

RESEARCH ARTICLE

Open Access

Conserved expression of natural antisense transcripts in mammals

Maurice HT Ling¹, Yuguang Ban¹, Hongxiu Wen², San Ming Wang² and Steven X Ge^{1*}

Abstract

Background: Recent studies had found thousands of natural antisense transcripts originating from the same genomic loci of protein coding genes but from the opposite strand. It is unclear whether the majority of antisense transcripts are functional or merely transcriptional noise.

Results: Using the Affymetrix Exon array with a modified cDNA synthesis protocol that enables genome-wide detection of antisense transcription, we conducted large-scale expression analysis of antisense transcripts in nine corresponding tissues from human, mouse and rat. We detected thousands of antisense transcripts, some of which show tissue-specific expression that could be subjected to further study for their potential function in the corresponding tissues/organs. The expression patterns of many antisense transcripts are conserved across species, suggesting selective pressure on these transcripts. When compared to protein-coding genes, antisense transcripts show a lesser degree of expression conservation. We also found a positive correlation between the sense and antisense expression across tissues.

Conclusion: Our results suggest that natural antisense transcripts are subjected to selective pressure but to a lesser degree compared to sense transcripts in mammals.

Background

Recent studies suggest that a substantial portion of mammalian genomes are transcribed as non-coding RNA [1-4], including *cis*-natural antisense transcripts (*cis*-NATs) [5-8]. *Cis*-NATs are transcribed from the antisense counterpart of protein coding sequences, which may result in post-transcriptional gene silencing [9]. However, the extent to which *cis*-NATs are biologically functional and actively regulated remains a subject of debate [10]. Some studies had suggested that *cis*-NATs represent transcriptional noise [11,12], while others had reported supportive evidence for the function of various *cis*-NATs [13-16], especially in RNA editing [17], stability [18], and translation [19]. Growing evidence implicating a role of NATs in medical conditions, such as hypertension [20] and immune disorders [21], suggests a functional role for *cis*-NATs. However, it cannot be assumed that *cis*-NATs are as actively regulated as its sense counterparts.

Through our previous study, we developed an Antisense Transcriptome analysis using Exon array (ATE) approach for high-throughput expression analysis of NATs by using commercial oligonucleotide DNA microarrays [22]. The Affymetrix Exon array is an inexpensive high-density oligonucleotide microarray that has two unique features: (1) it has multiple probes for each of known or predicted exons, and (2) its signals are strand-specific because of the generation and labeling of single-stranded DNA targets. By modifying the recommended cDNA synthesis protocol, we demonstrated that it is possible to label targets in reverse orientation as what would be labeled according to the standard protocol (See Additional file 1: Figure S1). Thus, the cDNAs from known genes can no longer hybridize with these probes. Instead, any true hybridization signal must come from transcripts on the opposite strand, i.e., *cis*-NATs. Our preliminary microarray data on human Jurkat cells showed that the modified protocol can successfully detect a large number of NATs transcribed from known exonic loci [22]. Although limited to exonic NATs, using the Affymetrix exon arrays with our modified protocol provides a cost-efficient method to study the expression of NATs on ~ 1 million exonic loci.

* Correspondence: Xijin.Ge@sdstate.edu

¹Department of Mathematics and Statistics, South Dakota State University, Brookings, South Dakota SD 57007, USA

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

The expression patterns of protein-coding genes [23,24] and orthologous genes that are essential to the organism [25] are evolutionarily conserved. Hence, it can be implied that orthologous transcripts that demonstrate expression conservation are likely to be biologically functional. Expression divergence of randomly assigned pairs of genes, by means of permutation, had been used as a baseline to approximate a neutral evolution of gene expression [23]. If orthologous *cis*-NATs show more correlated expression patterns when compared to randomly permuted *cis*-NAT pairs, it would provide evidence that *cis*-NATs are actively regulated or subject to selective pressure.

In this study, we measured the expression of antisense transcript across human, mouse, and rat using the ATE procedure [22]. Coupled with expression analysis of sense transcripts in the same samples, this will define a “double stranded” expression profile at the exon level. We report significant differences in expression divergence between antisense orthologous transcripts when compared to permuted pairs. However, the expression divergence of sense transcripts is significantly lower than that of antisense transcripts, suggesting that *cis*-NATs are subjected to selective pressure but to a lesser degree compared to sense transcripts.

Methods

Microarray data

Total RNA were purchased from Ambion (Austin, TX, USA). The RNA samples consist of human colon, mouse embryo, rat embryo, and 9 orthologous tissues from all 3 organisms, namely, brain, heart, kidney, liver, lung, spleen, ovary, testes, and thymus. Sense and antisense expression was measured using Affymetrix Exon 1.0 ST array according to manufacturer’s protocol and previously described modification for measuring antisense expression [22] respectively. Sense and antisense arrays were normalized separately using RMA method. The log expression values and exon annotations for core exon set were extracted using Affymetrix Expression Console. The final log expression values were averaged from two replicates.

Defining orthology

Orthologous genes between human, rat, and mouse were identified from NCBI HomoloGene, build 65. The protein accession numbers were converted to gene accession numbers [26] and mapped to the microarray annotations for gene-level orthology. For exon-level orthology, three sets of exon sequences; human, mouse, and rat; were downloaded from USCS Genome Browser and used to generate BLAST databases. The exon sequence sets were compared using *blastn* and exon pairs with global sequence identity of more than 80% were considered to be orthologous. In event where one exon was found to have global sequence identity of more than 80% with more than one orthologous exons

in the same organism, the exon with the highest global sequence identity were considered to be orthologous. In addition, exons with overlapping RefSeq transcripts on both sense and antisense strands were removed.

Permutation test for expression divergence

Permutation test was used to evaluate whether the expression divergence of orthologous exons are statistically different from random. Expression divergence between 2 probesets is defined by Euclidean distance of the relative abundance, converted from microarray log expression value, using 9 orthologous tissues. At least one of the tissues must have expression higher than 6.5. Relative abundance is defined as the quotient of the microarray log expression value of the sample and the sum of the log expression values of the 9 tissues in the same set. Student’s t-test was used to test the expression divergence between the orthologous probesets and permuted pairs. Permuted pairs were generated by randomly assigning pairs of probesets from different organisms within the orthologous set. As a result, the number of orthologous pairs and permuted pairs are equal.

Identifying tissue-specific NATs

Tissue-specific probesets were identified based on a previously described method [27] using 3 empirical criteria. Firstly, the log expression value of the highest expressing tissue must be higher than 6.5 which is the threshold for a detection p-value of 0.01 above background. Secondly, the Z-score of the log expression value for the highest expressing tissue must be higher than 2. Finally, the expression level of the highest expressing tissue must be at least one log higher than that of the second highest expressing tissue. Mouse and rat embryo samples were removed from the data set before identifying tissue-specific NAT probesets for the 9 orthologous tissues in rat and mouse.

Identifying novel NATs

Novel NAT probesets were identified from the core Affymetrix exon probeset based on a previously described method [28] using BLASTN (version 2.2.25+). Each probeset was queried against RefSeq database (downloaded from NCBI on August 9, 2012) and EST database (downloaded from NCBI on May 15, 2012) for perfect matches. Query sequences without perfect matches in RefSeq database and EST database were considered to be novel. The strand option was set to “minus” to query only the reverse-complement of the query sequence (personal communication, Wayne Matten, NIH).

Strand-specific RT-PCR

Strand-specific RT-PCR was used to validate the sense-antisense transcripts candidates. Total RNA samples were purchased from Clontech and tested for genomic

DNA contamination by direct PCR at the cycling conditions of 95°C 30 seconds, 56°C 30 seconds, 72°C 45 - seconds for 38 cycles before visualized in a 2% agarose gel. For RT-PCR, sense and antisense primers were designed for each candidate using Primer3. Strand-specific reverse transcription were performed for each candidate on the conditions: RNA (100 ng) were reverse-transcribed using M-MLV Reverse Transcriptase (Invitrogen) and gene-specific sense primer (for anti-sense detection) or antisense primer (for sense detection) (2 pmole) in 20 μl volume at 65°C for 5 minutes, 37°C for 50 minutes, 70°C for 15 minutes. 5 μl of cDNA was amplified in 50 μl containing sense primer (10 pmole), antisense primer (10 pmole), Taq polymerase (1.25 U, Promega), MgCl₂ (1.5 mM) in ABI 9700 cycler under cycling conditions: 95°C for 7 minutes, 25 or 38 - cycles of 95°C for 30 seconds, 56°C for 30 seconds, 72°C for 45 seconds, 72°C for 7 minutes, and visualized in 2% agarose gels.

Results and discussion

Using the ATE protocol [22], we analyzed antisense transcription in 10 corresponding tissues for human, mouse and rat. The Affymetrix GeneChip Human Exon ST array includes 1.4 million probesets targeting exonic loci, of which 287,329 "core" probes are supported by full-length RNAs. Similar arrays for mouse and rat have 1.2 and 1 million probesets, and 231,465 and 92,751 core probesets, respectively. The same RNA sample and microarray were used to detect sense gene expression using the standard protocol. Two technical replicates were performed for each biological sample, resulting in a total of 120 hybridizations.

Thousands of antisense transcripts are detected

67,649 (23.5%), 46,050 (19.9%), and 21,742 (23.4%) of human, mouse, and rat antisense exon probesets detected expression in at least one tissue, respectively. Of these, 33,313, 31,362, 14,180 antisense probesets do not overlap

with RefSeq and EST sequences (Table 1), suggesting that these antisense transcripts are novel. This is similar to [28] who defined novelty as non-repetitive sequences that were not found in RefSeq database. The novel probeset list and sequences are given as Additional file 2: Dataset 1.

To identify antisense transcripts with tissue-specific pattern of expression, we used empirical criteria similar to [27]. Our results suggest 14,485 (5.0% of core probesets) human antisense exon probesets, 12,673 (5.5% of core probesets) mouse antisense exon probesets and 11,358 (12.2% of core probesets) rat antisense exon probesets are tissue-specific. This may suggest a biological role for these antisense transcripts and warrant further investigation.

A number of the mouse tissue-specific antisense transcripts are confirmed using strand-specific RNA-seq data [29-31] in UCSC Genome Browser. As an example, the mouse probeset ID 4628230, corresponding to the 4th exon of *inositol oxygenase* (*Miox*) gene, shows kidney-specific expression in both the Affymetrix Exon Array data and RNA-seq (Figure 1). Using strand-specific PCR targeting the 4th exon of *Miox* gene, we confirmed the kidney-specific expression in both sense and antisense transcripts (Figure 2, Additional file 1: Figure S2). However, antisense expressions are much lower than sense expression and require larger quantities of RNA samples for antisense array [22]. This resulted in the need to normalize sense and antisense arrays separately. Therefore, it is difficult to quantify the ratio of sense to antisense expression on a global scale using microarray experiments. Inositol oxygenase had been reported to be involved in osmoregulation which accounts for its kidney-specificity [32].

We choose 2 additional tissue-specific probes for strand-specific PCR verification, namely, probeset ID 4544718 which is embryo-specific, and probeset ID 4772779 which is heart-specific (Figure 3, Additional file 1: Figure S3 for primer design methodology). PCR from RNA samples without undergoing reverse transcription

Table 1 Number of novel and tissue-specific antisense transcripts

	Human	Mouse	Rat
Total number of probesets supported by RefSeq annotations (Core probesets)	287,329 (100%)	231,465 (100%)	92,751 (100%)
Antisense transcripts that are not found in RefSeq database	278,827 (97.0%)	226,996 (98.1%)	91,648 (98.8%)
Antisense transcripts that are not found in EST database	140,496 (48.9%)	162,379 (70.2%)	66,943 (72.2%)
Antisense transcripts that are not found in both RefSeq or EST databases	139,918 (48.7%)	161,865 (69.9%)	66,865 (72.1%)
Antisense transcripts that are expressed in at least one tissue and not found in RefSeq or EST databases (novel antisense transcripts)	33,313 (11.6%)	31,362 (13.5%)	14,180 (15.3%)
Tissue-specific transcripts	14,485 (5.0%)	12,673 (5.5%)	11,358 (12.2%)
Tissue-specific transcripts and not found in RefSeq or EST database (novel, tissue-specific transcripts)	5,287 (1.8%)	7,652 (3.3%)	7,411 (8.0%)
Antisense transcripts that are expressed in at least one tissue and not found in RefSeq database (This is used for expression divergence analysis, Figure 5)	67,649 (23.5%)	46,050 (19.9%)	21,742 (23.4%)

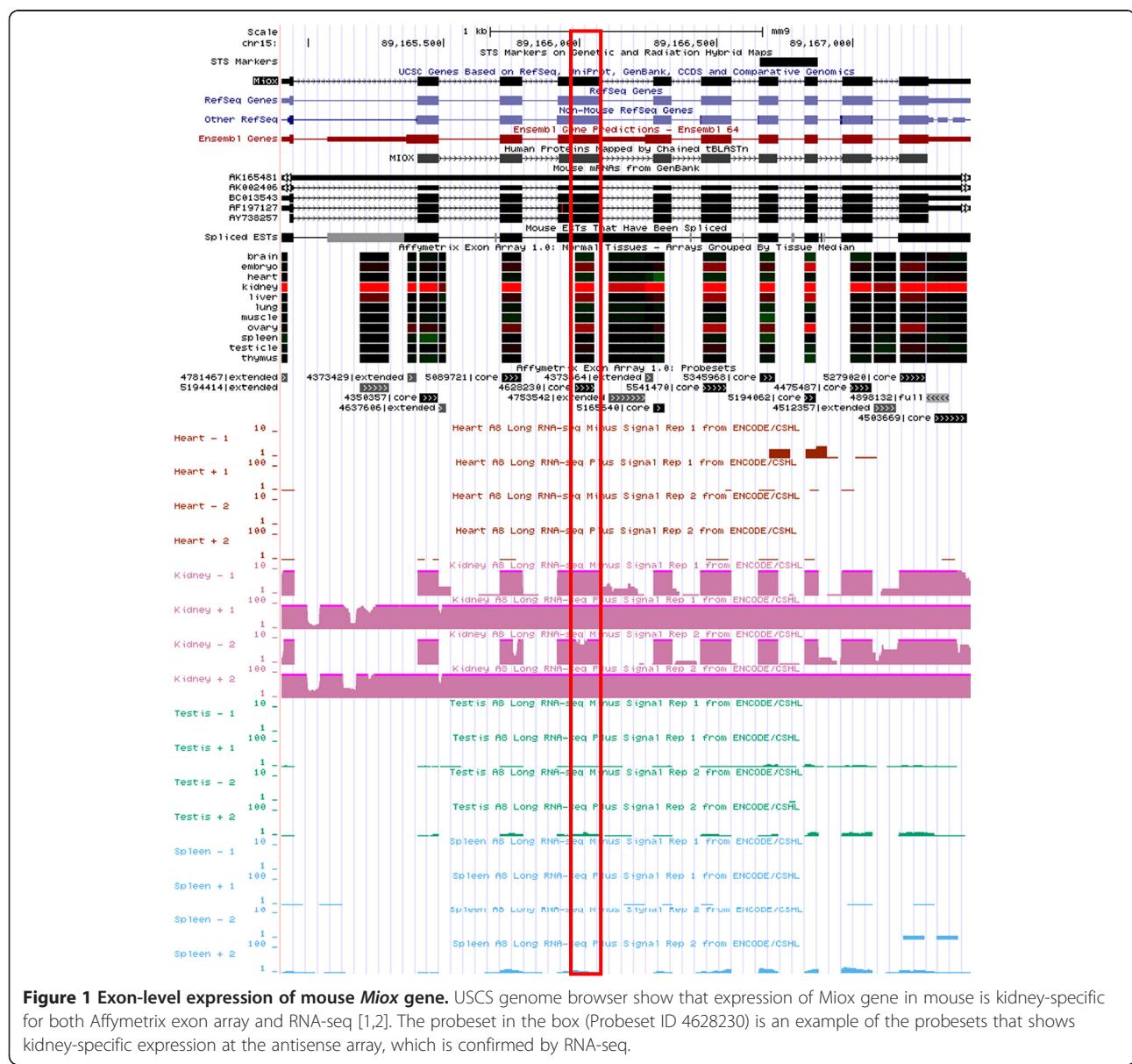
does not yield any visible bands (Additional file 1: Figure S4), which indicates no genomic DNA contamination in the RNA samples. Our results show that probeset ID 4772779 is only expressed in mouse heart and all 3 probes did not detect expression in mouse brain at the expected size. The PCR was performed with 25 and 38 extension cycles; hence, the lack of a strong band acts as strong support for lack of expression. Thus, experimental validation using strand-specific PCR supports the validity of our empirical criteria for the selection of tissue-specific antisense expression.

We had identified an example of human long noncoding RNA expression (Additional file 1: Figure S5) identified in NONCODE [33], which forms a sense-antisense pair at the 5'-end of *Ubiquitin-specific protease 25*

(*USP25*). *USP25* is involved in degrading mis-folded proteins in the endoplasmic reticulum [34]. This suggests that the expression of *USP25* may be modulated by anti-sense transcript.

Further validation of antisense transcript detection protocol

The average expression of antisense transcripts is lower than that of sense transcripts at both the gene and exon level (Additional file 1: Figures S6-S11). This is consistent with that of [22] who reported that the hybridization control probes were higher in antisense arrays, suggesting that the actual expression levels of transcripts on the antisense arrays were lower than sense arrays. The proportion of core exon probesets detected above background (DABG)



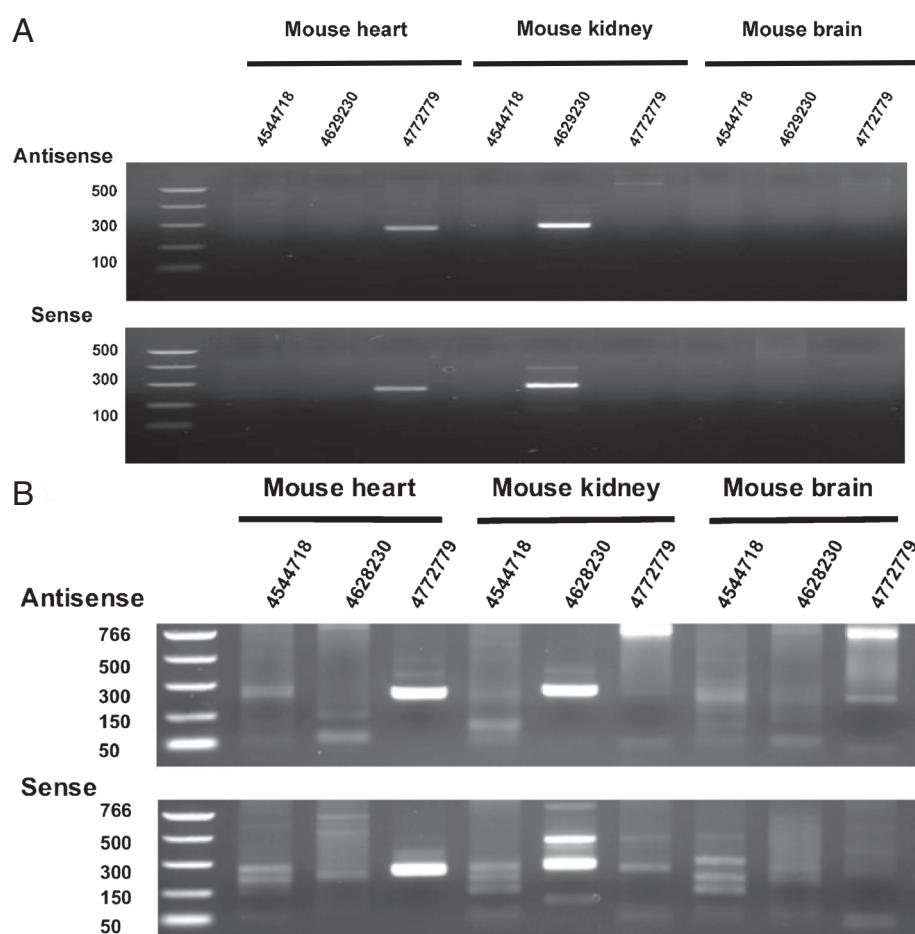


Figure 2 Strand-specific PCR of 3 tissue-specific mouse probesets. Panel **A** shows 25 cycles of amplification. Panel **B** showed 38 cycles of amplification. Probesets ID 4544718, 4628230, and 4772779 showed tissue-specific expressions to mouse embryo, mouse kidney, and mouse heart respectively. Probesets ID 4544718 targets the last exon of *myosin regulatory light chain 2, skeletal muscle (Mylpf)* gene. Probesets ID 4628230 targets the 4th exon of *inositol oxygenase (Miox)* gene. Probesets ID 4772779 targets the 2nd exon of *creatine kinase M-type (Ckm)* gene. Microarray expression data for these 3 probesets are given in Figure 3.

with a detection p-value less than 0.01 (Additional file 1: Figure S12) were consistent with previous report [22]. Our results also suggest that probesets with DABG p-value of less than 0.01 have an intensity of at least 6.5.

