underlying *Limma* and *T-test* (**Supplementary Figure 4A&B**) showed no intersected ones.

We further performed quantitative RT-PCR for the calls derived from *SFC* to validate the fold changes of the mRNA expressions. Nineteen mRNAs, which were up-regulated in both of cardiac and hepatic allografts compared with isografts, were randomly selected (**Table 2 & Supplementary Table 3**). Being consistent with the results of microarray, a significantly higher amount of mRNA expression was detected in allografts versus isografts in cardiac (**Figure 4A)** and hepatic (**Figure 4B**) allografts.

Discussion

Microarray is widely used and accepted as a stable, well established and less costing technology to investigate gene expression data [1, 7-9]. In this study based on the microarray data, we established a novel method *SFC* to detect differential expressions and compared it with *T-test* and *Limma*. According to the Formula (1), the parameter b can be adjusted to control the nearby number of probes, which contribute the variant of the central probe. We set 1000 as default, and users can customize its value based on a different size of microarray probes. For the simulation data, the parameter configurations (θ and k) of the null hypothesis and alternative hypothesis also can be adjusted (Formula 2&3) [10]. Moreover, we calculated the FPR and FNR based on different significant levels (p-value = 0.01 and 0.001) for different values of θ and k. With a more stringent significant level (from 0.05 to 0.001), the FPRs were decreasing while the FN rates were increasing, which was observed by all three methods (**Figure 1, Supplementary Figure 2&3, Table 1, and Supplementary Table 1&2**). Notably, when p-value equals to 0.001, *T-test* performed a high FPR (48%, θ = 50%, and true positive gene percent = 10%) and *Limma* performed with a high FNR (sometimes more than 90%). These suggested the *T-test* choose to give more positive hits with high FPR, while *Limma* will report fewer hits to reduce the FPR but miss some TP ones. Importantly, *SFC* can make a good balance of FPR and FNR, and perform well both in FPR and FNR with stringent significant level.

Statistical correction (e.g. Bonferroni correction) is often introduced for multiple comparisons to adjust p-value and control the false discovery rate [11]. We also analyzed the mice transplantation data by the other two methods (*Limma* and *T-test*) on the different significant levels (p-value = 0.05, 0.001, and 0.05 with Bonferroni correction). *Limma* and *T-test* had a huge number of positive hits when the p-value is 0.05 in three phases (**Supplementary Figure 5&6)**. When p-value is 0.001, the positive hits by *Limma* and *T-test* decreased a lot while by *SFC* the number stayed respectively stable. When p-value was stringent to 0.05 with Bonferroni correction (**Figure 3 & Supplementary Figure 4**), *SFC* still reported 52 significances

overlapping with three phases, but *Limma* and *T-test* showed no overlapping significance. The results of *T-test* showed no sharing significance with *SFC*. Intriguingly, in 67 significances of cardiac POD5 reported by *Limma* (**Supplementary Figure 4**), thirty genes showed in the cardiac POD5 result of *SFC*, and 16 showed in the 52 significances. Besides, for hepatic POD5 and POD8 by *Limma*, 4 out of 7 (POD5) and 19 out of 36 (POD8) were observed in the corresponding results of *SFC*, and 2 out of 5 (overlapping in POD5 and POD8) appear in the 52 genes of *SFC*. As 19 in 52 genes from *SFC* were randomly selected and all passed the validation of qRT-PCR, these results indicated that *SFC* had a more stable result than *T-test* and *Limma*.

We, therefore, investigated the functions of these 52 genes (**Supplementary Table 4**), revealing the most significant pathways were graft-versus-host disease (mmu05332) and allograft rejection (mmu05330). Moreover, immune system response (e.g. mmu04612, mmu04660, GO: 0006955) and positive regulation (e.g. GO: 0050863, GO: 0051249, GO: 0050870) were also activated. All these enrichment analyses indicated a reaction of transplantation rejection in vivo and functional changes of immune system both in cardiac and hepatic level after five days of allografts [12-14].

In conclusion, based on the quality control experimental data and simulated data, *SFC* performed some better than *Limma* and much better than *T-test* by using the nearby information of one probe in pooled probes. We utilized *SFC* into the real data of mouse transplantation models, and it reported a more stable and convincible set with 52 significant genes, revealing the insights of pathway and gene expression changes after both cardiac and hepatic allografts. Nineteen genes were further randomly picked up and validated by qRT-PCR. We suggested *SFC* is a new and effective approach, which can detect differential expressions and help to find more reliable information in microarray studies.

Materials and Methods

Animal

Male B10.BR (BR, H-2k), B10.D2 (D2, H-2d), C57BL/ 10 (B10, H-2b), and CBA (H-2k) mice (weighing 25-30 g) were purchased from the Shizuoka Laboratory Animal Center (Shizuoka, Japan) and housed and cared for in agreement with the guidelines of the Institutional Animal Care and Use Committee and the National Research Institute for Child Health and Development guidelines on laboratory animal welfare. The Committee on the Care and Use of Laboratory Animals at the National Research Institute accepted the experimental protocol for Child Health and Development (Permission Number: 2002-003). All surgical procedures were conducted by anesthetization with isoflurane/oxygen, and all attempts were carried out to minimize suffering.