Two technical replicates were performed for each tissue and our results show that core probeset intensities of the technical replicates are strongly correlated ($0.86 < r < 0.99$) in both sense and antisense arrays (Additional file 1: Figures S13-S24), suggesting that the experimental procedure is reproducible. The average intensity values from the two technical replicates were used for analysis.

For further validation of ATE procedure, we examined pre-designed probes for known antisense transcripts. We searched for extended probesets at the same genic location to core probesets but are located on the opposite strand. Some antisense transcripts should be detected

by these extended probesets in the sense arrays, as well as the core probesets in the antisense array. We found positive correlation ($0.50 < r < 0.58$) between the expression level of core probesets in antisense array and the overlapping extended probeset in sense array (Additional file 1: Figure S25, S26). This is expected as the extended probesets in the sense array and the sense probesets in the antisense array are detecting the same antisense transcript.

Sense gene expression is conserved as previously reported

To quantify expression divergence, we used Euclidean distance to measure the dissimilarity between the expression patterns across multiple tissues [23,35]. The expression divergence between orthologous genes is significantly

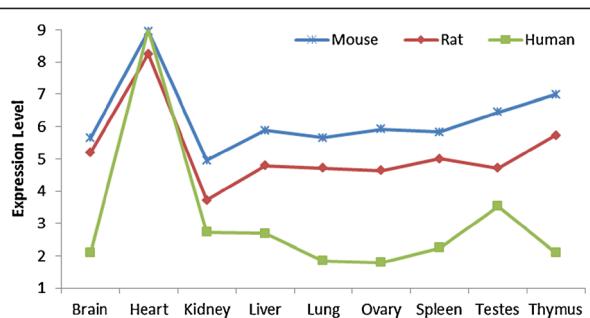


Figure 3 Expression profile of an antisense heart-specific transcript (*creatine kinase M-type gene*) across all 3 species. The mouse, rat, and human probesets are probesets ID 4772779, 5823307, and 3865287 respectively. The correlations between mouse/rat and mouse/human are 0.96 and 0.85 respectively, demonstrating expression conservation in all 3 species.

lower than permuted pairs (Figure 4, Figure 5A) and the difference is correlated to evolutionary time between the two species (Additional file 1: Figure S27). In addition, we found stronger correlation between expression profiles of orthologous tissues than non-orthologous tissues (Figure 4). This corroborates [23] and validates our experimental and analytical procedures. Our results showed poor correlation ($r < 0.167$; Additional file 1: Figure S28)

between the 2 measures used to calculate expression divergence, namely, Euclidean distance and Pearson's correlation. Nevertheless, our results are robust (Additional file 1: Figure S27) despite using different expression divergence measures (Euclidean distance and Pearson's correlation) or different intensity thresholds (full set of orthologous genes without threshold, or intensity of at least one orthologous gene to be above 6.5).

We extended the same analysis to exon-level expression data. Despite using different intensity thresholds to calculate expression divergence (Figure 4, Figure 5B, and Additional file 1: Figure S30), our results show that the expression divergence between orthologous sense exon transcripts is significantly lower than permuted pairs. Our results consistently demonstrate a stronger correlation between expression profiles of orthologous tissues than non-orthologous tissues as shown by the diagonal lines representing orthologous tissues between 2 organisms on the heatmap (Figure 4). As an example, the Pearson's correlation between mouse and rat kidney is 0.72 while the correlation between mouse kidney and human brain is 0.44 (Additional file 1: Figure S31). Thus, our results validated our experiment and suggested that our analytical procedures are suitable for studying expression divergence in exon transcripts.

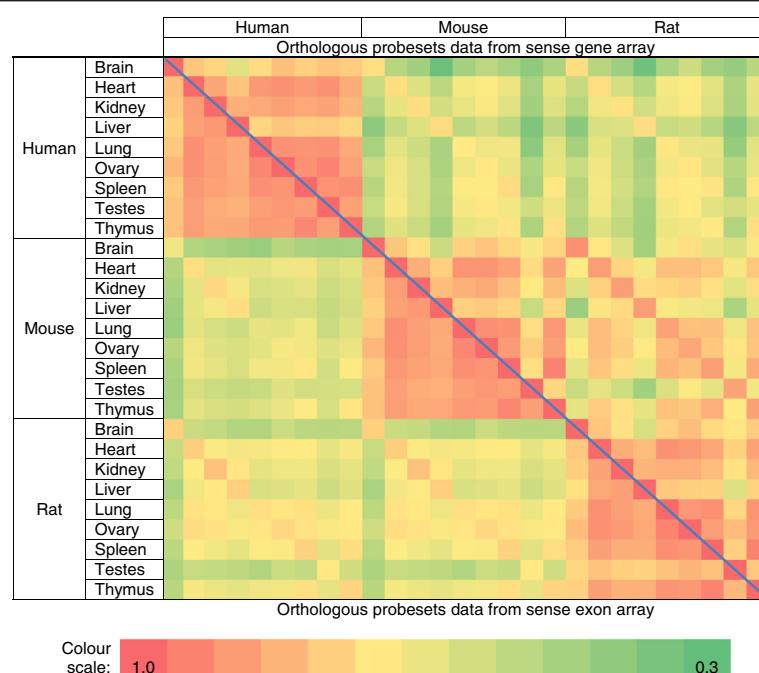


Figure 4 Heatmaps of orthologous sense exon and gene probesets. The colours represent the Pearson's correlation coefficient of the orthologous probesets across any 2 tissues. The correlations for orthologous exons were calculated from exons probesets that are present across all 3 species based on 80% DNA identity. The correlations for orthologous genes were calculated from gene probesets that are present across all 3 species based on NCBI HomoloGene build 65. Pearson's correlation ranged from 0.3 (green) to 1.0 (orange). The lower and upper triangles are heatmaps using data from exon and gene probesets respectively.

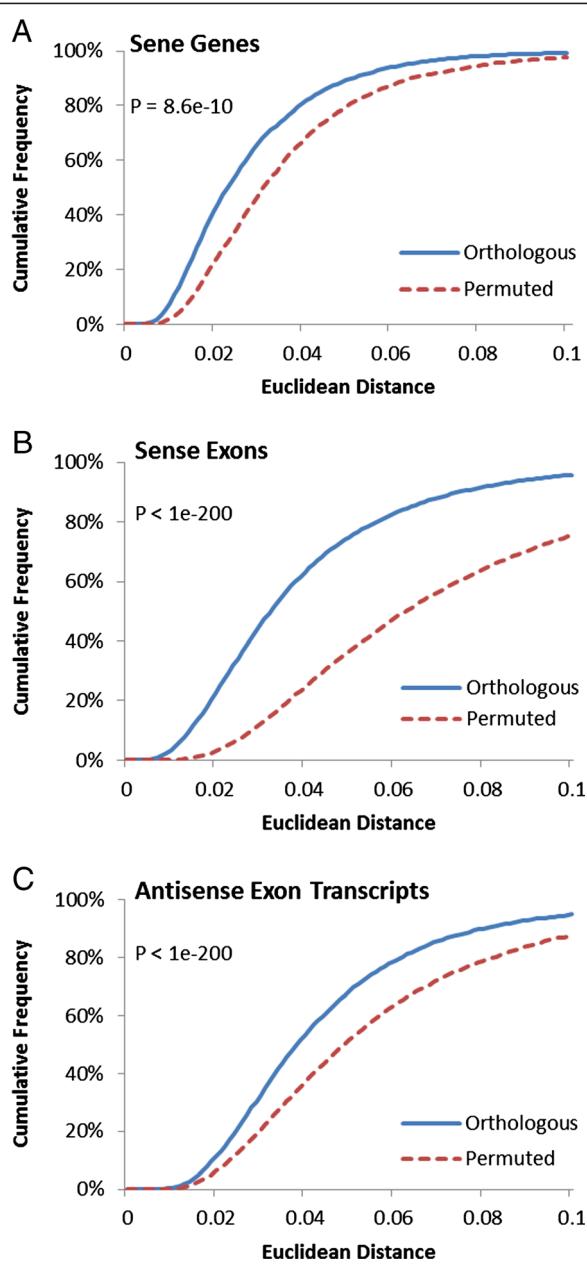
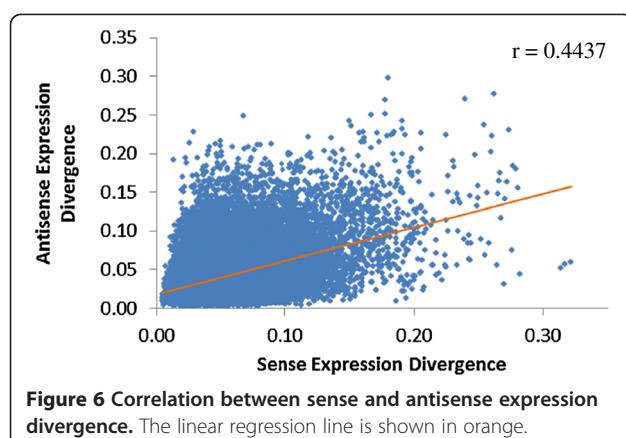


Figure 5 Cumulative frequency of mouse/rat orthologous transcripts at gene and exon level against permuted pairs. Vertical axis represents cumulative frequency. Horizontal axis represents expression divergence, calculated using Euclidean distance. Panel **A** shows the expression divergence distribution of protein-coding genes as defined by NCBI HomoloGene build 65 and its permuted pairs. Panel **B** shows the orthologous sense exons. Panel **C** shows antisense exon transcripts at the exon level excluding exons with antisense RefSeq transcripts. Probability density plots are given in Additional file 1: Figure S33.

Antisense transcript expressions suggest selective pressure

The same method for analyzing sense expression divergence at the exon level was used to analyze antisense expression divergence. Expression signal detected by each exon-specific probeset on the antisense array was treated as one antisense transcript. Using BLAST, we identified 7,460, 5,836 and 18,346 orthologous antisense transcripts for human/mouse, human/rat and mouse/rat comparisons, respectively. Student's t-test with unequal variance between the expression divergence of orthologous antisense transcripts and that of permuted pairs suggests that the expression divergence of orthologous antisense transcripts are significantly lower from their respective permuted pairs in all 3 comparisons (p -value $< 10^{-8}$; Figure 5C; Additional file 1: Figures S32A, S33A, and S35). Using different intensity thresholds (full set of orthologous exons without threshold, intensity of at least one orthologous probeset to be above 6.5 which is the intensity threshold for DABG p -value of less than 0.01, or intensity of at least one orthologous probeset to be above 8) also showed that the expression divergence of orthologous exons for antisense transcripts are significantly lower. At the same time, the Pearson's correlations of expression divergence between permuted pairs (Additional file 1: Figures S32B, and S33B) are close to zero which is similar to that reported by [36]. However, it had been suggested that Pearson's correlation and Euclidean distance can produce different results [37] and a novel randomization procedure had been proposed recently [38]. Using the randomization procedure proposed by [38], we arrive at the same conclusion that the expression divergence of orthologous antisense transcripts are significantly lower than their respective permuted pairs regardless of intensity thresholds (Additional file 1: Figure S34). This suggests that our results are not artifacts due to the use of Euclidean distance, different randomization methods or intensity thresholds. Therefore, our results suggest that the expressions of antisense transcripts are under selective pressure and their expressions are evolutionarily conserved which in agreement with the roles of antisense transcripts proposed by a large number of recent studies across different species (reviewed in [39,40]).

Our results show that the average difference between the expression divergence of the sense orthologs and its permuted counterparts is larger than the average difference between that of the antisense orthologs and its permuted counterparts (Figures 5B and 5C). This may imply on the extent of selective pressure on antisense transcripts. Permuted pairs approximate a purely neutral evolution of gene expression without selective pressure [23]. Hence, the deviation from permuted pairs may approximate the strength of selective pressure on top of a neutral background. Therefore, our results suggest that antisense expression is subjected to less selective pressure compared to sense expression.



Positive correlation between sense and antisense expression
Our results showed positive correlation between the sense and antisense expressions for all 30 tissues (Additional file 1: Figure S36-S39), demonstrating consistency across the 3 species. The Pearson's correlation coefficient ranged from 0.333 (human brain) to 0.627 (rat thymus). Our result also showed positive correlation between sense and antisense divergence (Figure 6, $r = 0.4437$). This corroborates studies suggesting that sense and antisense expressions are positively correlated [36,41]. [36] reported that more sense/antisense pairs were positively correlated than negatively correlated, leading [36] to a conclusion that negative correlation between sense/antisense pairs are rare events. This is supported by [41] showing positive correlation between sense/antisense pairs in human breast epithelium and verified these findings with strand-specific quantitative PCR. However, our results corroborates [36], showing a large range of correlations between sense and antisense expressions. This may suggest a degree of independence between the regulation of sense and antisense expression. In addition, [36] reported that sense transcript detection occurred about 10 cycles earlier than antisense transcript detection using strand-specific quantitative PCR and [31] had reported lower antisense expression than their protein coding counterparts. This suggests even though both sense and antisense transcripts are present at the same time, there is a net abundance of sense transcripts.

Consistent positive correlation between sense and antisense divergence may suggest an over-arching regulatory mechanism for both sense and antisense expression while the large proportion of uncorrelated sense and antisense expression may suggest a layer of independent regulation. Chromatin structure had been shown to affect the accessibility of RNA polymerase II and other transcription factors to the site of transcription by means of methylation and acetylation [42,43]; thereby, playing an important factor in regulating gene expression. However, the role of chromatin structure in antisense transcript regulation had only been recently reported by [10] using chromatin

immuno-precipitation and demonstrated positive correlation between *cis*-NAT promoter activity, the presence of RNA polymerase II histones modification and the resulting antisense RNA-seq read density, suggesting that chromatin structure also may be involved in *cis*-NAT transcription.

Conclusion

In summary, we analyzed the expression of antisense transcripts in corresponding tissues across 3 species and found evidence for the conserved expression of these transcripts, similar to what have been observed in protein-coding genes. This supports the idea that expression of antisense transcripts is regulated and subject to selection pressure. Our result is based on a large number of antisense transcripts, supplementing previous studies of the functions of specific antisense transcripts. In addition, the tissue-specific expression pattern of some antisense transcripts might guide future in-depth study of their potential function.

One could argue that the conserved expression of natural antisense transcripts might be a by-product of the regulated and conserved expression of the corresponding protein-coding genes. Chromatin remodelling is one mechanism of gene regulation that makes the DNA sequences accessible to transcriptional protein complexes. It is possible for antisense transcripts to get a "free ride" when this happens in a regulated manner. Further studies are necessary to fully address these possibilities, especially using new technologies like strand-specific RNA-Seq.

Additional files

Additional file 1: Figures S1-S39, Supplementary methods and additional figures.

Additional file 2: Supplementary Dataset 1. List of novel antisense transcripts, with associated sequences, identified in this study.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

ML performed the bioinformatic analysis and drafted the manuscript. YB contributed the randomization procedures and participated in bioinformatics analyses. HW validated the results using PCR. SMW and SXG conceived and oversaw this project. All authors read and approved the final manuscript.

Acknowledgement

The authors would like to thank Administrative and Research Computing at South Dakota State University for providing computational resources. We also thank Dr. Jaejung Kim of the Genomics Core Facility at University of Chicago for conducting the customized DNA microarray experiments.

Data access

Microarray data for this work had been deposited in NCBI GEO as GSE41462 (antisense data) and GSE41464 (sense data). The sense and antisense expression data for various normal tissues in human, mouse and rat will also be accessed at UCSC Genome Browser as a custom track.

Funding

This work was supported by National Institute of Health (GM083226) to SXG.

Author details

¹Department of Mathematics and Statistics, South Dakota State University, Brookings, South Dakota SD 57007, USA. ²Department of Genetics, Cell Biology and Anatomy, University of Nebraska Medical Center, Omaha, Nebraska NE 68198, United States of America.