Transplantation and RNA extraction

The cardiac transplantation was performed on the sex-matched B10 donor to the CBA recipient by microsurgical techniques. Intra-abdominal vascularized heterotopic mouse cardiac transplantation was performed [15]. The cardiac graft survival was determined using daily palpation of the recipient's abdomen. Three case samples on the fifth day were obtained. BR mice were used as donors and D2 mice were used as recipients in the orthotopic hepatic transplantation. We performed transplantation surgery on the mice [12]. For orthotopic liver transplantation, BR mice were used as donors and D2 mice were used as recipients. We subsequently transplanted the hepatics into the recipient mice using the cuff technique [12]. Grafts were harvested at 5 days post operation (POD5) or POD8 after transplantation and were submerged in RNAlater® stabilization solution (Life Technologies, Carlsbad, CA) for freezing. Total RNA was extracted from frozen tissue samples using ISOGEN (NipponGene, Tokyo, Japan). We also designed control groups of three normal cardiac tissues and three hepatic tissues.

Standardized fold change method

The probe signals from microarray data were firstly transformed by *Ln* and then manipulated with quantitative normalization. To assess the differential expressions

among cases and controls, the statistics *SFC* is defined as:

$$SFC_i = \frac{T - C}{STDEV(T - C)} = \frac{T_i - C_i}{\sqrt{Var(T - C)}} = \frac{Median(t_1, t_2, ..t_i) - Median(c_1, c_2, ..c_i)}{\sqrt{Median((T_{i-b/2} - C_{i-b/2})^2, .. (T_i - C_i)^2, .. (T_{i+b/2} - C_{i+b/2})^2) / 0.455}} \quad (1)$$

For the variance of each probe, we ranked all probes by the mean values of signals from all samples and then took the median value of its *b* nearest neighbors as the variance, where the default bin size of *b* here is 1000. The SFC software now implements this algorithm in Linux system at https://github.com/WeichenZhou/SFC.

Simulation data study

We generated the simulated data from simple formulas with the Gaussian noise (mean = 0, variance = 1) as a default distribution for gene expression data [16]. The control and case samples in the null hypothesis are shown as follows:

$$H0\ control: y_0 = x_0 + N(0,1)(\sqrt{kx_0} + 1)$$
$$H0\ case: y_0' = (1 + \theta_0)x_0' + N(0,1)(\sqrt{kx_0'} + 1) \quad (2)$$

The θ represented the differential expression underlying cases versus controls and we defined θ_0 is 0% and *k* is 1. The control and case samples in the alternative hypothesis are shown as follows:

$$H1\ control: y_1 = x_1 + N(0,1)(\sqrt{kx_1} + 1)$$
$$H1\ case: y_1' = (1 + \theta_1)x_1' + N(0,1)(\sqrt{kx_1'} + 1) \quad (3)$$

We defined θ_1 as 10%, 25% and 50%, respectively. The size of real positive calls consists 1%, 5%, 10% of the whole simulated data, respectively. Following these, a 100-time simulation was conducted to assess FPR and FNR.

MAQC data and the reproducibility analysis

MAQC data was conducted by the US Food and Drug Administration (FDA) to develop standards and quality control metrics, which involved six Centers (ABI, AFX, AGI, GEH, ILM, and NCI) major providers of microarray platforms and RNA

samples [1, 7]. The reproducibility of top 100 significant genes and 1000 significant genes were estimated in the inter- and intra-platform by the three statistical methods and the heat maps were drawn with the matrix of each batch. For the expression data from mouse transplant model, we picked up two out of three cases and controls to build one batch and made a 9x9 matrix heat map to estimate the reproducibility. The significant level of mice microarray data was 0.05.

Application on mouse transplantation data

We detected the gene differential expressions between cases and controls in three phases: POD5 of cardiac transplantation, POD5 of hepatic transplantation and POD8 of hepatic transplantation. All *p*-values from expression data are adjusted by the Bonferroni correction. After getting all significant probes from *SFC*, we converted these probe level significance to gene level using annotation file. The Venn diagrams showed the significant genes with differential expression. Pathway and GO enrichment analyses were performed by using the Database for Annotation, Visualization and Integrated Discovery (DAVID) database (http://david.abcc.ncifcrf.gov/) with the Bonferroni correction adjusted *p*-values less than 0.05 [17]. Mouse transplantation data have been deposited in NCBI's Gene Expression Omnibus [18] and are accessible through GEO Series accession number GSE89340 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=qrmliwycxlovtuf&acc=GSE89340). All data were conducted by quantile normalization before processed by different methods. *Limma* can be found as the R package ***Limma*** [2, 3] and the heat maps were conducted by ***gplots***. All R packages can be downloaded from Bioconductor (www.bioconductor.org).

Quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR)

The RNA was reverse-transcribed to cDNA using a PrimeScript® RT Reagent Kit (Takara Bio, Shiga, Japan) as described previously [19]. The sequences used in our study are shown in **Supplementary Table 3**. Quantitative RT-PCR was performed

using a SYBR green system on the Applied Biosystem PRISM7700 instrument (Applied Biosystems, Carlsbad, CA), and experiments were conducted using 0.4 μM of each primer in a final reaction volume of 20 μl of KAPA SYBR® FAST qPCR kit (Kapa Biosystems, South Africa). The PCR cycling conditions were as follows: 95°C for 30 sec, and 50 cycles of 95°C for 5 sec, 60°C for 1 min. The normalized threshold cycle (Ct) value of each gene was obtained by subtracting the Ct value obtained for 18S rRNA. The cardiac mRNA levels were analyzed on POD5. **Figure 4** indicates the number of copies of each of the three representative mRNAs measured in the syngeneic grafts or allografts obtained from three individuals. The relative amount of each mRNA was normalized to that of 18S rRNA. All experiments were analyzed in three mice per each time point and expressed as the mean ± SEM. *$p < 0.05$ compared with syngeneic on day 5.

Acknowledgments/ FUNDING

This study was supported by research grants from National Natural Science Foundation of China (31521003) and the 111 Project (B13016), and the National Center for Child Health and Development (26-6), Ministry of Education, Culture, Sports, Science and Technology of Japan (Grants-in-Aid 15F15756 and 15K10043).

Author Contributions

W.Z., Y.W., X.L. and J.W. designed the project. L.S. supported the MAQC data, and M.F. and X.L. supported the mouse model and validations. W.Z. carried out the analysis and simulations. W.Z., X.L. and J.W. wrote the manuscript. L.J., X.L. and J.W. contributed to the final revision the paper. All authors read and approved the final manuscript.

References:

1. Consortium, M.Shi, L.Reid, L. H.Jones, W. D.Shippy, R.Warrington, J. A.Baker, S. C.Collins, P. J.de Longueville, F.Kawasaki, E. S.Lee, K. Y.Luo, Y.Sun, Y. A.Willey, J. C.Setterquist, R. A.Fischer, G. M.Tong, W.Dragan, Y. P.Dix, D. J.Frueh, F. W.Goodsaid, F. M.Herman, D.Jensen, R. V.Johnson, C. D.Lobenhofer, E.

354 K.Puri, R. K.Schrf, U.Thierry-Mieg, J.Wang, C.Wilson, M.Wolber, P. K.Zhang, L.Amur, S.Bao, W.Barbacioru,
355 C. C.Lucas, A. B.Bertholet, V.Boysen, C.Bromley, B.Brown, D.Brunner, A.Canales, R.Cao, X. M.Cebula, T.
356 A.Chen, J. J.Cheng, J.Chu, T. M.Chudin, E.Corson, J.Corton, J. C.Croner, L. J.Davies, C.Davison, T.
357 S.Delenstarr, G.Deng, X.Dorris, D.Eklund, A. C.Fan, X. H.Fang, H.Fulmer-Smentek, S.Fuscoe, J.
358 C.Gallagher, K.Ge, W.Guo, L.Guo, X.Hager, J.Haje, P. K.Han, J.Han, T.Harbottle, H. C.Harris, S.
359 C.Hatchwell, E.Hauser, C. A.Hester, S.Hong, H.Hurban, P.Jackson, S. A.Ji, H.Knight, C. R.Kuo, W.
360 P.LeClerc, J. E.Levy, S.Li, Q. Z.Liu, C.Liu, Y.Lombardi, M. J.Ma, Y.Magnuson, S. R.Maqsodi, B.McDaniel,
361 T.Mei, N.Myklebost, O.Ning, B.Novoradovskaya, N.Orr, M. S.Osborn, T. W.Papallo, A.Patterson, T.
362 A.Perkins, R. G.Peters, E. H., et al. (2006) The MicroArray Quality Control (MAQC) project shows inter-
363 and intraplatform reproducibility of gene expression measurements, *Nature biotechnology.* **24**,
364 1151-61.
365 2. Smyth, G. K. (2004) Linear models and empirical bayes methods for assessing differential
366 expression in microarray experiments, *Statistical applications in genetics and molecular biology.* **3**,
367 Article3.
368 3. Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W. & Smyth, G. K. (2015) limma powers
369 differential expression analyses for RNA-sequencing and microarray studies, *Nucleic acids research.* **43**,
370 e47.
371 4. Rice, J. A. (2007) *Mathematical statistics and data analysis*, 3rd edn, Thomson/Brooks/Cole,
372 Belmont, CA.
373 5. Vaux, D. L., Fidler, F. & Cumming, G. (2012) Replicates and repeats--what is the difference and is it
374 significant? A brief discussion of statistics and experimental design, *EMBO reports.* **13**, 291-6.
375 6. Fomel, S. & Claerbout, J. F. (2009) Guest Editors' Introduction: Reproducible Research, *Computing*
376 *in Science & Engineering.* **11**, 5-7.
377 7. Shi, L.Campbell, G.Jones, W. D.Campagne, F.Wen, Z.Walker, S. J.Su, Z.Chu, T. M.Goodsaid, F.
378 M.Pusztai, L.Shaughnessy, J. D., Jr.Oberthuer, A.Thomas, R. S.Paules, R. S.Fielden, M.Barlogie, B.Chen,
379 W.Du, P.Fischer, M.Furlanello, C.Gallas, B. D.Ge, X.Megherbi, D. B.Symmans, W. F.Wang, M. D.Zhang,
380 J.Bitter, H.Brors, B.Bushel, P. R.Bylesjo, M.Chen, M.Cheng, J.Cheng, J.Chou, J.Davison, T. S.Delorenzi,
381 M.Deng, Y.Devanarayan, V.Dix, D. J.Dopazo, J.Dorff, K. C.Elloumi, F.Fan, J.Fan, S.Fan, X.Fang,
382 H.Gonzaludo, N.Hess, K. R.Hong, H.Huan, J.Irizarry, R. A.Judson, R.Juraeva, D.Lababidi, S.Lambert, C.
383 G.Li, L.Li, Y.Li, Z.Lin, S. M.Liu, G.Lobenhofer, E. K.Luo, J.Luo, W.McCall, M. N.Nikolsky, Y.Pennello, G.

A.Perkins, R. G.Philip, R.Popovici, V.Price, N. D.Qian, F.Scherer, A.Shi, T.Shi, W.Sung, J.Thierry-Mieg, D.Thierry-Mieg, J.Thodima, V.Trygg, J.Vishnuvajjala, L.Wang, S. J.Wu, J.Wu, Y.Xie, Q.Yousef, W. A.Zhang, L.Zhang, X.Zhong, S.Zhou, Y.Zhu, S.Arasappan, D.Bao, W.Lucas, A. B.Berthold, F.Brennan, R. J.Buness, A.Catalano, J. G.Chang, C.Chen, R.Cheng, Y., et al. (2010) The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models, *Nature biotechnology.* **28**, 827-38.

8. Barrett, T. & Edgar, R. (2006) Mining microarray data at NCBI's Gene Expression Omnibus (GEO)*, *Methods in molecular biology.* **338**, 175-90.

9. Roy, N. C., Altermann, E., Park, Z. A. & McNabb, W. C. (2011) A comparison of analog and Next-Generation transcriptomic tools for mammalian studies, *Briefings in functional genomics.* **10**, 135-50.

10. Wang, Y., Li, Y., Cao, H., Xiong, M., Shugart, Y. Y. & Jin, L. (2015) Efficient test for nonlinear dependence of two continuous variables, *BMC bioinformatics.* **16**, 260.

11. Cui, X. & Churchill, G. A. (2003) Statistical tests for differential expression in cDNA microarray experiments, *Genome biology.* **4**, 210.

12. Morita, M., Chen, J., Fujino, M., Kitazawa, Y., Sugioka, A., Zhong, L. & Li, X. K. (2014) Identification of microRNAs involved in acute rejection and spontaneous tolerance in murine hepatic allografts, *Scientific reports.* **4**, 6649.

13. Chen, X., Chang, C. H., Stein, R., Cardillo, T. M., Gold, D. V. & Goldenberg, D. M. (2013) Prevention of acute graft-versus-host disease in a xenogeneic SCID mouse model by the humanized anti-CD74 antagonistic antibody milatuzumab, *Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation.* **19**, 28-39.

14. Fujiwara, H., Maeda, Y., Kobayashi, K., Nishimori, H., Matsuoka, K., Fujii, N., Kondo, E., Tanaka, T., Chen, L., Azuma, M., Yagita, H. & Tanimoto, M. (2014) Programmed death-1 pathway in host tissues ameliorates Th17/Th1-mediated experimental chronic graft-versus-host disease, *Journal of immunology.* **193**, 2565-73.

15. Hou, J., Zhang, Q., Fujino, M., Cai, S., Ito, H., Takahashi, K., Abe, F., Nakajima, M., Tanaka, T., Xu, J., Zou, H., Ding, Q. & Li, X. K. (2015) 5-Aminolevulinic acid with ferrous iron induces permanent cardiac allograft acceptance in mice via induction of regulatory cells, *The Journal of heart and lung transplantation : the official publication of the International Society for Heart Transplantation.* **34**,

254-63.

16. Wang, Y., Li, Y., Cao, H., Xiong, M., Shugart, Y. Y. & Jin, L. (2015) Efficient test for nonlinear dependence of two continuous variables, *BMC bioinformatics.* **16**, 260.

17. Dennis, G., Jr., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C. & Lempicki, R. A. (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery, *Genome biology.* **4**, P3.

18. Edgar, R., Domrachev, M. & Lash, A. E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository, *Nucleic acids research.* **30**, 207-10.

19. Nishio, Y., Fujino, M., Cai, S., Kitajima, Y., Saito, T., Tsumura, H., Ito, M., Ito, Y., Nagahara, Y. & Li, X. K. (2016) Impaired CD98 signaling protects against graft-versus-host disease by increasing regulatory T cells, *Transplant immunology.* **35**, 34-9.