Received: 5 November 2012 Accepted: 6 March 2013

Published: 12 April 2013

References

- Werner A, Carlile M, Swan D: What do natural antisense transcripts regulate? *RNA Biol* 2009, **6**:43–48.
- Gustincich S, Sandelin A, Plessy C, Katayama S, Simone R, Lazarevic D, Hayashizaki Y, Carninci P: The complexity of the mammalian transcriptome. *J Physiol* 2006, **575**:321–332.
- Zheng R, Shen Z, Tripathi V, Xuan Z, Freier SM, Bennett CF, Prasanth SG, Prasanth KV: Polypurine-repeat-containing RNAs: a novel class of long non-coding RNA in mammalian cells. *J Cell Sci* 2010, **123**:3734–3744.
- Au PC, Zhu QH, Dennis ES, Wang MB: Long non-coding RNA-mediated mechanisms independent of the RNAi pathway in animals and plants. *RNA Biol* 2011, **8**:404–414.
- Stamatoyannopoulos JA, Snyder M, Hardison R, Ren B, Gingras T, Gilbert DM, Groudine M, Bender M, Kaul R, Canfield T, et al: An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biol* 2012, **13**:418.
- Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, et al: GENCODE: the reference human genome annotation for the ENCODE project. *Genome Res* 2012, **22**:1760–1774.
- Chiba M, Miura T, Kasai K, Monzen S, Kashiwakura I, Yasue H, Nakamura T: Identification of up-regulated and down-regulated cis-natural antisense transcripts in the human B lymphoblastic cell line IM-9 after X-ray irradiation. *Mol Med Report* 2012, **5**:1151–1157.
- Zhang Y, Liu XS, Liu QR, Wei L: Genome-wide in silico identification and analysis of cis natural antisense transcripts (cis-NATs) in ten species. *Nucleic Acids Res* 2006, **34**:3465–3475.
- Li K, Ramchandran R: Natural antisense transcript: a concomitant engagement with protein-coding transcript. *Oncotarget* 2010, **1**:447–452.
- Conley AB, Jordan IK: Epigenetic regulation of human cis-natural antisense transcripts. *Nucleic Acids Res* 2012, **40**:1438–1445.
- Wang J, Zhang J, Zheng H, Li J, Liu D, Li H, Samudrala R, Yu J, Wong GK: Mouse transcriptome: neutral evolution of ‘non-coding’ complementary DNAs. *Nature* 2004, **431**:1–757. discussion following 757.
- Struhl K: Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat Struct Mol Biol* 2007, **14**:103–105.
- Werner A, Sayer JA: Naturally occurring antisense RNA: function and mechanisms of action. *Curr Opin Nephrol Hypertens* 2009, **18**:343–349.
- Cui I, Cui H: Antisense RNAs and epigenetic regulation. *Epigenomics* 2010, **2**:139–150.
- Nishizawa M, Okumura T, Ikeya Y, Kimura T: Regulation of inducible gene expression by natural antisense transcripts. *Front Biosci* 2012, **17**:938–958.
- Faghhi MA, Kocera J, Modarresi F, Engstrom PG, Chalk AM, Brothers SP, Koesema E, St Laurent G, Wahlestedt C: RNAi screen indicates widespread biological function for human natural antisense transcripts. *PLoS One* 2010, **5**:e13177.
- Le Mercier P, Garcin D, Hausmann S, Kolakofsky D: Ambisense sendai viruses are inherently unstable but are useful to study viral RNA synthesis. *J Virol* 2002, **76**:5492–5502.
- Stazic D, Lindell D, Steglich C: Antisense RNA protects mRNA from RNase E degradation by RNA-RNA duplex formation during phage infection. *Nucleic Acids Res* 2011, **39**:4890–4899.
- Hu X, Hipolito S, Lynn R, Abraham V, Ramos S, Wong-Staal F: Relative gene-silencing efficiencies of small interfering RNAs targeting sense and antisense transcripts from the same genetic locus. *Nucleic Acids Res* 2004, **32**:4609–4617.
- Michael DR, Phillips AO, Krupa A, Martin J, Redman JE, Altaher A, Neville RD, Webber J, Kim MY, Bowen T: The human hyaluronan synthase 2 (HAS2) gene and its natural antisense RNA exhibit coordinated expression in the renal proximal tubular epithelial cell. *J Biol Chem* 2011, **286**:19523–19532.
- Hatzoglou A, Deshayes F, Madry C, Lapree G, Castanas E, Tsapis A: Natural antisense RNA inhibits the expression of BCMA, a tumour necrosis factor receptor homologue. *BMC Mol Biol* 2002, **3**:4.
- Ge X, Rubinstein WS, Jung YC, Wu Q: Genome-wide analysis of antisense transcription with Affymetrix exon array. *BMC Genomics* 2008, **9**:27.
- Jordan IK, Marino-Ramirez L, Koonin EV: Evolutionary significance of gene expression divergence. *Gene* 2005, **345**:119–126.
- Schreiber AW, Sutton T, Caldo RA, Kalashyan E, Lovell B, Mayo G, Muehlbauer GJ, Druka A, Waugh R, Wise RP, et al: Comparative transcriptomics in the triticeae. *BMC Genomics* 2009, **10**:285.
- Bergmiller T, Ackermann M, Silander OK: Patterns of evolutionary conservation of essential genes correlate with their compensability. *PLoS Genet* 2012, **8**:e1002803.
- Alibes A, Yanklevich P, Canada A, Diaz-Uriarte R: IDconverter and IDClight: conversion and annotation of gene and protein IDs. *BMC Bioinforma* 2007, **8**:9.
- Ge X, Yamamoto S, Tsutsumi S, Midorikawa Y, Ihara S, Wang SM, Aburatani H: Interpreting expression profiles of cancers by genome-wide survey of breadth of expression in normal tissues. *Genomics* 2005, **86**:127–141.
- Li YY, Qin L, Guo ZM, Liu L, Xu H, Hao P, Su J, Shi Y, He WZ, Li YX: In silico discovery of human natural antisense transcripts. *BMC Bioinforma* 2006, **7**:18.
- Jiang L, Schlesinger F, Davis CA, Zhang Y, Li R, Salit M, Gingras TR, Oliver B: Synthetic spike-in standards for RNA-seq experiments. *Genome Res* 2011, **21**:1543–1551.
- Rosenblom KR, Dreszer TR, Pheasant M, Barber GP, Meyer LR, Pohl A, Raney BJ, Wang T, Hinrichs AS, Zweig AS, et al: ENCODE whole-genome data in the UCSC genome browser. *Nucleic Acids Res* 2010, **38**:D620–625.
- Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, et al: The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* 2012, **22**:1775–1789.
- Nayak B, Xie P, Akagi S, Yang Q, Sun L, Wada J, Thakur A, Danesh FR, Chugh SS, Kanwar YS: Modulation of renal-specific oxidoreductase/myo-inositol oxygenase by high-glucose ambience. *Proc Natl Acad Sci USA* 2005, **102**:17952–17957.
- Bu D, Yu K, Sun S, Xie C, Skogerbö G, Miao R, Xiao H, Liao Q, Luo H, Zhao G, et al: NONCODE v3.0: integrative annotation of long noncoding RNAs. *Nucleic Acids Res* 2012, **40**:D210–215.
- Blount JR, Burr AA, Denuc A, Marfany G, Todi SV: Ubiquitin-specific protease 25 functions in endoplasmic reticulum-associated degradation. *PLoS One* 2012, **7**:e36542.
- Liao BY, Zhang J: Evolutionary conservation of expression profiles between human and mouse orthologous genes. *Mol Biol Evol* 2006, **23**:530–540.
- Oeder S, Mages J, Flieck P, Lang R: Uncovering information on expression of natural antisense transcripts in Affymetrix MOE430 datasets. *BMC Genomics* 2007, **8**:200.
- Pereira V, Waxman D, Eyre-Walker A: A problem with the correlation coefficient as a measure of gene expression divergence. *Genetics* 2009, **183**:1597–1600.
- Piascik B, Robinson-Rechavi M, Bergmann S: Correcting for the bias due to expression specificity improves the estimation of constrained evolution of expression between mouse and human. *Bioinformatics* 2012, **28**:1865–1872.
- Chen HM, Neiman AM: A conserved regulatory role for antisense RNA in meiotic gene expression in yeast. *Curr Opin Microbiol* 2011, **14**:655–659.
- Werner A, Swan D: What are natural antisense transcripts good for? *Biochem Soc Trans* 2010, **38**:1144–1149.
- Grigoriadis A, Oliver GR, Tanney A, Kendrick H, Smalley MJ, Jat P, Neville AM: Identification of differentially expressed sense and antisense transcript pairs in breast epithelial tissues. *BMC Genomics* 2009, **10**:324.
- Paranjape SM, Kamakaka RT, Kadonaga JT: Role of chromatin structure in the regulation of transcription by RNA polymerase II. *Annu Rev Biochem* 1994, **63**:265–297.
- Kamakaka RT, Kadonaga JT: Biochemical analysis of the role of chromatin structure in the regulation of transcription by RNA polymerase II. *Cold Spring Harb Symp Quant Biol* 1993, **58**:205–212.

doi:10.1186/1471-2164-14-243

Cite this article as: Ling et al.: Conserved expression of natural antisense transcripts in mammals. *BMC Genomics* 2013 **14**:243.

Supplementary Figures

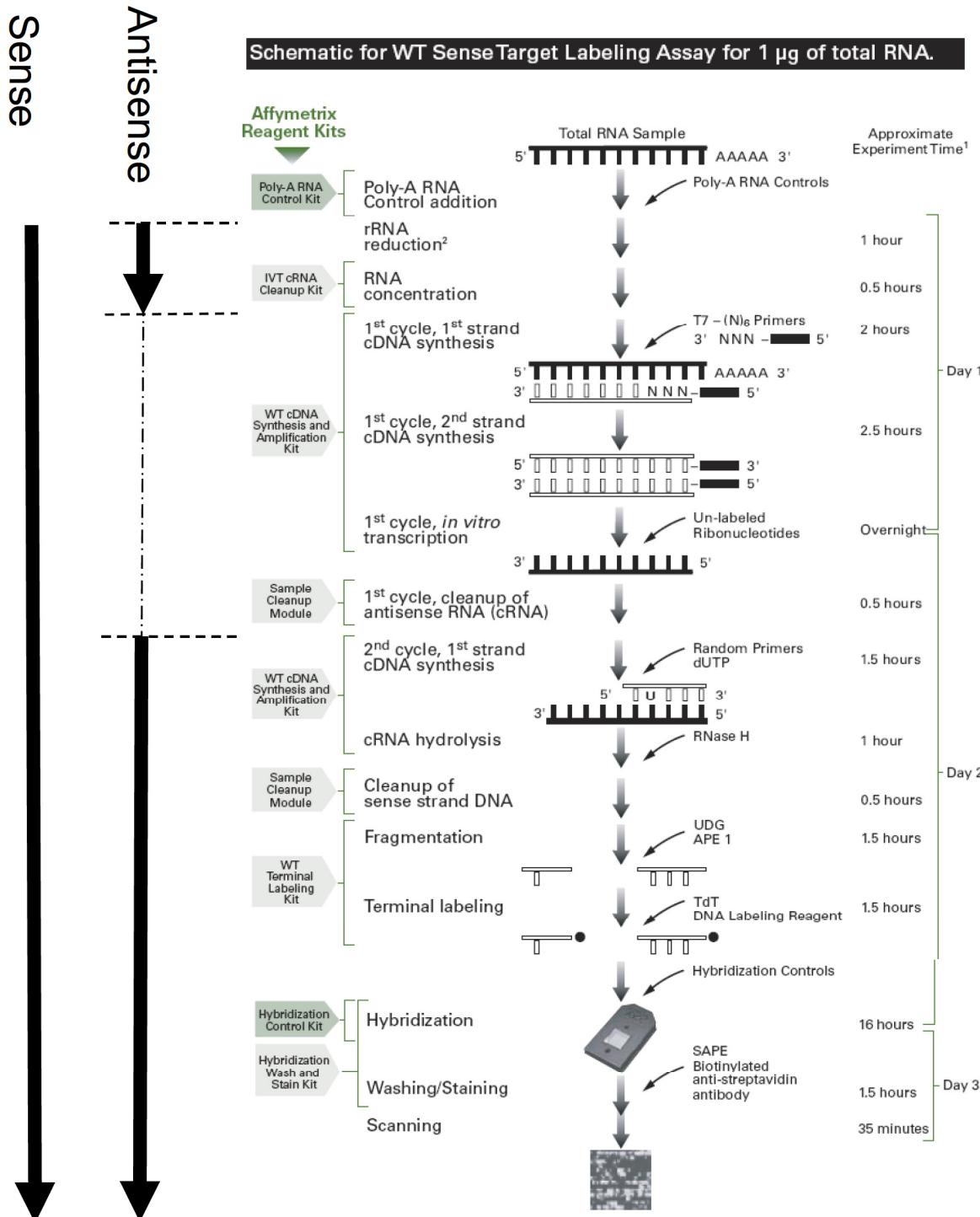


Figure S1. The modified protocol compared with standard Affymetrix protocol. In our antisense transcriptome analysis using Exon array (ATE) protocol, the first cycle cDNA synthesis and *in vitro* transcription is skipped. The image on the right is courtesy of Affymetrix, Inc.

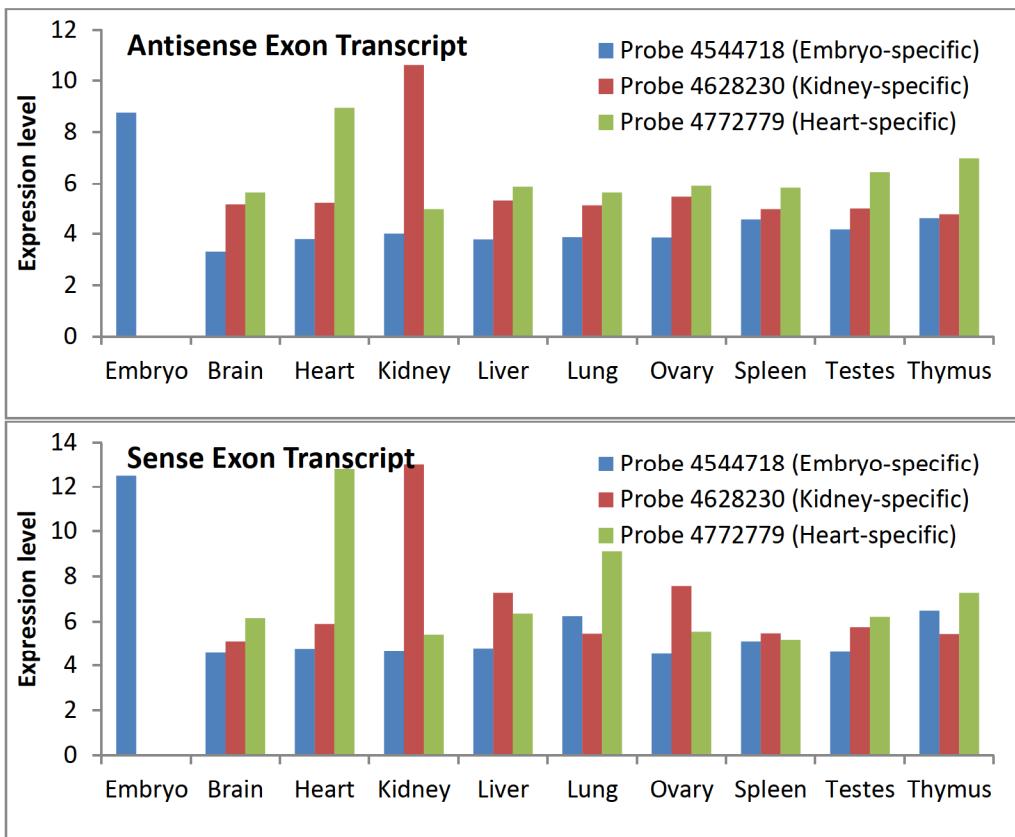
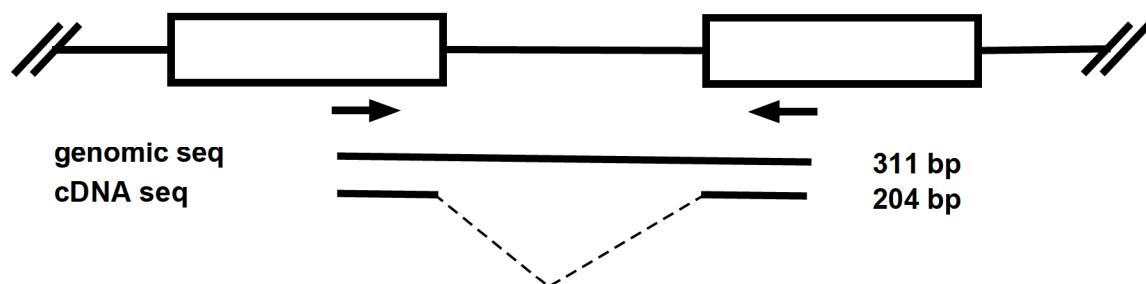
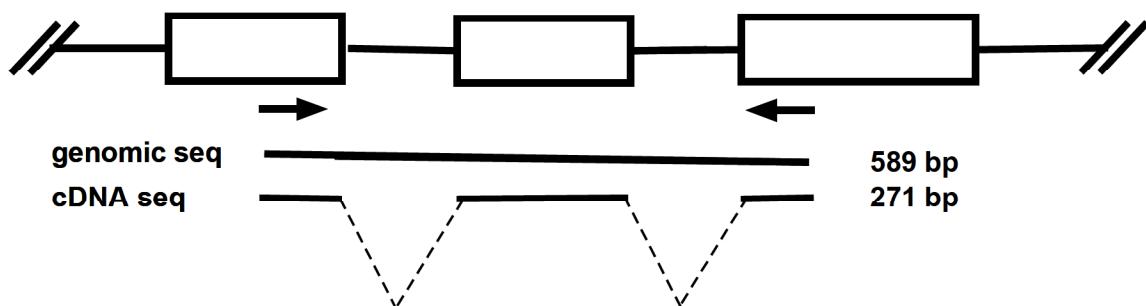


Figure S2. Expression data from microarray for probesets used in strand-specific PCR validation (Figure 4). All 3 probesets are from mouse exon array.

A. Probe set ID 4544718



B. Probe set ID 4628230



C. Probe set ID 4772779

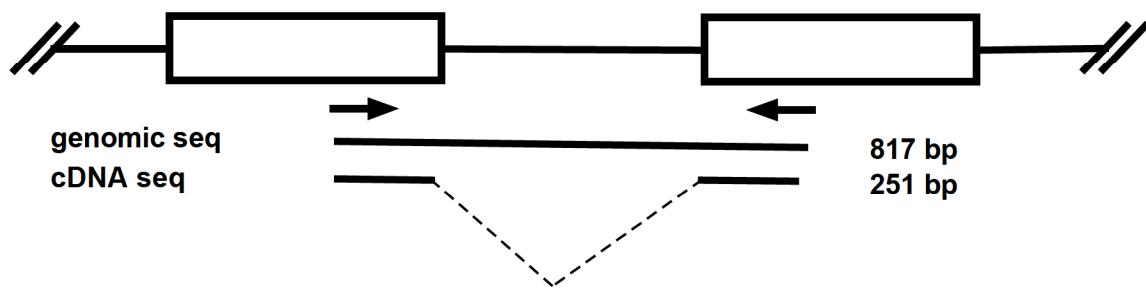


Figure S3. Primer design for strand-specific RT-PCR. For each targeted transcript, the corresponding probe sequence in array was used to map to the reference genome. Sense primer and antisense primer were designed in different exons across one intron (probe 4544718 and 4772779) or two introns/one exon (probe 4628230). Different size from mRNA and genomic DNA for each targeted sequence will determine if the RT-PCR products were from mRNA templates or from genomic DNA contamination. In fact, the sizes of strand-specific RT-PCR fit with the size of corresponding mRNA; therefore, rule out the possibility of genomic DNA contamination.

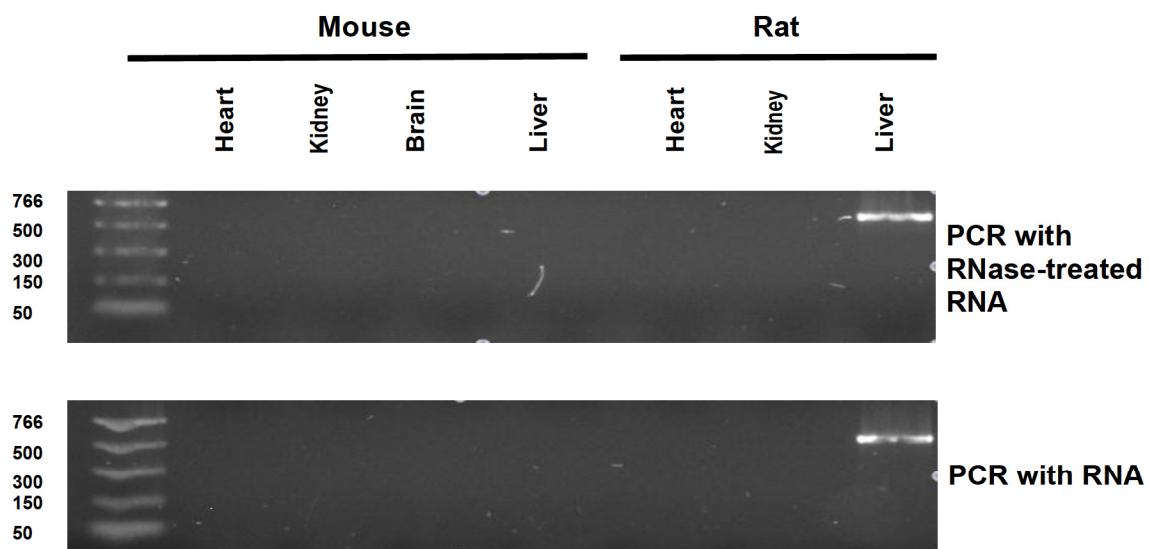


Figure S4. PCR of RNA samples. The issue of genomic DNA contamination: All the RNA samples used for the study were pre-tested for possible DNA contamination. The upper panel shows PCR result using RNase A-treated RNA samples as the templates, the lower panel shows PCR results using RNA directly as the templates. In both tests, all RNA samples show no sign of genomic DNA contamination except one (rat liver RNA), which was not used in validation study.

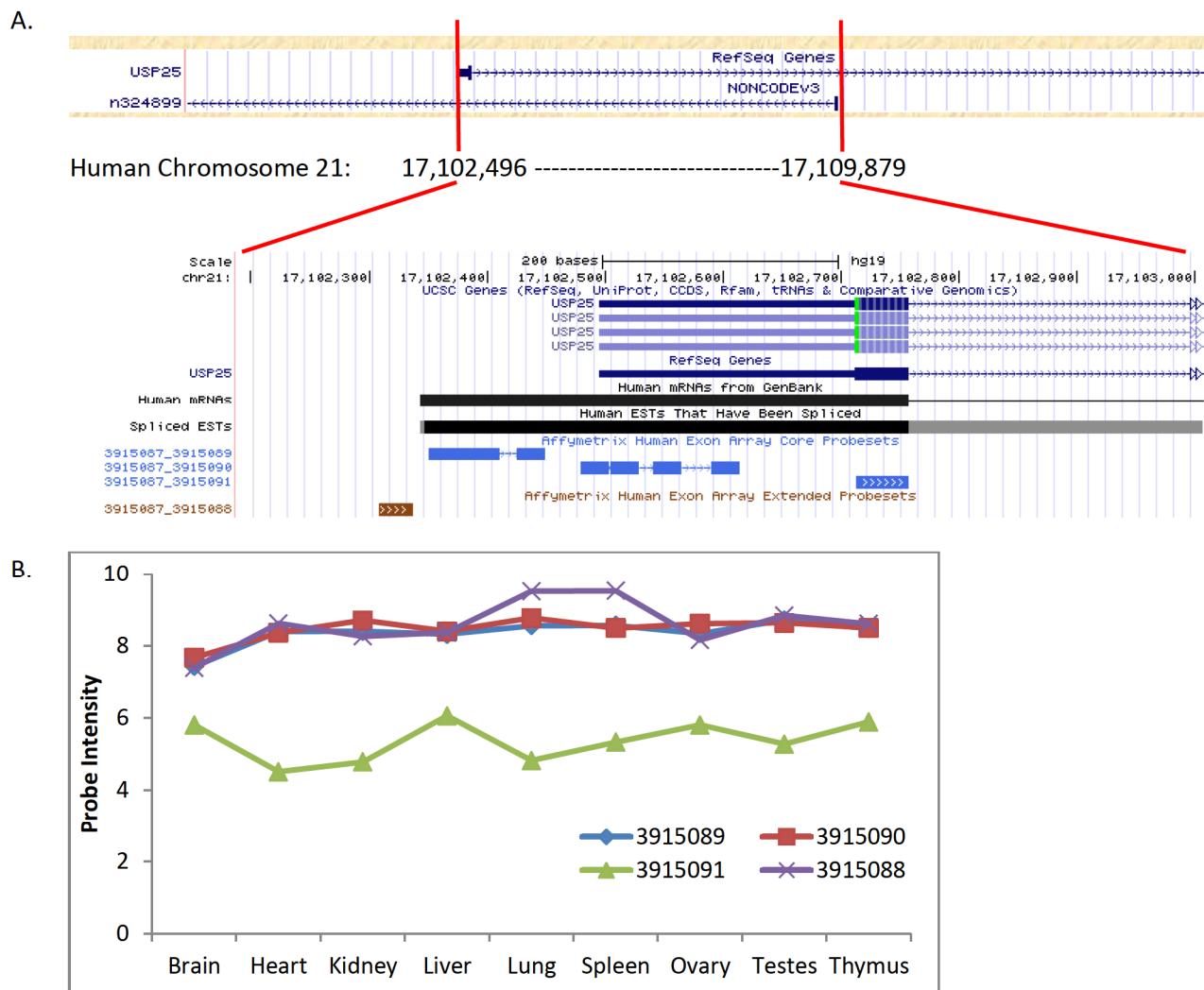


Figure S5. Expression profile of a human long noncoding RNA. Panel A show an overlapping long noncoding RNA identified in NONCODE version 3 (Bu et al., 2012), n324899, which overlaps with human USP25 in chromosome 21 and in the opposite orientation. The region of overlap is between 17,102,496 to 17,109,879 base-pairs, which contain 3 Affymetrix core probesets (probeset IDs 3915089, 3915090 and 3915091) and 1 extended probeset, probeset ID 3915088. Panel B shows the expression profile of the antisense transcript which corresponds to NONCODE transcript ID n324899. **Main Findings:** Our antisense microarray results confirm the presence of expected NONCODE transcript.