Figure Legends:

Figure 1. *Histograms of False Positive Rate (FPR) and False Negative Rate (FNR) from the three methods under the null hypothesis (H0) and the alternative hypothesis (H1).*

(A) The FPR histogram under the null hypothesis (FN = 0). (B) The histograms under different alternative hypotheses, in which θ is equal to 10%, 25% and 50% and the simulated real positive calls is 1%, 5% and 10% of the whole simulated data, respectively. The significant threshold is 0.05.

Figure 2. *Heat maps of reproducibility analysis.*

(A) The reproducibility of top 100 significant genes by *T-test*, *Limma* and *SFC*, respectively, based on the MAQC data; (B) The reproducibility of top 1000 significant genes by the three methods based on the MAQC data; (C) the reproducibility of significant genes by the three methods based on pairwise analysis of data from the mouse cardiac graft model.

Figure 3. *Venn diagram of significant genes analyzed by SFC with the significant p-value under 0.05 after the Bonferroni correction.*

The overall numbers of significant genes in three phases are shown outside, which are

followed by numbers in the brackets showing the counts of over-expressed genes versus under-expressed ones. The circle on the top represents POD5 for heart; the circle on the left bottom represents POD5 for liver and the one on the right bottom represents POD8 for liver.

Figure 4. *Validation of the microarray data using a qRT-PCR assay in the mouse cardiac graft model and hepatic graft model.*

(A) The graph shows the cardiac mRNA levels analyzed on POD5, indicating the values of mRNAs measured in the syngeneic grafts (CONT) or allografts (D5) obtained from three individuals; (B) The graph shows the hepatic mRNA levels analyzed on POD5 and POD8, indicating the value of mRNAs measured in the syngeneic grafts (CONT) or allografts (D5 or D8) obtained from three individuals. A two-tailed unpaired t-test was used to calculate p-values comparing syngeneic grafts to allografts.

Table 1. Evaluation of three methods with the significant p-value under 0.05.

	T-test	Limma	SFC	
		H0		
FPR (%)	5.043	5.222	5.694	
FNR (%)	0.000	0.000	0.000	
Calls in total (%)	5.043	5.222	5.694	
		H1 : simulated real positive calls = 1%		
FPR (%)	6.043	5.455	5.350	θ=10%
	8.763	6.306	5.038	θ=25%
	14.255	8.600	3.990	θ=50%
FNR (%)	6.825	15.367	6.958	θ=10%
	0.783	1.933	0.058	θ=25%
	0.808	0.025	0.000	θ=50%

	6.908	6.240	6.220	θ=10%
Calls in total (%)	9.661	7.217	5.980	θ=25%
	15.098	9.507	4.943	θ=50%
H1 : simulated real positive calls = 5%				
	13.306	7.987	3.616	θ=10%
FPR (%)	32.978	17.856	1.818	θ=25%
	52.026	34.301	1.057	θ=50%
	6.942	15.283	8.224	θ=10%
FNR (%)	0.492	2.108	0.075	θ=25%
	0.699	0.020	0.000	θ=50%
	17.290	11.820	8.020	θ=10%
Calls in total (%)	36.301	21.854	6.714	θ=25%
	54.388	37.5817	5.999	θ=50%
H1 : simulated real positive calls = 10%				
	27.850	13.782	1.615	θ=10%
FPR (%)	56.305	35.345	0.626	θ=25%
	73.170	57.081	0.266	θ=50%
	7.282	15.334	9.830	θ=10%
FNR (%)	0.551	2.042	0.277	θ=25%
	0.652	0.019	0.000	θ=50%
	34.336	20.870	10.470	θ=10%
Calls in total (%)	60.619	41.606	10.535	θ=25%
	75.787	61.371	10.238	θ=50%

Table 2. The list of validated genes.

Access No.	Gene	Gene name	Fold-heart	Flod-liverD5	Fold-liverD8
NM_008337	*Ifng*	interferon gamma	1593.863	54.675	72.591
NM_010259	*Gbp2b*	guanylate binding protein 2b	1263.049	12.951	18.460
NM_013542	*Gzmb*	granzyme B	185.351	147.035	114.736