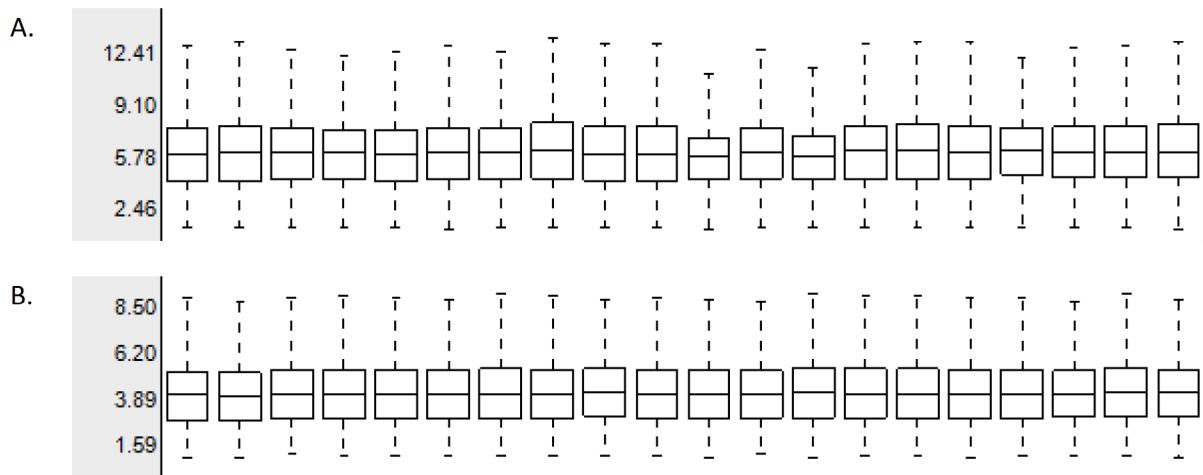


Figure S6. Distribution of core probeset intensity after RMA normalization: human at gene probeset level. Panel A shows intensity of sense gene probesets. Panel B shows intensity of antisense gene probesets. The y-axis represents the probeset intensities. The samples are (from left to right): brain replicate 1, brain replicate 2, colon replicate 1, colon replicate 2, heart replicate 1, heart replicate 2, kidney replicate 1, kidney replicate 2, liver replicate 1, liver replicate 2, lung replicate 1, lung replicate 2, ovary replicate 1, ovary replicate 2, spleen replicate 1, spleen replicate 2, testes replicate 1, testes replicate 2, thymus replicate 1, and thymus replicate 2.

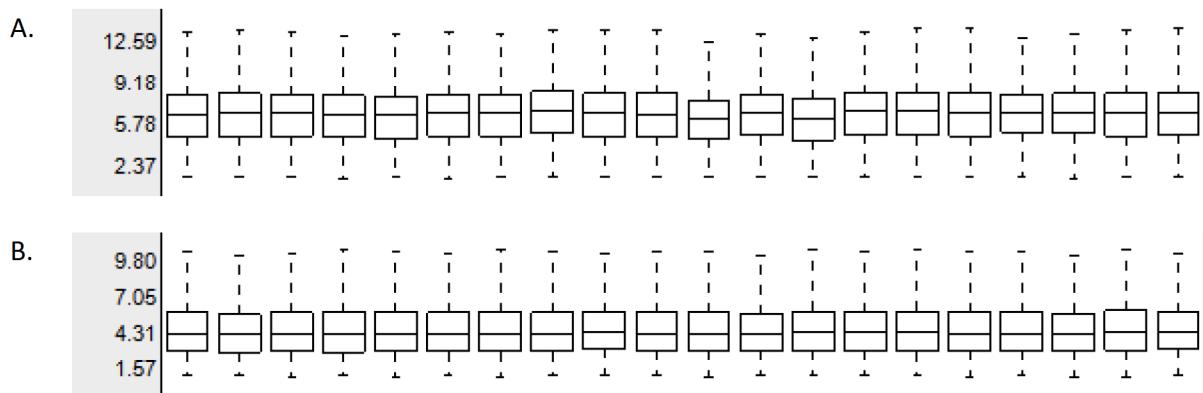


Figure S7. Distribution of core probeset intensity after RMA normalization: human at exon probeset level. Panel A shows intensity of sense exon probesets. Panel B shows intensity of antisense exon probesets. The y-axis represents the probeset intensities. The samples are (from left to right): brain replicate 1, brain replicate 2, colon replicate 1, colon replicate 2, heart replicate 1, heart replicate 2, kidney replicate 1, kidney replicate 2, liver replicate 1, liver replicate 2, lung replicate 1, lung replicate 2, ovary replicate 1, ovary replicate 2, spleen replicate 1, spleen replicate 2, testes replicate 1, testes replicate 2, thymus replicate 1, and thymus replicate 2.

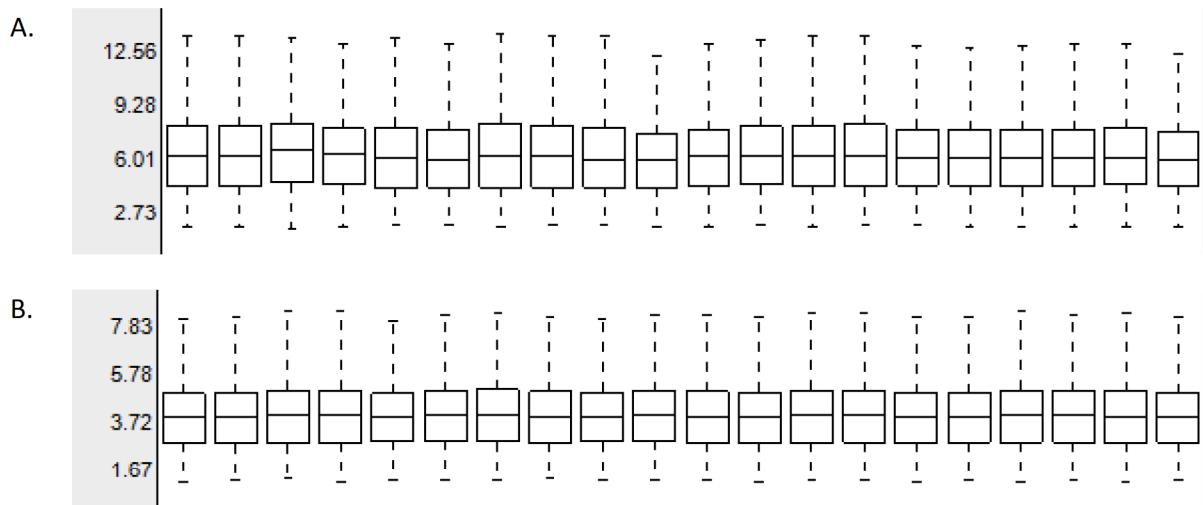


Figure S8. Distribution of core probeset intensity after RMA normalization: mouse at gene probeset level. Panel A shows intensity of sense gene probesets. Panel B shows intensity of antisense gene probesets. The y-axis represents the probeset intensities. The samples are (from left to right): brain replicate 1, brain replicate 2, embryo replicate 1, embryo replicate 2, heart replicate 1, heart replicate 2, kidney replicate 1, kidney replicate 2, liver replicate 1, liver replicate 2, lung replicate 1, lung replicate 2, ovary replicate 1, ovary replicate 2, spleen replicate 1, spleen replicate 2, testes replicate 1, testes replicate 2, thymus replicate 1, and thymus replicate 2.

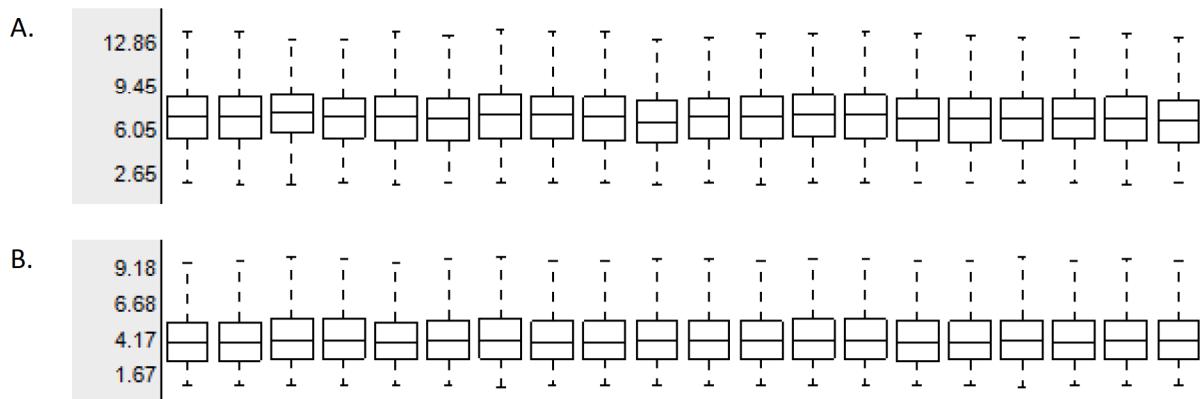


Figure S9. Distribution of core probeset intensity after RMA normalization: mouse at exon probeset level. Panel A shows intensity of sense exon probesets. Panel B shows intensity of antisense exon probesets. The y-axis represents the probeset intensities. The samples are (from left to right): brain replicate 1, brain replicate 2, embryo replicate 1, embryo replicate 2, heart replicate 1, heart replicate 2, kidney replicate 1, kidney replicate 2, liver replicate 1, liver replicate 2, lung replicate 1, lung replicate 2, ovary replicate 1, ovary replicate 2, spleen replicate 1, spleen replicate 2, testes replicate 1, testes replicate 2, thymus replicate 1, and thymus replicate 2.

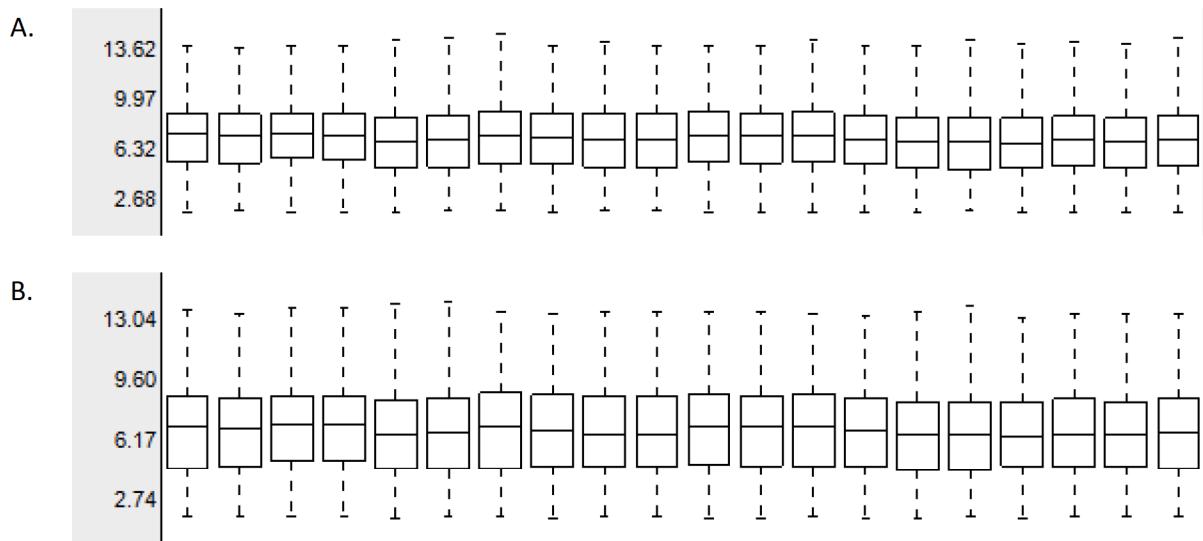


Figure S10. Distribution of core probeset intensity after RMA normalization: rat at gene probeset level.
 Panel A shows intensity of sense gene probesets. Panel B shows intensity of antisense gene probesets. The y-axis represents the probeset intensities. The samples are (from left to right): brain replicate 1, brain replicate 2, embryo replicate 1, embryo replicate 2, heart replicate 1, heart replicate 2, kidney replicate 1, kidney replicate 2, liver replicate 1, liver replicate 2, lung replicate 1, lung replicate 2, ovary replicate 1, ovary replicate 2, spleen replicate 1, spleen replicate 2, testes replicate 1, testes replicate 2, thymus replicate 1, and thymus replicate 2.

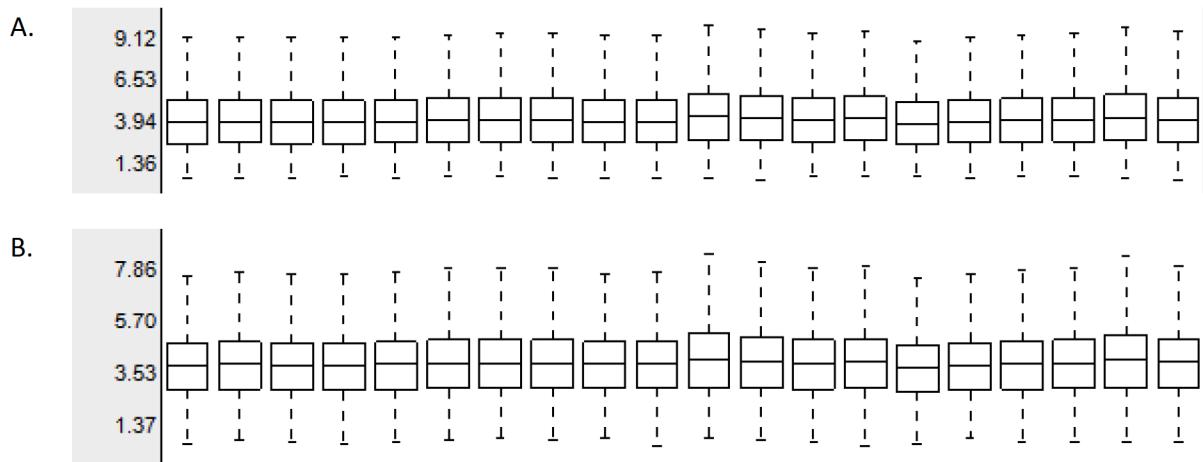


Figure S11. Distribution of core probeset intensity after RMA normalization: rat at exon probeset level.
 Panel A shows intensity of sense exon probesets. Panel B shows intensity of antisense exon probesets. The y-axis represents the probeset intensities. The samples are (from left to right): brain replicate 1, brain replicate 2, embryo replicate 1, embryo replicate 2, heart replicate 1, heart replicate 2, kidney replicate 1, kidney replicate 2, liver replicate 1, liver replicate 2, lung replicate 1, lung replicate 2, ovary replicate 1, ovary replicate 2, spleen replicate 1, spleen replicate 2, testes replicate 1, testes replicate 2, thymus replicate 1, and thymus replicate 2.

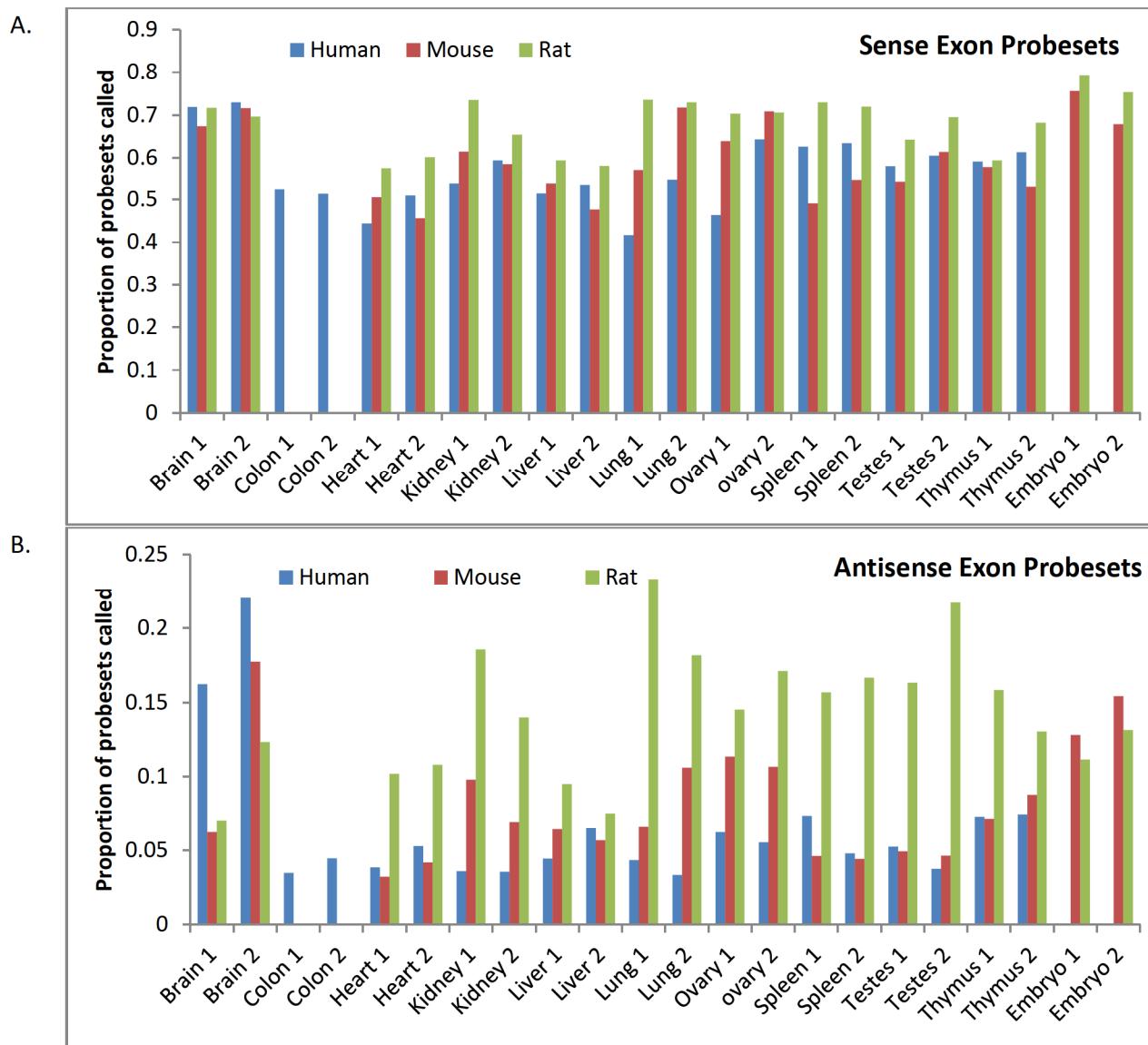


Figure S12. Proportion of exon-level probeset detected above background (DABG), with a detection p-value of < 0.01. Panel A shows sense arrays. Panel B shows antisense arrays. **Main Findings:** The proportion of detected probesets in the sense and antisense exon arrays were 61.5% (standard deviation of 9.2%) and 9.6% (standard deviation of 5.5%) respectively, which is relatively consistent with Ge et al. (2008) reporting 41.4% sense detection and 13.4% antisense detection.

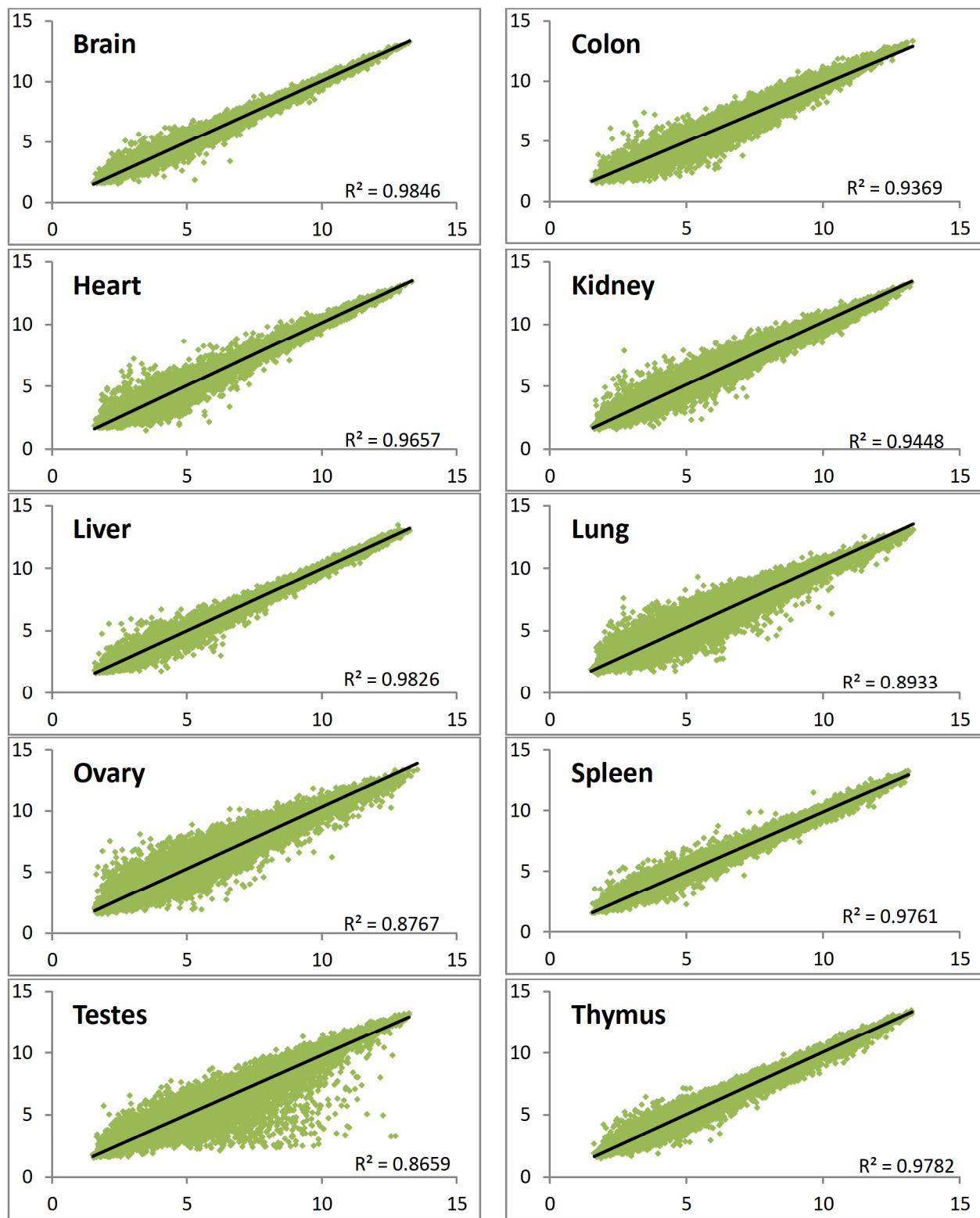


Figure S13. Correlation between intensity of technical replicates ($n = 2$) for human sense gene array. The y-axis and x-axis are the expression intensities (in logarithmic scale) of the first and second replicate respectively. **Main Findings:** All technical replicates are correlated.

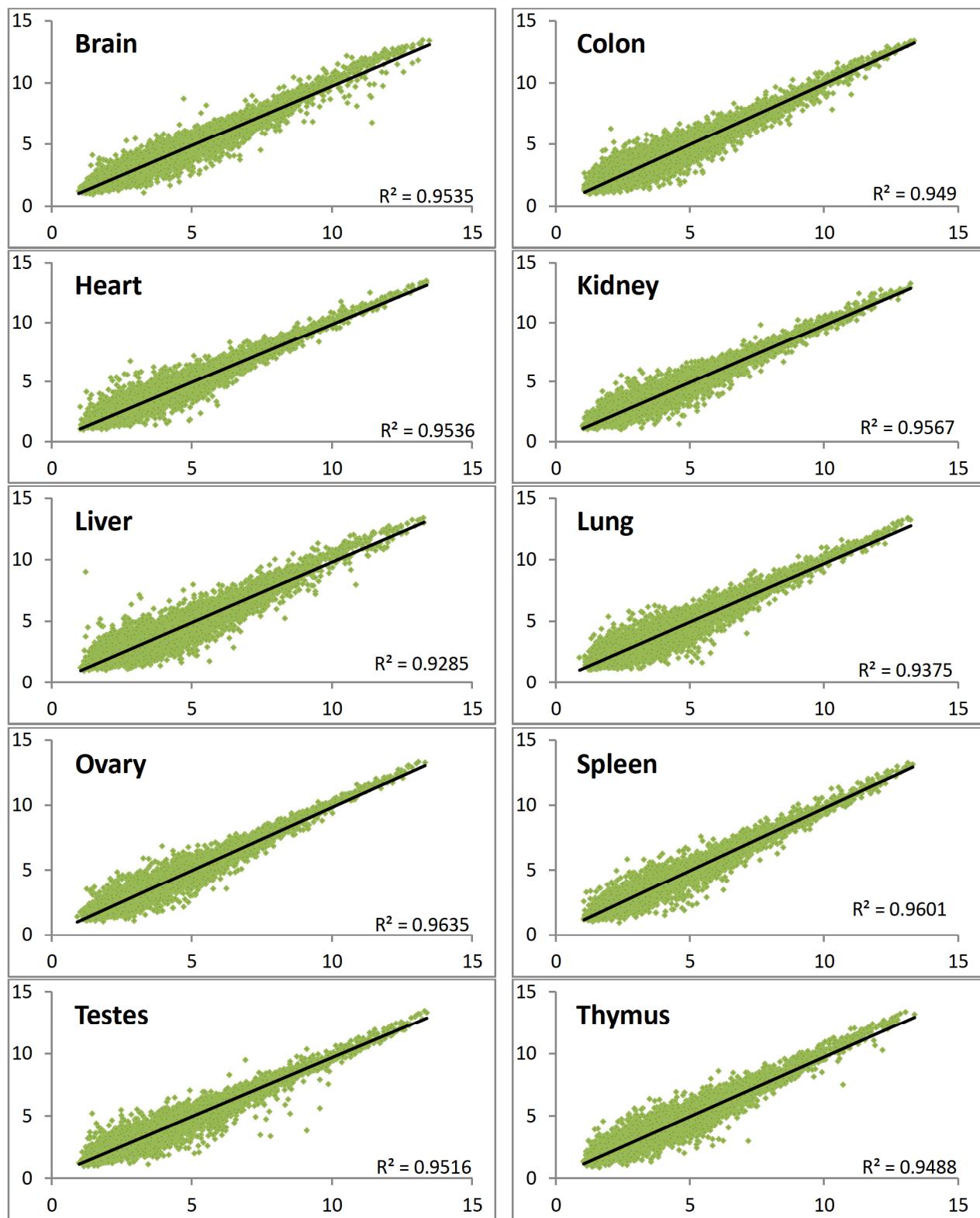


Figure S14. Correlation between intensity of technical replicates ($n = 2$) for human antisense gene array. The y-axis and x-axis are the expression intensities (in logarithmic scale) of the first and second replicate respectively. **Main Findings:** All technical replicates are correlated.

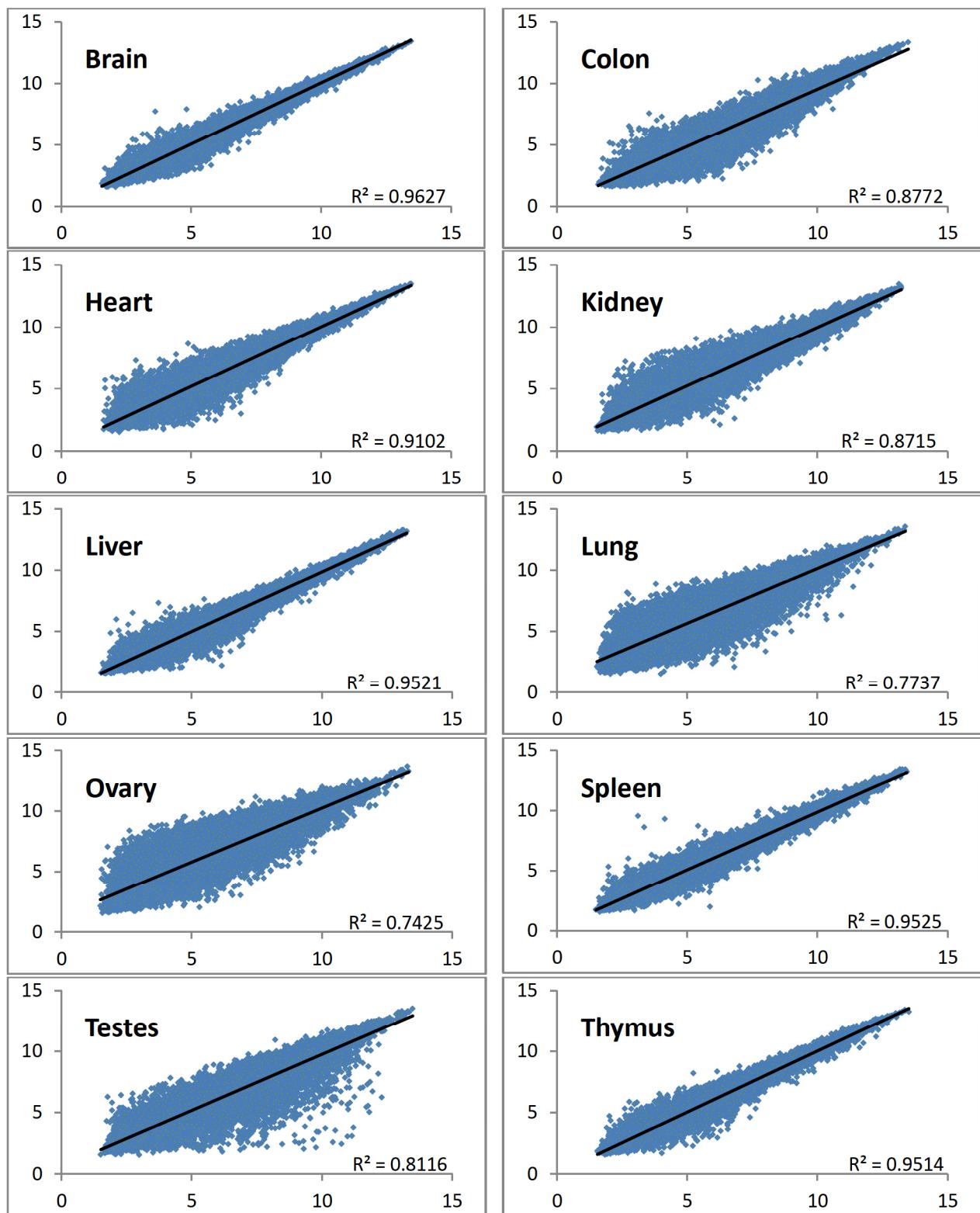


Figure S15. Correlation between intensity of technical replicates ($n = 2$) for human sense exon array. The y-axis and x-axis are the expression intensities (in logarithmic scale) of the first and second replicate respectively. **Main Findings:** All technical replicates are correlated.

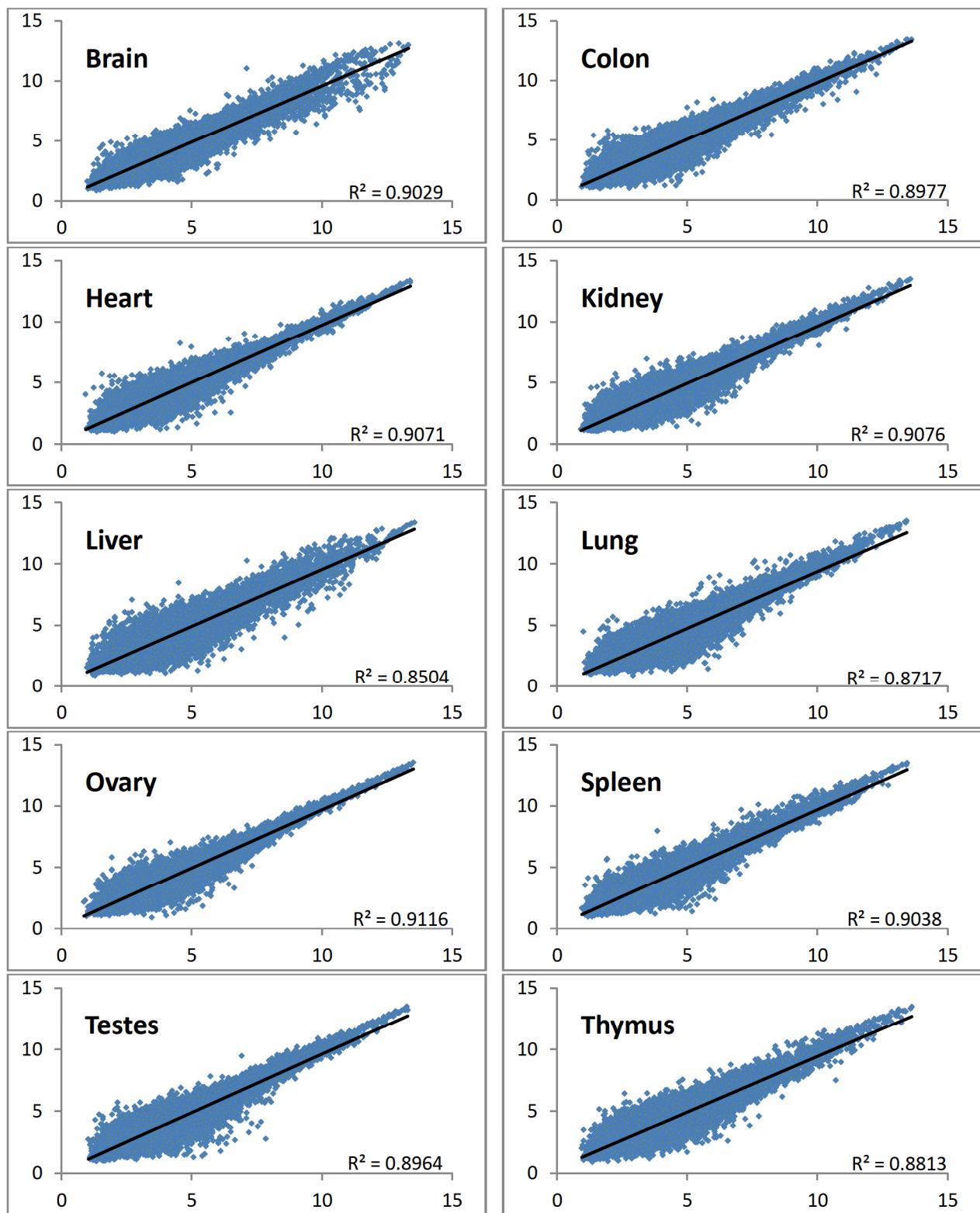


Figure S16. Correlation between intensity of technical replicates ($n = 2$) for human antisense exon array. The y-axis and x-axis are the expression intensities (in logarithmic scale) of the first and second replicate respectively. **Main Findings:** All technical replicates are correlated.

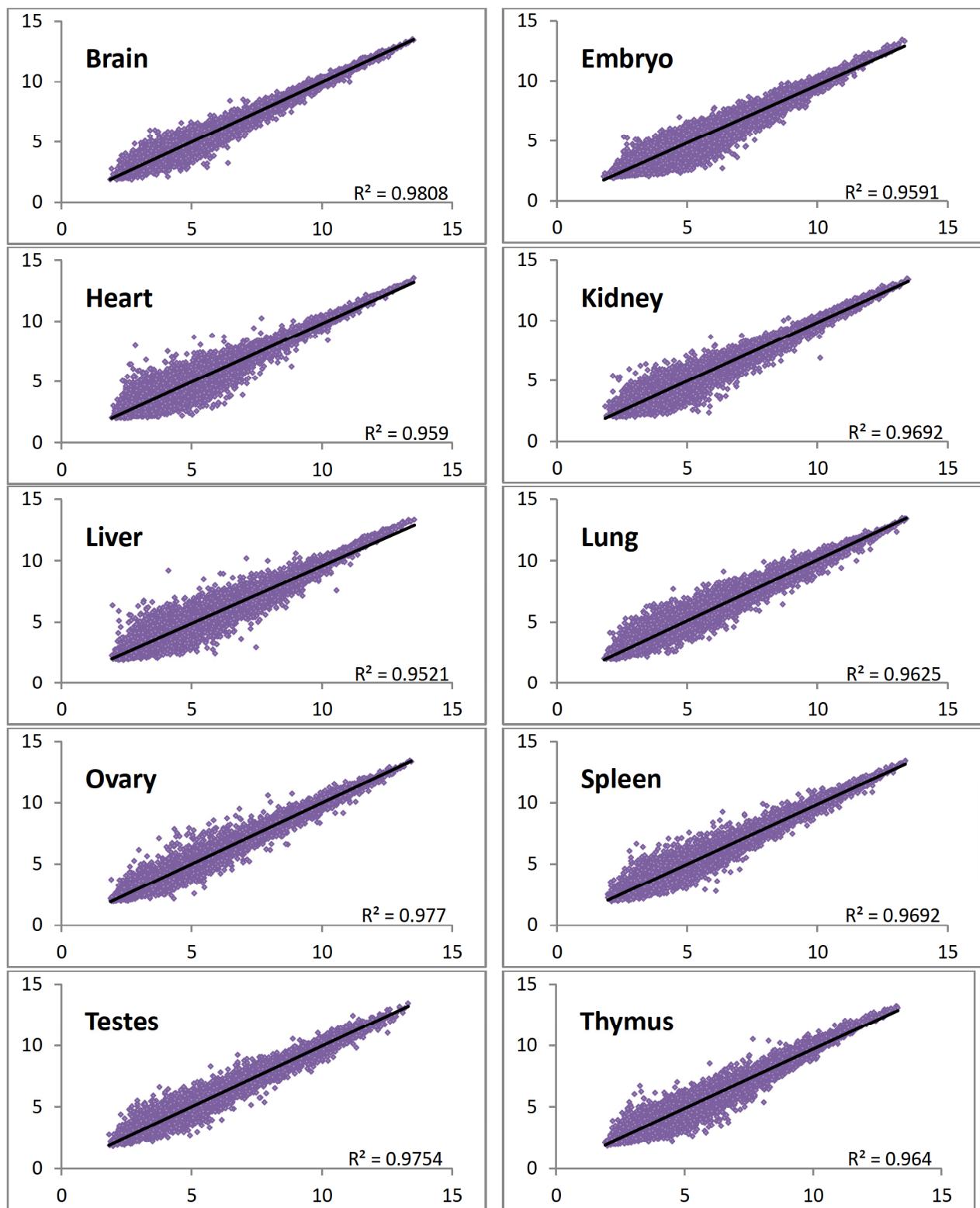


Figure S17. Correlation between intensity of technical replicates ($n = 2$) for mouse sense gene array. The y-axis and x-axis are the expression intensities (in logarithmic scale) of the first and second replicate respectively. **Main Findings:** All technical replicates are correlated.