NM_008324	*Ido1*	indoleamine 2,3-dioxygenase 1	103.729	38.474	47.050
NM_011073	*Prf1*	perforin 1 (pore forming protein)	99.539	38.016	37.767
NM_008510	*Xcl1*	chemokine (C motif) ligand 1	82.096	27.777	26.918
NM_011579	*Tgtp1*	T cell specific GTPase 1	76.367	33.074	59.197
NM_021396	*Pdcd1lg2*	programmed cell death 1 ligand 2	74.231	14.479	41.463
NM_001081110	*Cd8a*	CD8 antigen, alpha chain	60.400	33.458	32.012
NM_024253	*Nkg7*	natural killer cell group 7 sequence	47.828	38.247	30.322
NM_019465	*Crtam*	cytotoxic and regulatory T cell molecule	46.089	26.296	15.863
NM_001033126	*Cd27*	CD27 antigen	33.240	39.830	41.565
NM_008798	*Pdcd1*	programmed cell death 1	29.391	74.356	69.542
NM_033078	*Klrk1*	killer cell lectin-like receptor subfamily K, member 1	28.611	18.487	16.631
NM_008530	*Ly6f*	lymphocyte antigen 6 complex, locus F	27.006	56.930	29.637
NM_011612	*Tnfrsf9*	tumor necrosis factor receptor superfamily, member 9	26.947	30.625	29.872
NM_009977	*Cst7*	cystatin F (leukocystatin)	25.625	26.383	30.931
NM_011337	*Ccl3*	chemokine (C-C motif) ligand 3	21.102	47.883	82.279
NM_013652	*Ccl4*	chemokine (C-C motif) ligand 4	19.907	35.686	56.794

Supplementary Figure legends

Supplementary Figure 1. *Distribution of mean and variance of sample microarray signals in each probe derived from the MAQC data.*

The X-axis is the means of microarray signals derived from the MAQC data and transformed by *Ln*. The Y-axis is the *Ln* values of variance.

Supplementary Figure 2. *Histograms of False Positive Rate (FPR) and False Negative Rate (FNR) from the three methods under the null hypothesis (H0) and the alternative hypothesis (H1) with the significant p-value under 0.01.*

(A) The FPR histogram under the null hypothesis (FN = 0). (B) The histograms under different alternative hypotheses, in which θ is equal to 10%, 25% and 50% and the simulated real positive calls is 1%, 5% and 10% of the whole simulated data,

474 respectively.

475

476 Supplementary Figure 3. *Histograms of False Positive Rate (FPR) and False*
477 *Negative Rate (FNR) from the three methods under the null hypothesis (H0) and the*
478 *alternative hypothesis (H1) with the significant p-value under 0.001.*
479 (A) The FPR histogram under the null hypothesis (FN = 0). (B) The histograms under
480 different alternative hypotheses, in which θ is equal to 10%, 25% and 50% and the
481 simulated real positive calls is 1%, 5% and 10% of the whole simulated data,
482 respectively.

483

484 Supplementary Figure 4. *Venn diagrams of significant gene numbers analyzed by*
485 *T-test and Limma with the significant p-value under 0.05 after the Bonferroni*
486 *correction.*
487 The numbers of significant genes in three phases are shown outside, which are
488 followed by numbers in the brackets showing the counts of over-expressed genes
489 versus under-expressed ones. The circle on the top represents POD5 for heart; the
490 circle on the left bottom represents POD5 for liver and the one on the right bottom
491 represents POD8 for liver.

492

493 Supplementary Figure 5. *Venn diagrams of significant gene numbers analyzed by*
494 *T-test, Limma and SFC with the significant p-value under 0.05.*
495 The numbers of significant genes in three phases are shown outside, which are
496 followed by numbers in the brackets showing the counts of over-expressed genes
497 versus under-expressed ones. The circle on the top represents POD5 for heart; the
498 circle on the left bottom represents POD5 for liver and the one on the right bottom
499 represents POD8 for liver. Every part of the Venn also display a counting number.

500

501 Supplementary Figure 6. *Venn diagrams of significant gene numbers analyzed by*
502 *T-test, Limma and SFC with the significant p-value under 0.001.*
503 The numbers of significant genes in three phases are shown outside, which are

followed by numbers in the brackets showing the counts of over-expressed genes versus under-expressed ones. The circle on the top represents POD5 for heart; the circle on the left bottom represents POD5 for liver and the one on the right bottom represents POD8 for liver. Every part of the Venn also display a counting number.

Supplementary Tables

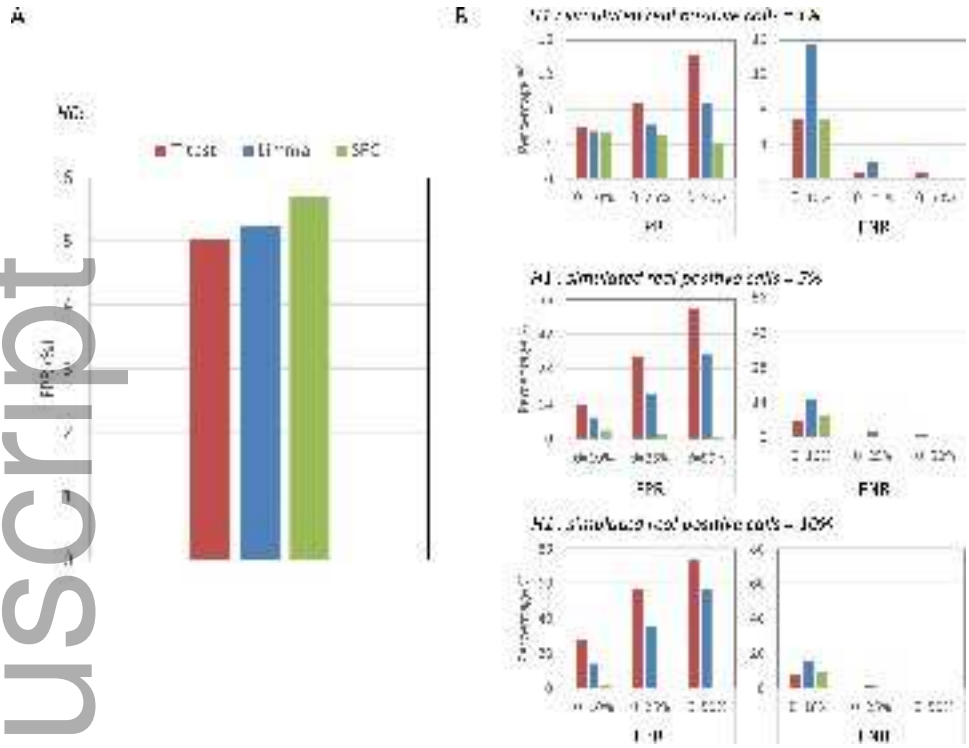
Supplementary Table 1. Evaluation of three methods with the significant p-value under 0.01.