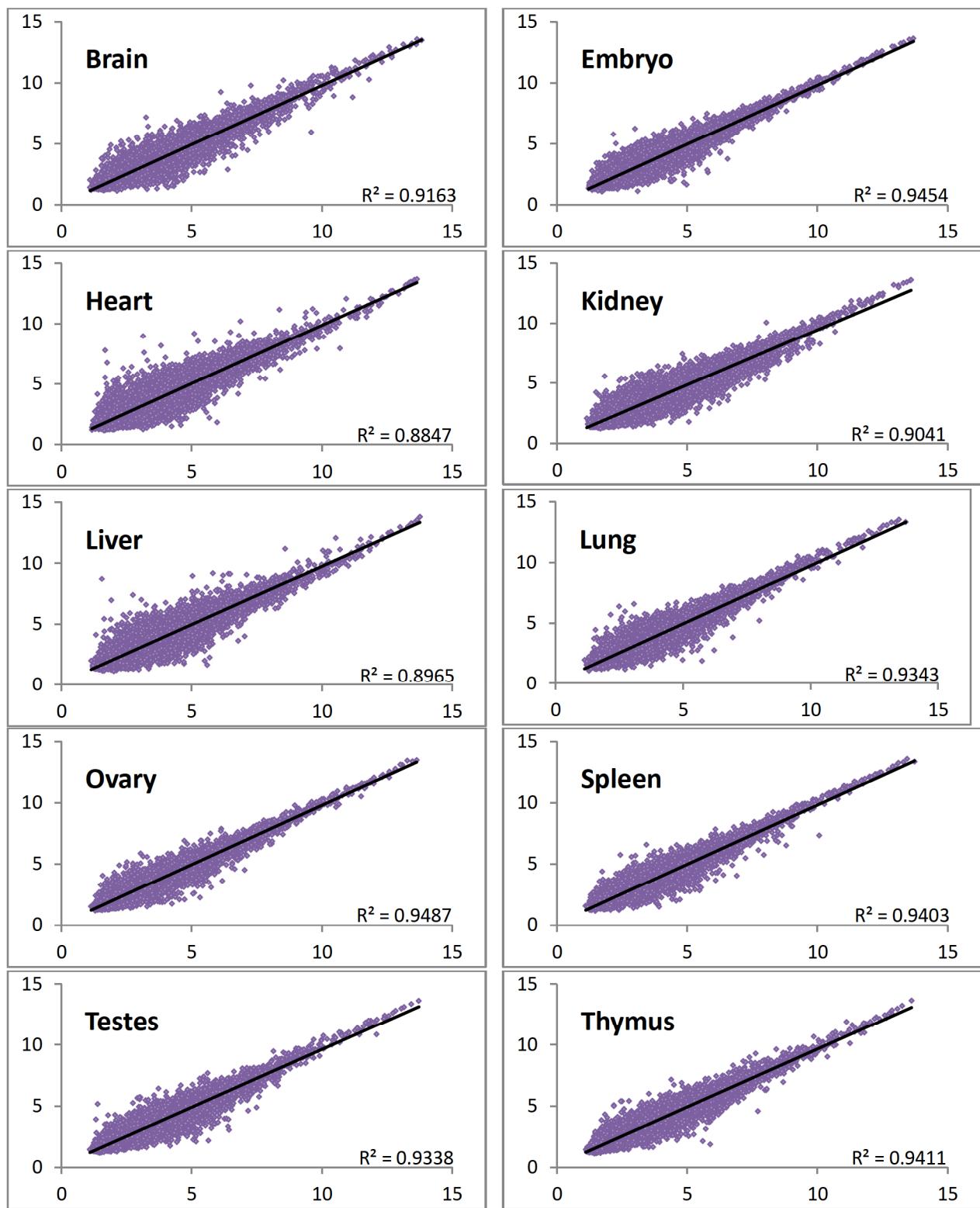


Figure S18. Correlation between intensity of technical replicates ($n = 2$) for mouse antisense gene array. The y-axis and x-axis are the expression intensities (in logarithmic scale) of the first and second replicate respectively. **Main Findings:** All technical replicates are correlated.

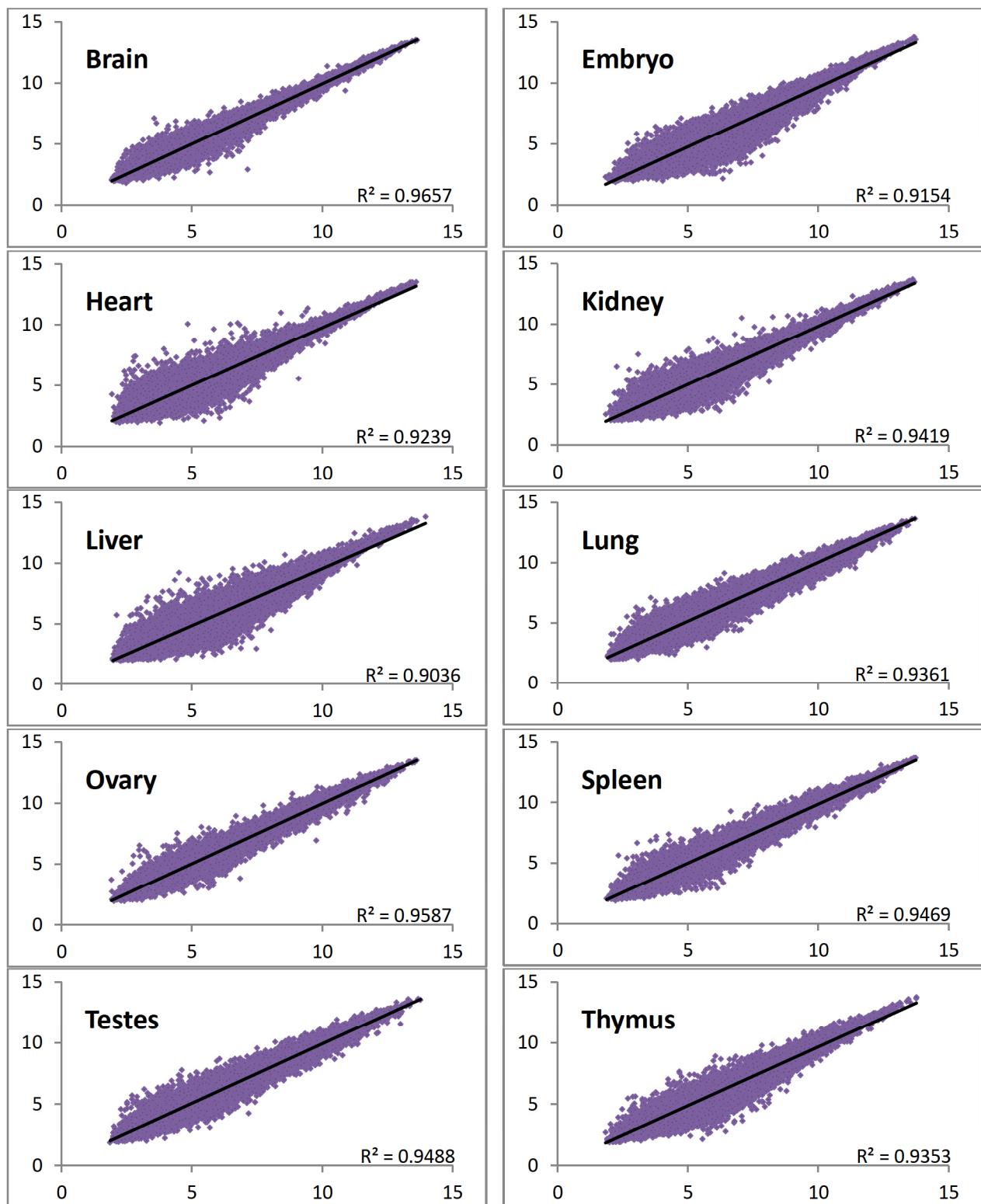


Figure S19. Correlation between intensity of technical replicates ($n = 2$) for mouse sense exon array. The y-axis and x-axis are the expression intensities (in logarithmic scale) of the first and second replicate respectively. **Main Findings:** All technical replicates are correlated.

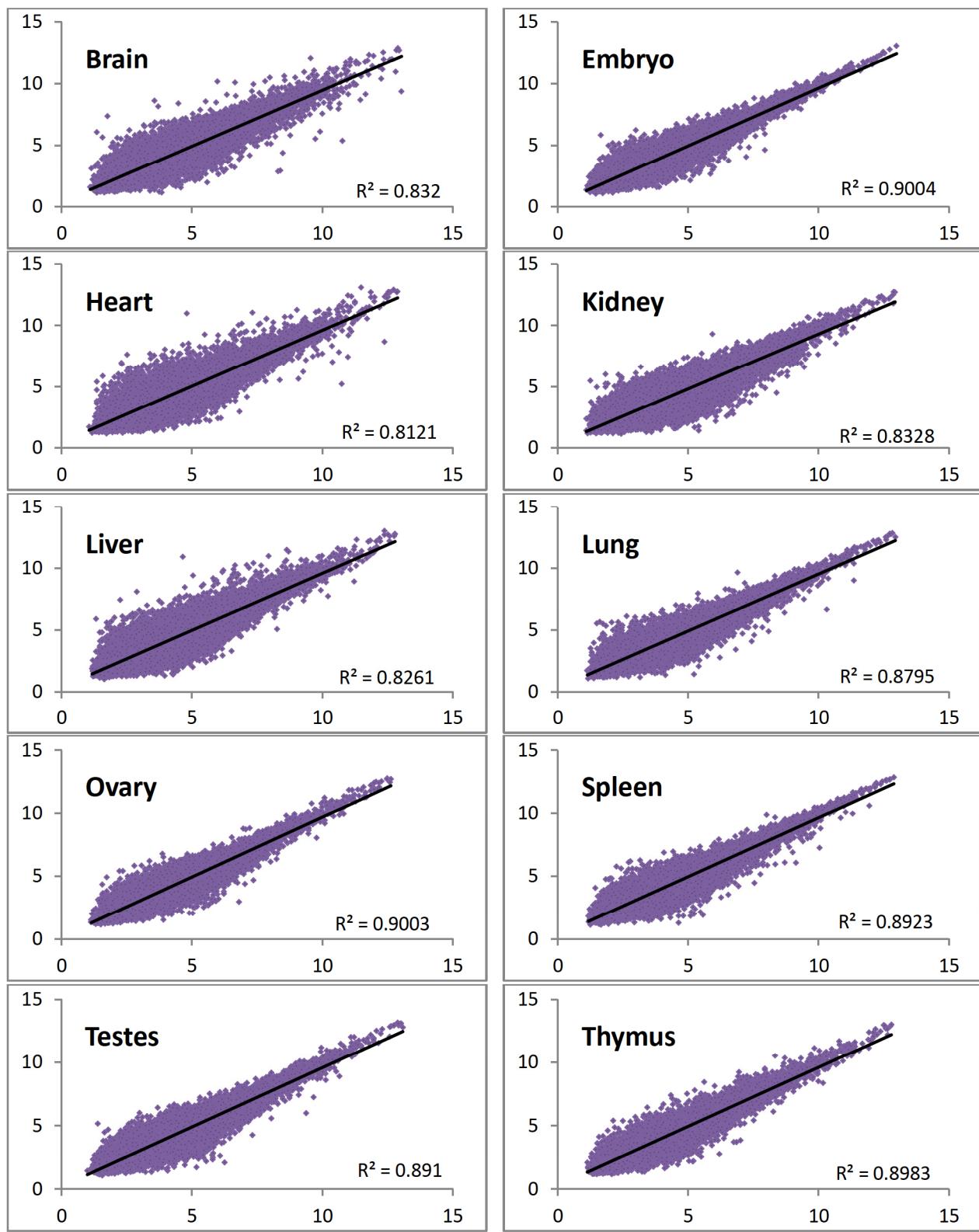


Figure S20. Correlation between intensity of technical replicates ($n = 2$) for mouse antisense exon array.
The y-axis and x-axis are the expression intensities (in logarithmic scale) of the first and second replicate respectively. **Main Findings: All technical replicates are correlated.**

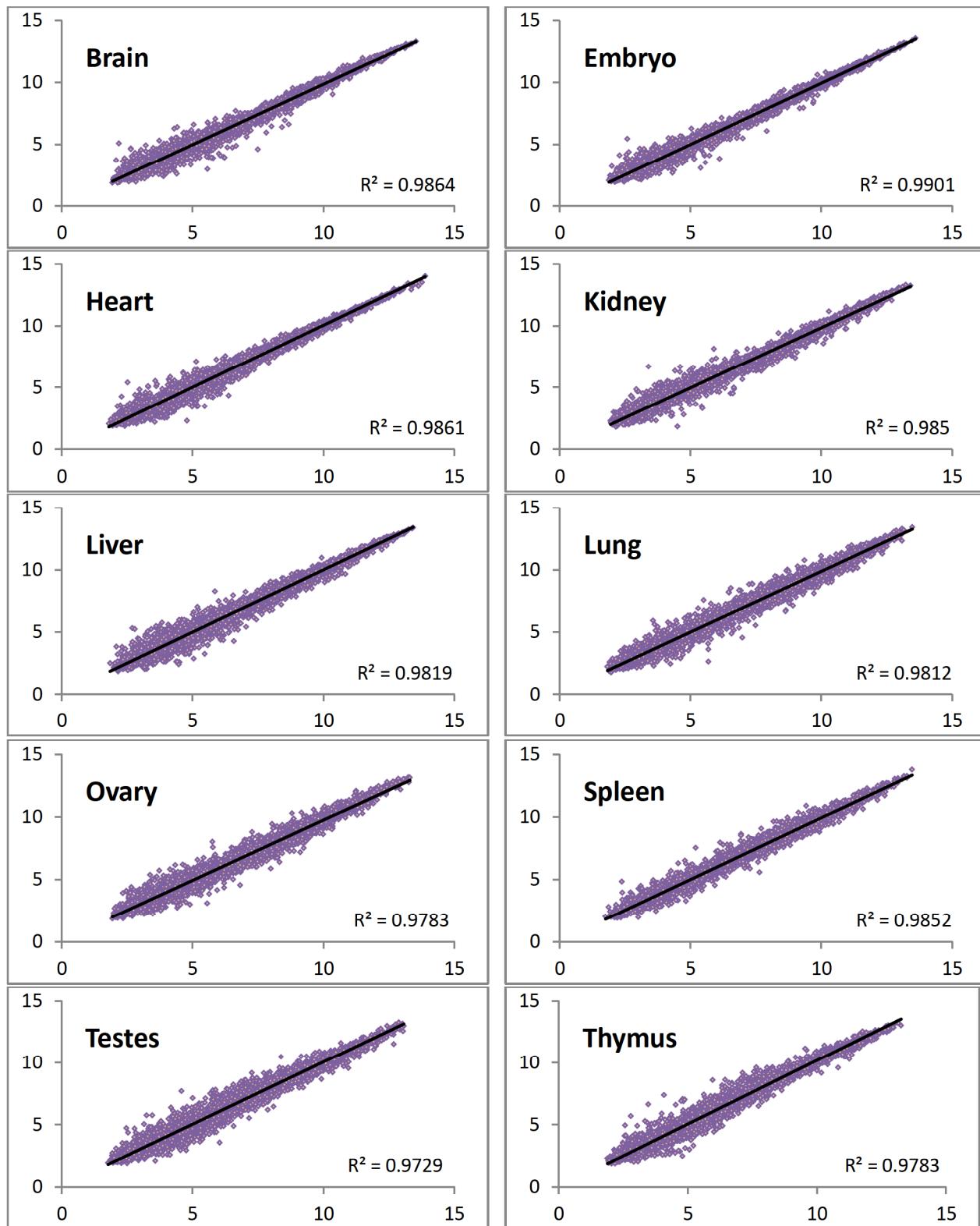


Figure S21. Correlation between intensity of technical replicates ($n = 2$) for rat sense gene array. The y-axis and x-axis are the expression intensities (in logarithmic scale) of the first and second replicate respectively. **Main Findings:** All technical replicates are correlated.

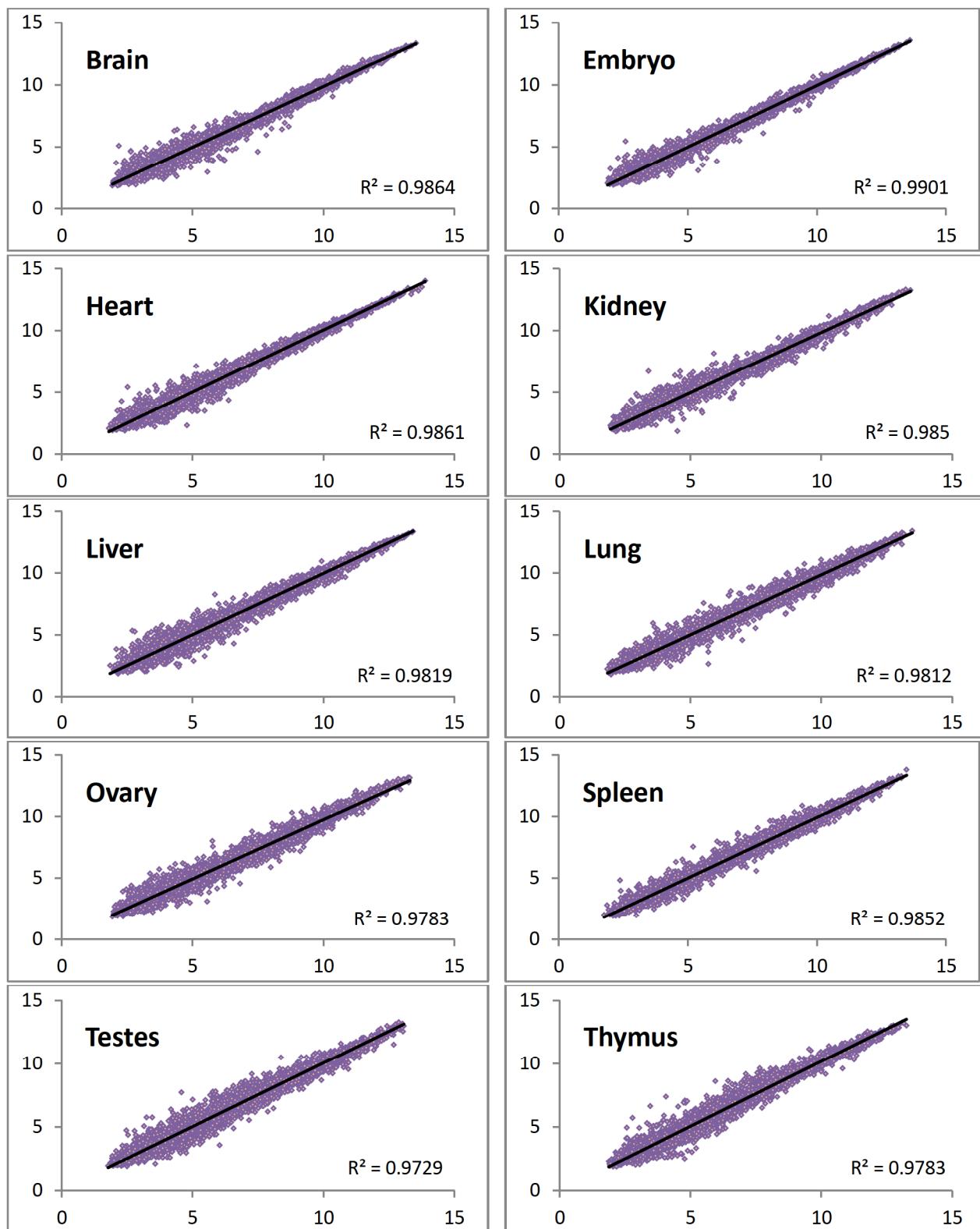


Figure S22. Correlation between intensity of technical replicates ($n = 2$) for rat antisense gene array. The y-axis and x-axis are the expression intensities (in logarithmic scale) of the first and second replicate respectively. **Main Findings:** All technical replicates are correlated.