Supplementary Table 2. Evaluation of three methods with the significant p-value under 0.001.

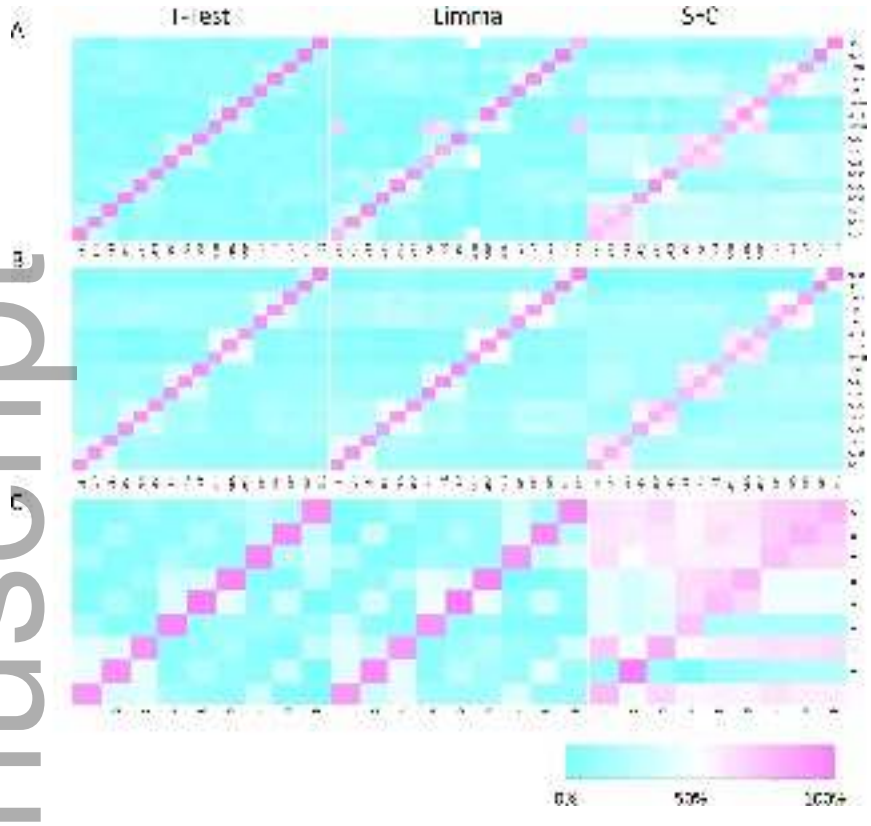
Supplementary Table 3. Primer sequences for qRT-PCR.

Supplementary Table 4. GO term and pathway enrichment analysis based on the 52 significant genes. Sheet 1: GO term enrichment analysis with three categories BP (biological process), MF (molecular function) and CC (cellular component).

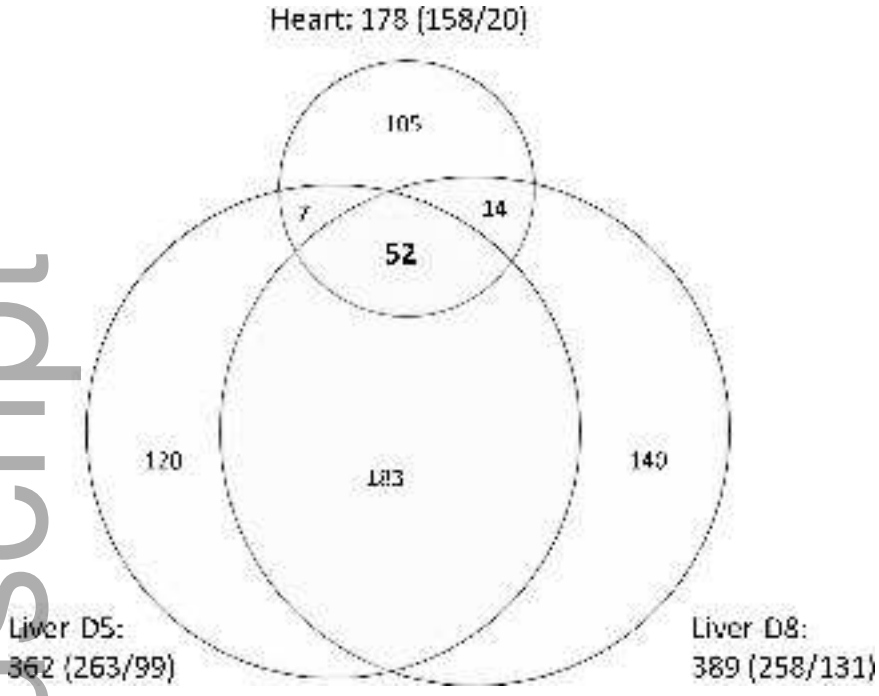
Sheet 2: Pathway enrichment analysis based on the KEGG dataset.



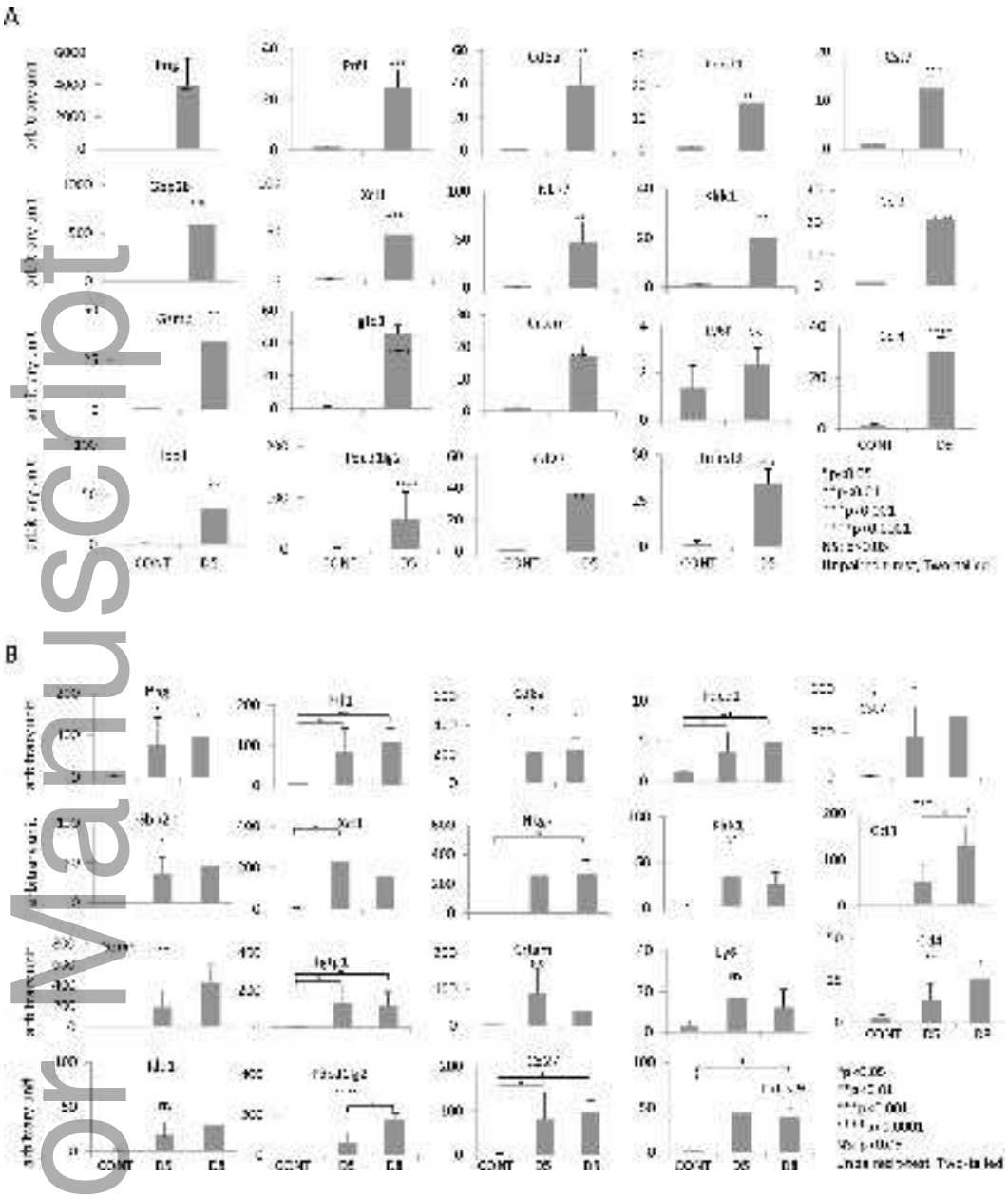
feb4_12343_f1.tiff



feb4_12343_f2.tiff



feb4_12343_f3.tiff



feb4_12343_f4.tiff