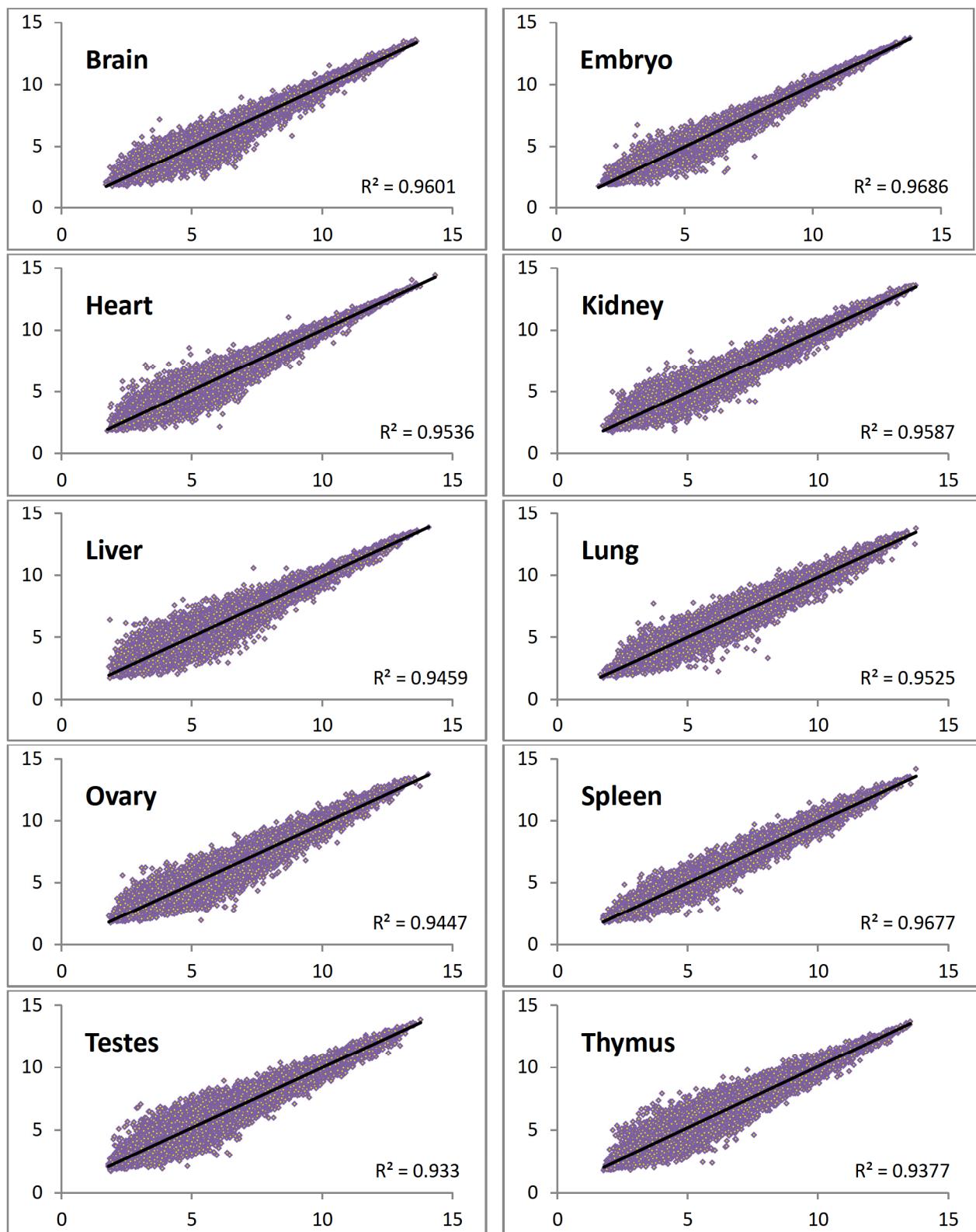


Figure S23. Correlation between intensity of technical replicates ($n = 2$) for rat sense exon array. The y-axis and x-axis are the expression intensities (in logarithmic scale) of the first and second replicate respectively. **Main Findings:** All technical replicates are correlated.

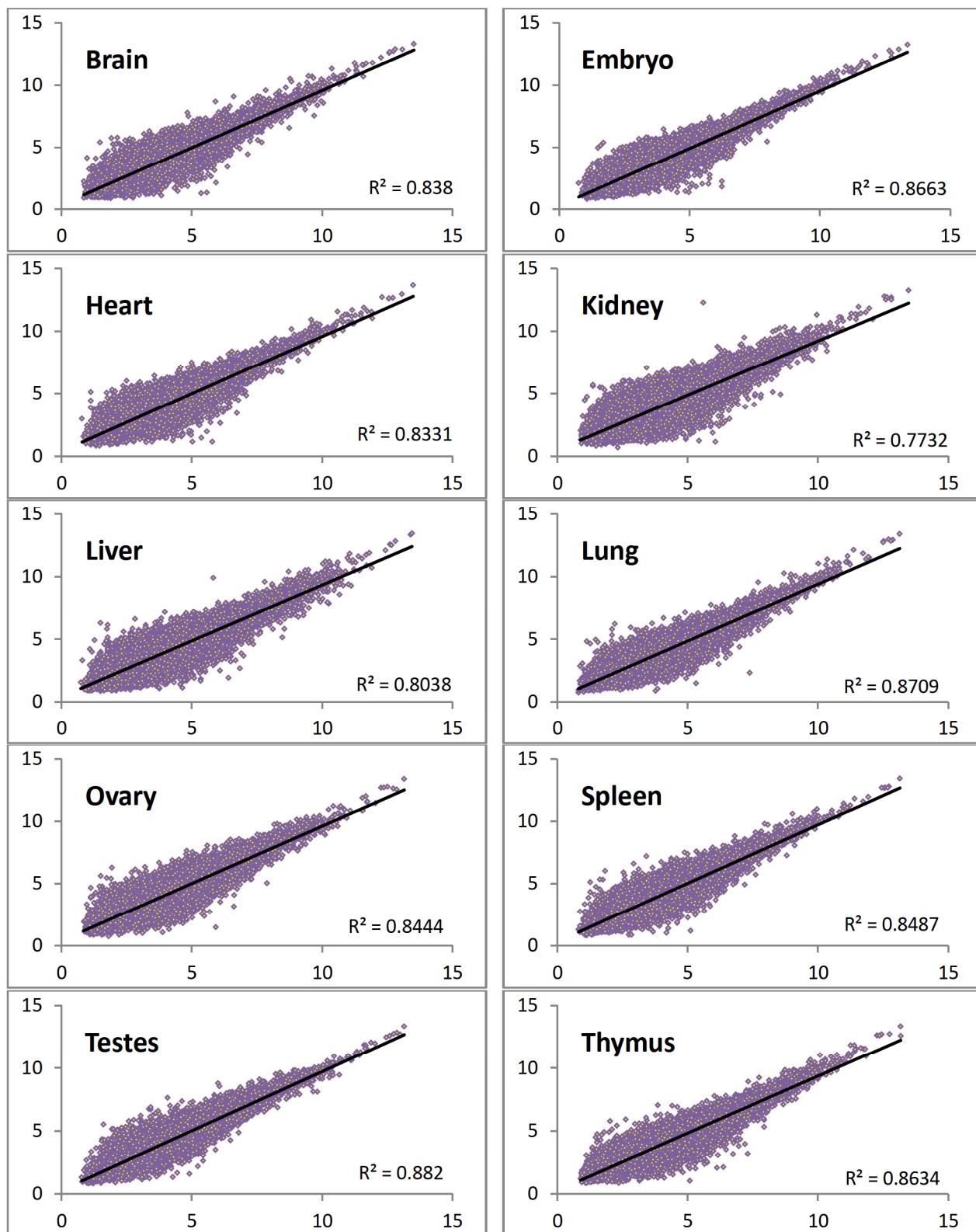


Figure S24. Correlation between intensity of technical replicates ($n = 2$) for rat antisense exon array. The y-axis and x-axis are the expression intensities (in logarithmic scale) of the first and second replicate respectively. **Main Findings:** All technical replicates are correlated.

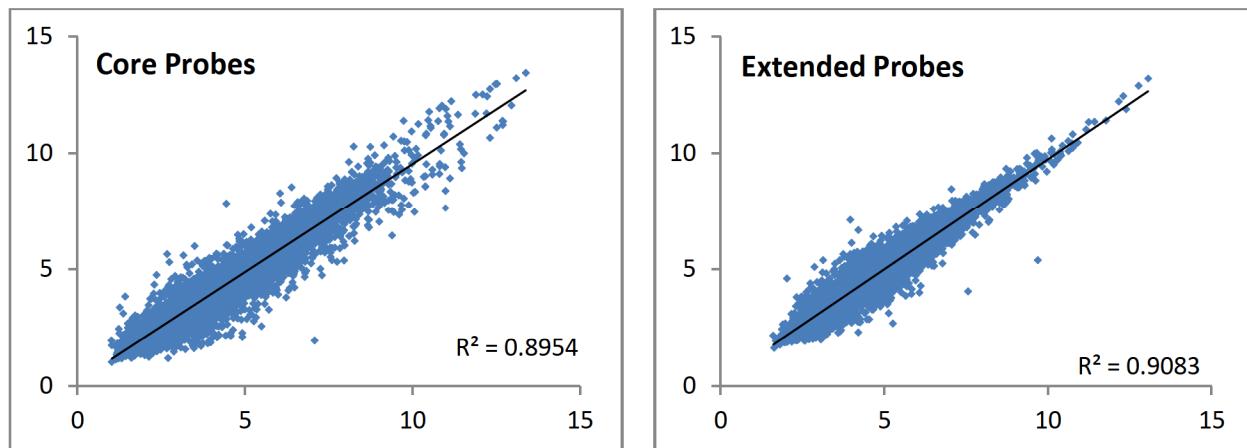


Figure S25. Comparing the correlations of the 2 technical replicates between overlapping core and extended probesets for human brain. 5283 extended probesets were found to be in opposite orientation compared to the core probesets with less than 25 bases between the centre of the opposite oriented probesets. This allowed for the validation of the antisense array protocol using the extended probesets in the corresponding sense array. Using these 5283 probesets, correlation between the technical replicates ($n = 2$) were estimated.

Main Findings: The technical replicates are strongly correlated in both core probesets in antisense array and extended probesets in sense array.

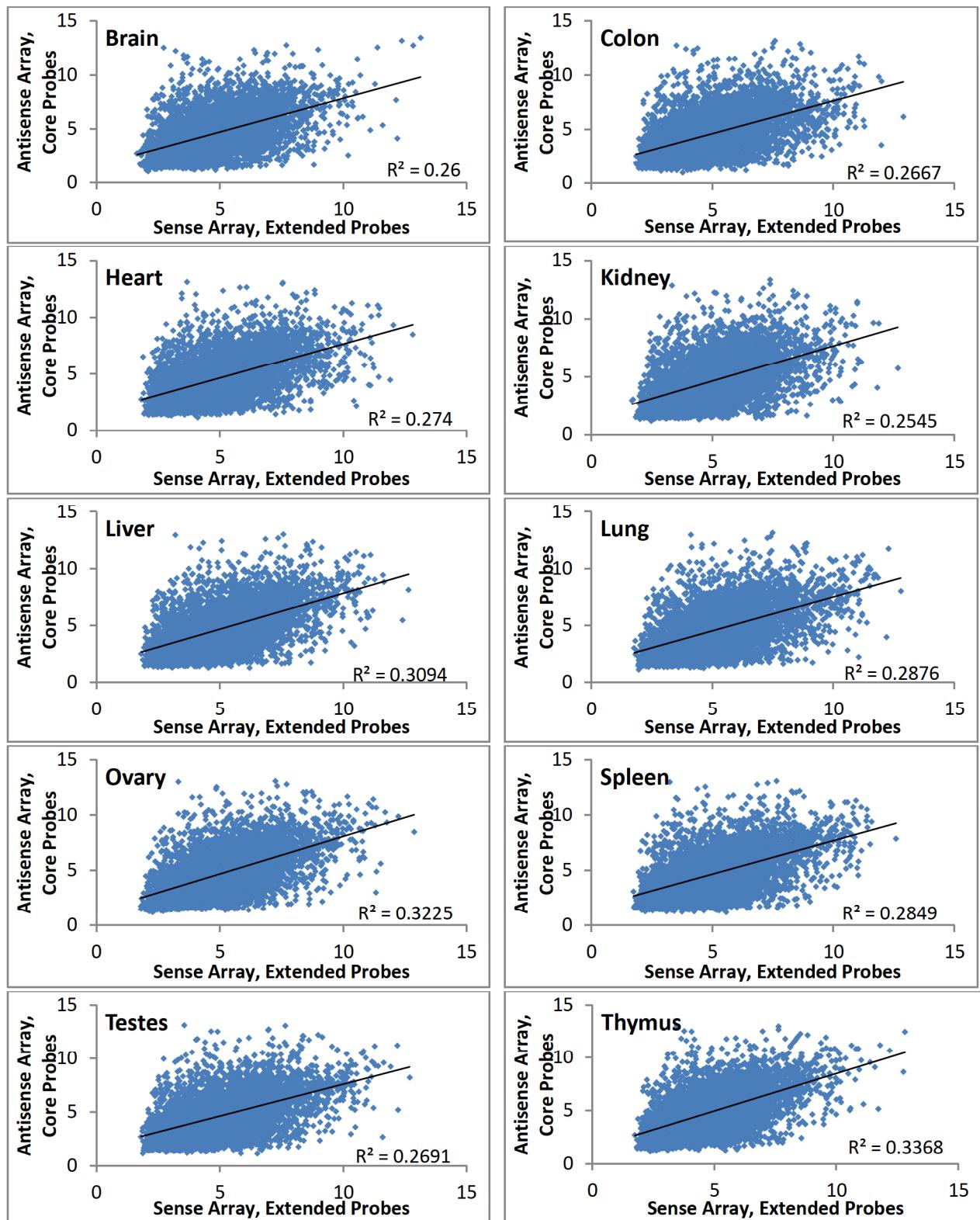


Figure S26. Correlation between overlapping core probesets and oppositely oriented extended probesets in human arrays. 5283 extended probesets were found to be in opposite orientation compared to the core probesets with less than 25 bases between the centre of the opposite oriented probesets. This allowed for the validation of the antisense array protocol using the extended probesets

in the corresponding sense array. The expression intensity of each probeset is the average of the 2 technical replicates. **Main Findings: Using 10 human tissues, the correlation between the overlapping core probesets and the oppositely oriented extended probesets is between 0.25 to 0.34.**

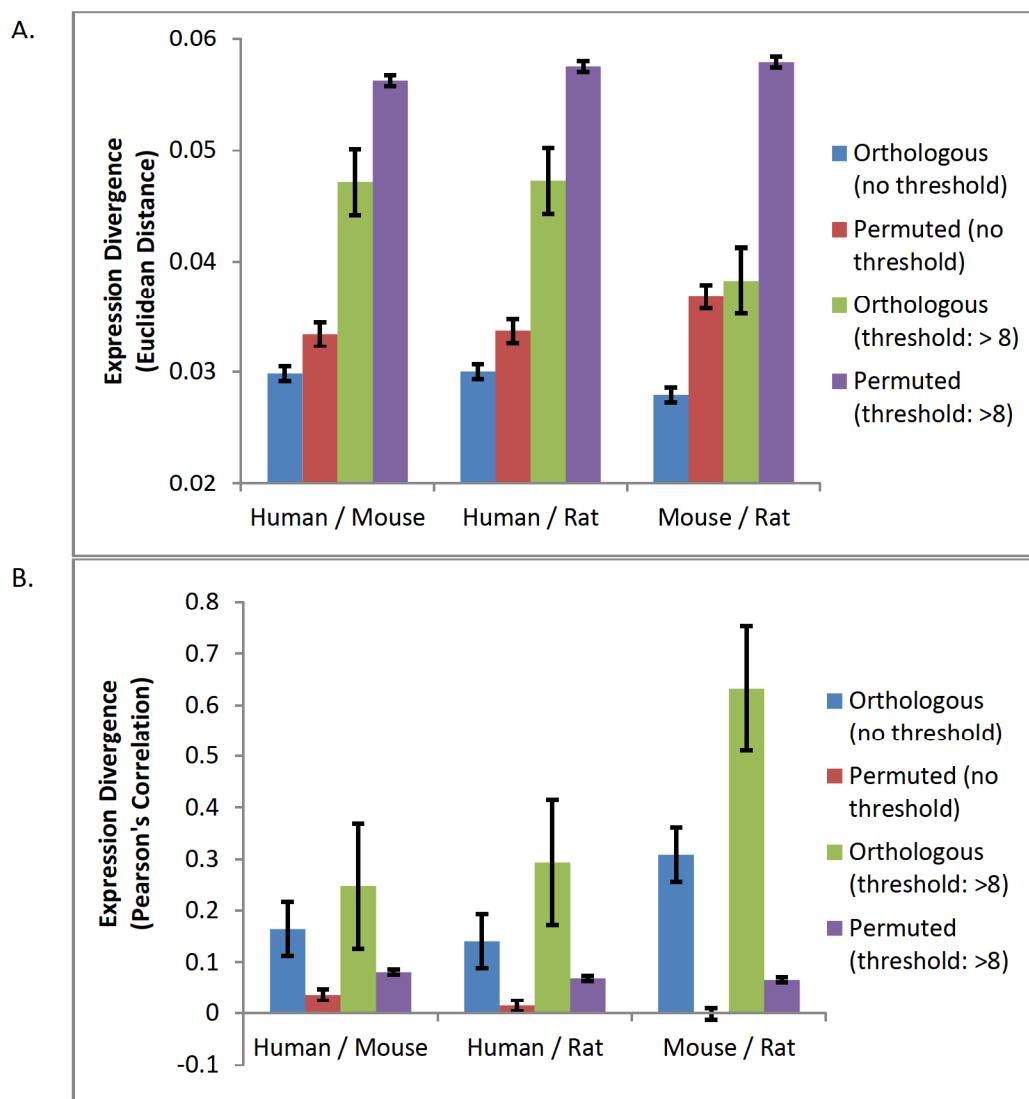


Figure S27. Average expression divergence between orthologous sense genes and permuted pairs. Panel A used Euclidean distance as a measure of expression divergence while Panel B used Pearson's correlation. "No threshold" refers to using full set of orthologous genes. "Threshold: >8" requires at least one of the orthologous probesets to have an expression intensity of more than 8, so as to reduce errors due to lowly expressed genes in both species. Error bars denote standard error.

Main Findings: Orthologous pairs show expression divergence which are significantly lower than that of permuted pairs regardless of threshold or measures, suggesting that different threshold or measures do not affect the conclusion. Mouse/rat orthologous genes show lower expression divergence compared to either human/mouse or human/rat orthologous genes. These findings are consistent with Jordan et al. (2005).

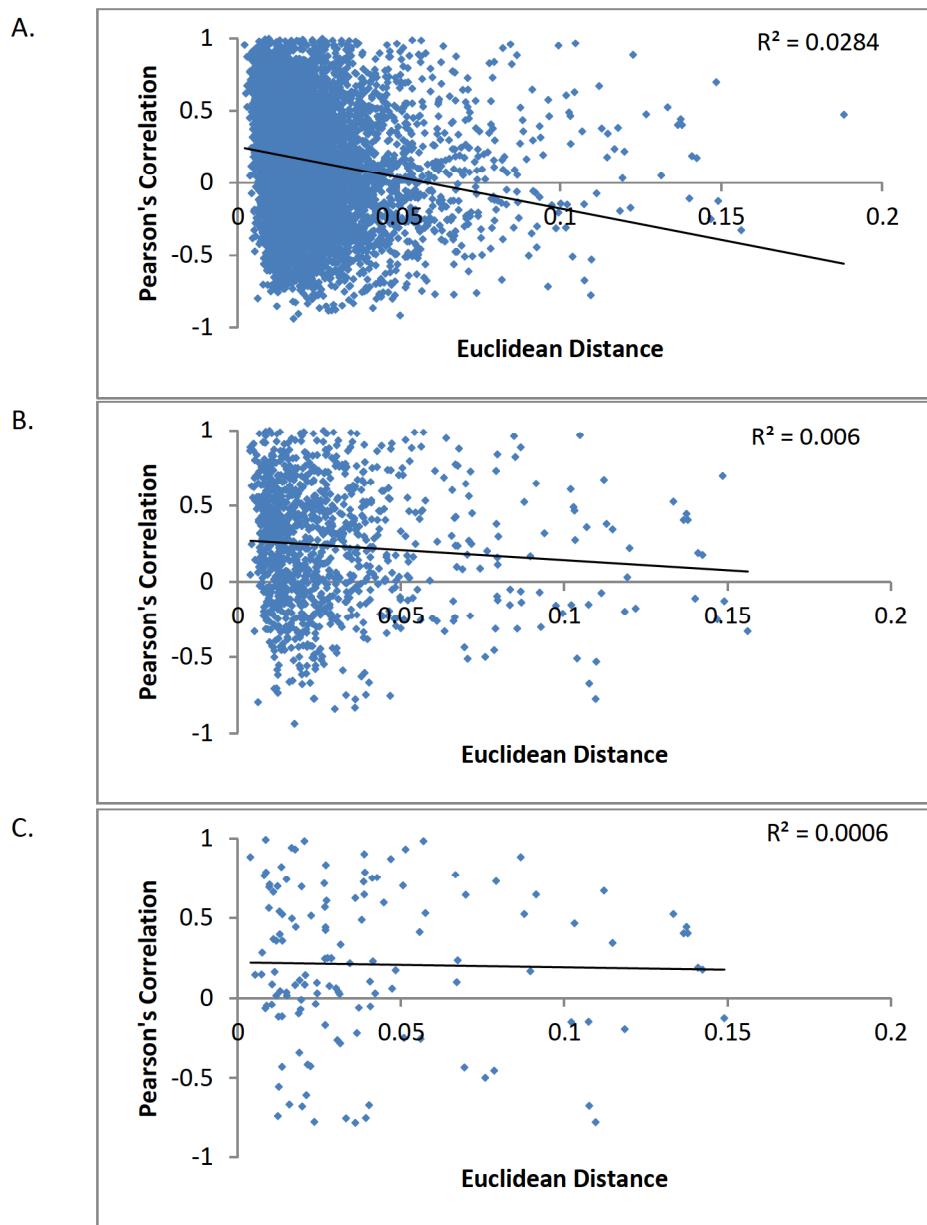


Figure S28. Correlation between Euclidean distance and Pearson's Correlation. The correlation between Euclidean distance and Pearson's Correlation was analyzed using orthologous sense gene expressions between human and mouse. Panel A shows the scatterplot using all orthologous gene expressions between human and mouse while Panels B and C requires at least one of the expressions in the orthologous tissues to be more than 6.5 and 8 respectively. **Main Findings: There is no correlation between Euclidean distance and Pearson's correlation.**

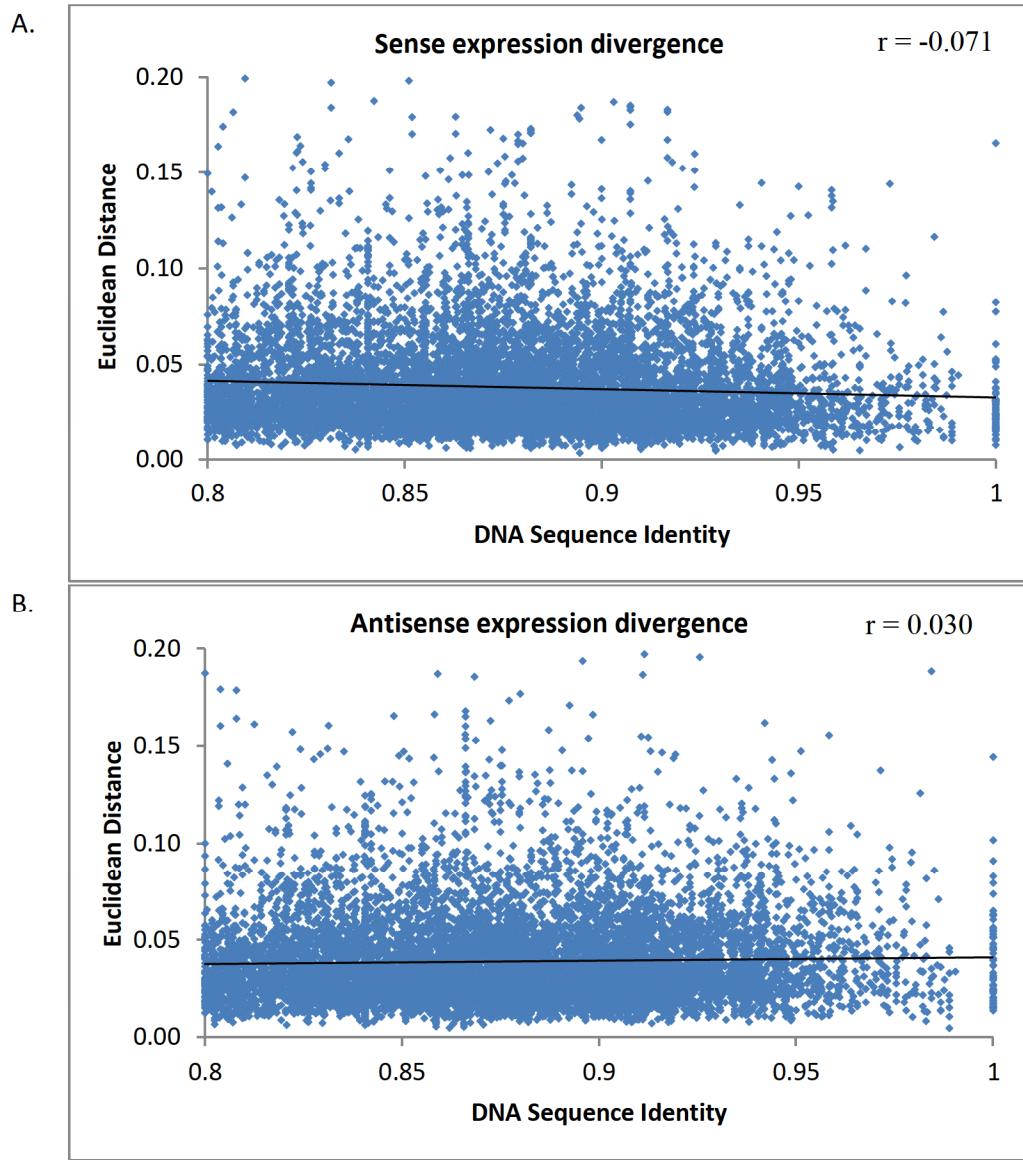


Figure S29. Correlation between expression divergence and DNA sequence identity.

Main Findings: Expression divergence is not correlated to DNA sequence identity for both sense and antisense transcripts.

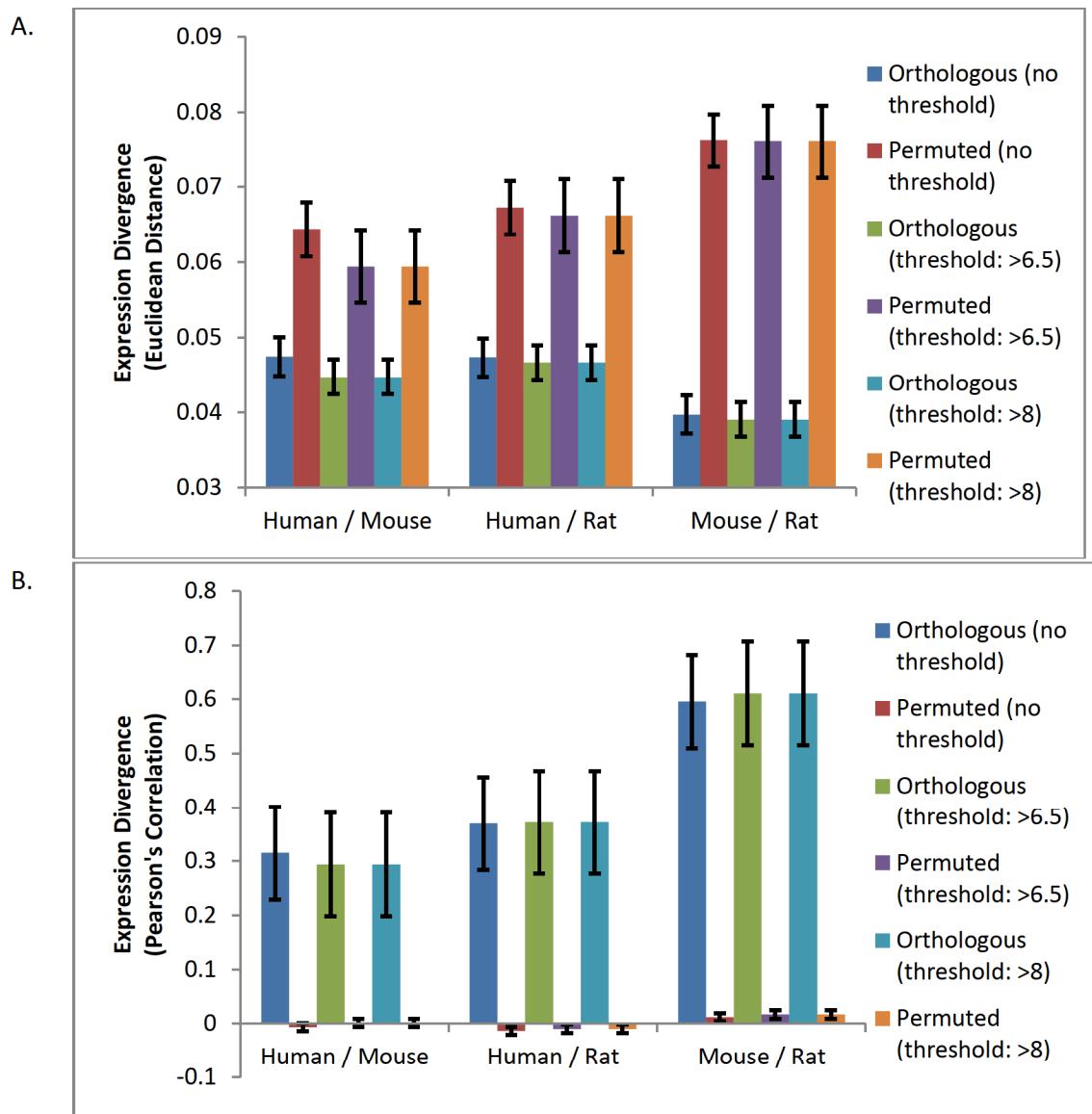


Figure S30. Average expression divergence between orthologous sense exon probesets and permuted pairs. Panel A used Euclidean distance as a measure of expression divergence while Panel B used Pearson's correlation. "No threshold" refers to using full set of orthologous genes. "Threshold: >6.5" and "Threshold: >8" requires at least one of the orthologous probesets to have an expression intensity of more than 6.5 and 8 respectively. Error bars denote standard error.

Main Findings: Orthologous pairs show expression divergence which are significantly lower than that of permuted pairs regardless of threshold or measures, suggesting that different threshold or measures do not affect the conclusion. Mouse/rat orthologous genes show lower expression divergence compared to either human/mouse or human/rat orthologous genes. These findings are consistent with Jordan et al. (2005).

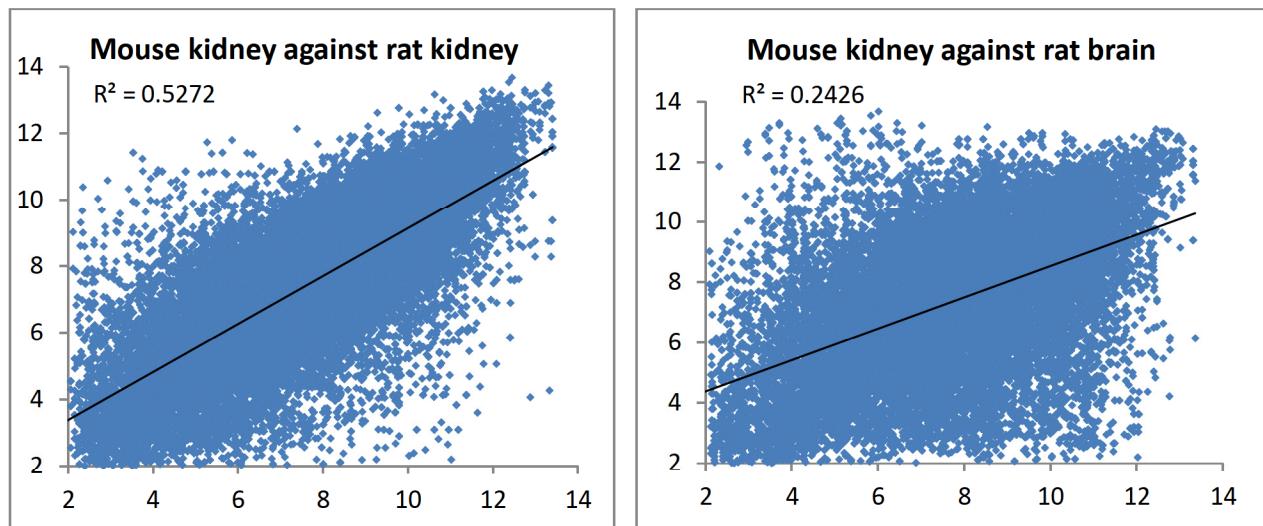


Figure S31. Scatterplot between sense exon expression of mouse kidney and rat kidney or rat brain.
Both axes are expression levels. **Main Findings:** Orthologous tissues are more correlated in expression levels than non-orthologous tissues for sense expression.

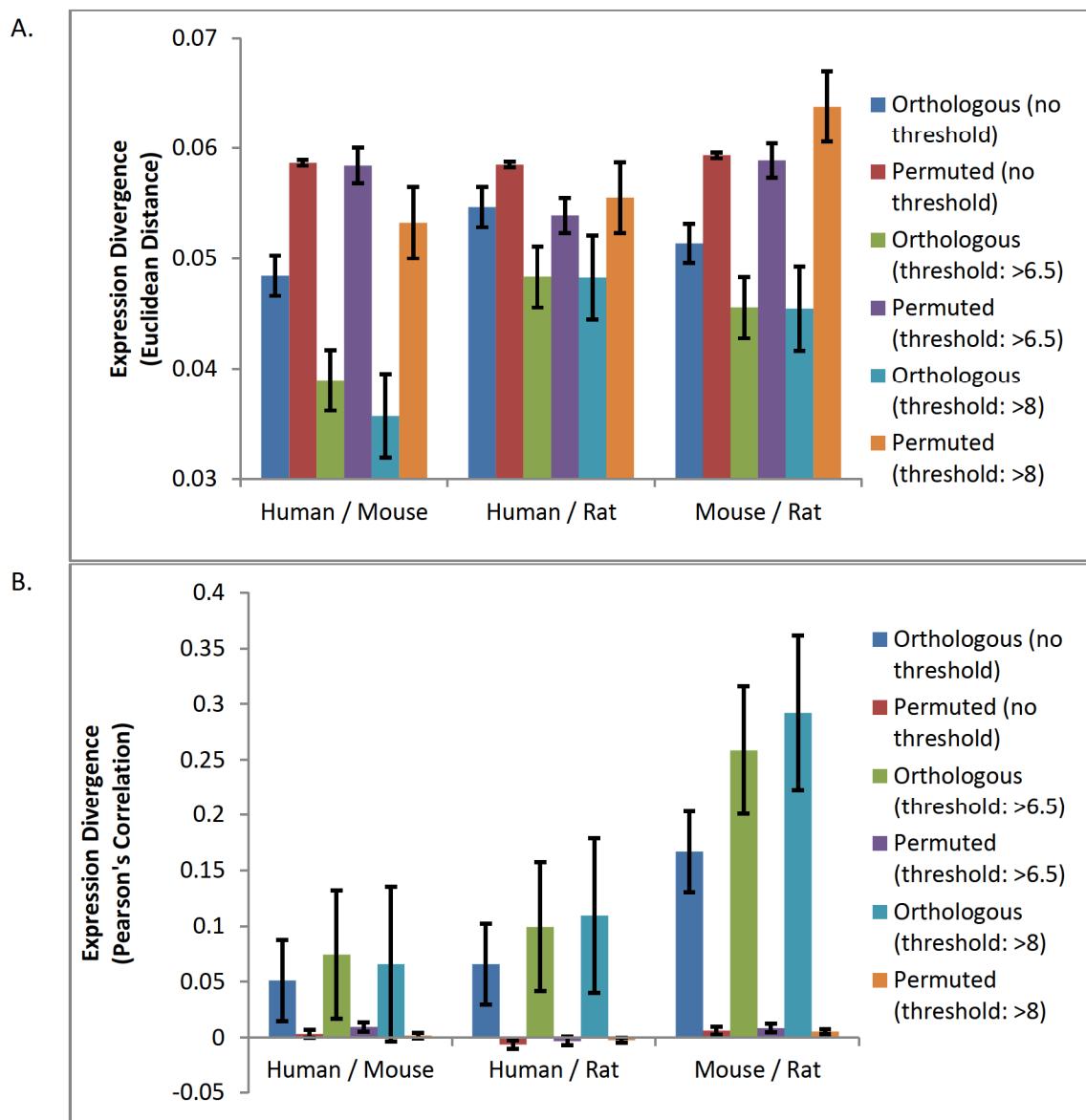


Figure S32. Average expression divergence between orthologous antisense transcript probesets at exon level and permuted pairs (with overlapping RefSeq transcripts as antisense transcripts). Panel A used Euclidean distance as a measure of expression divergence while Panel B used Pearson's correlation. "No threshold" refers to using full set of orthologous genes. "Threshold: >6.5" and "Threshold: >8" requires at least one of the orthologous probesets to have an expression intensity of more than 6.5 and 8 respectively. Error bars denote standard error.

Main Findings: Orthologous pairs show expression divergence which are significantly lower than that of permuted pairs regardless of threshold or measures, suggesting that different threshold or measures do not affect the conclusion.

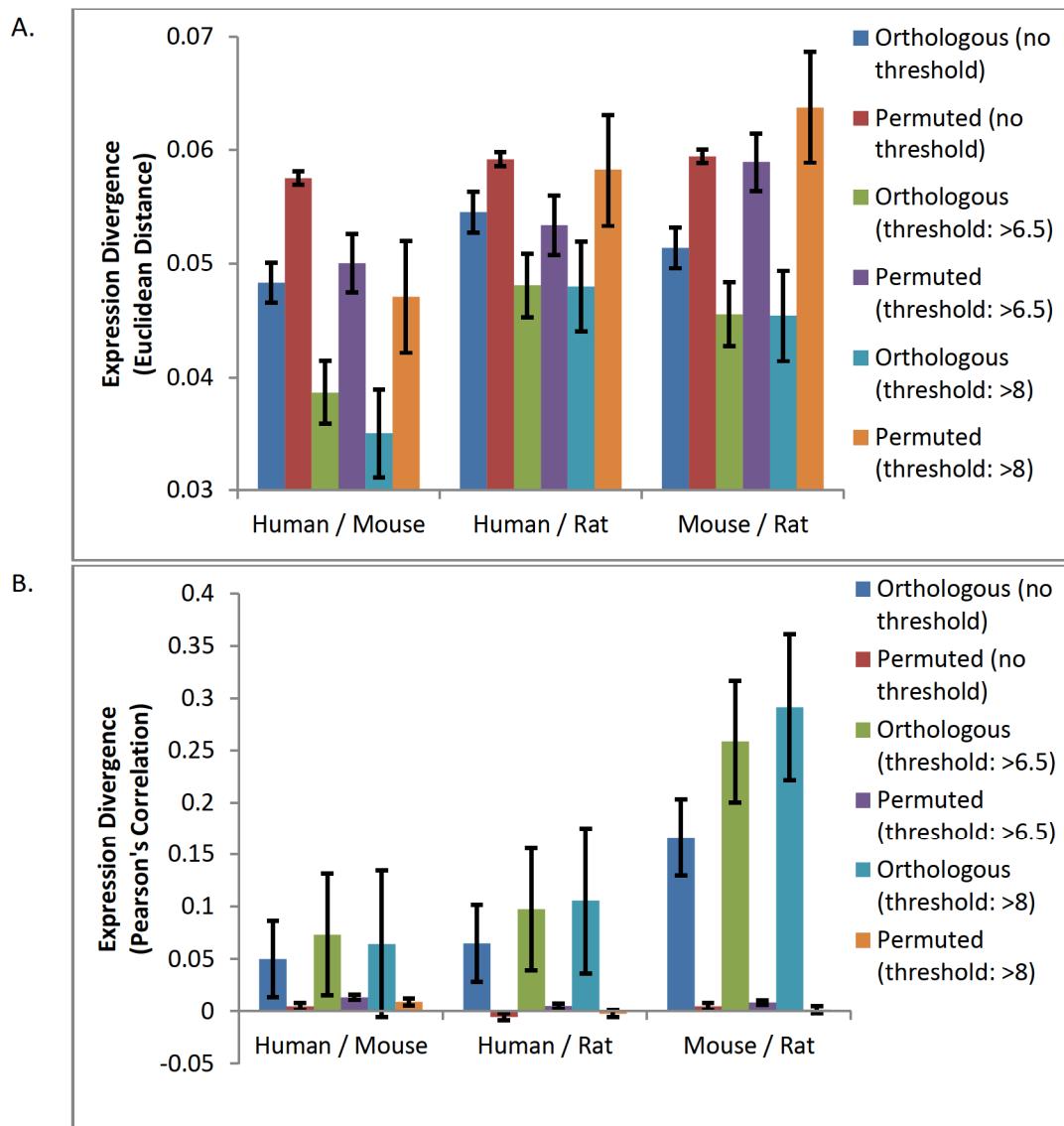


Figure S33. Average expression divergence between orthologous antisense transcript probesets at exon level and permuted pairs (after removal of overlapping RefSeq transcripts as antisense transcripts).

Panel A used Euclidean distance as a measure of expression divergence while Panel B used Pearson's correlation. "No threshold" refers to using full set of orthologous genes. "Threshold: >6.5" and "Threshold: >8" requires at least one of the orthologous probesets to have an expression intensity of more than 6.5 and 8 respectively. Error bars denote standard error.

Main Findings: Orthologous pairs show expression divergence which are significantly lower than that of permuted pairs regardless of threshold or measures, suggesting that different threshold or measures do not affect the conclusion. Comparing to Figure S26, the presence of interfering RefSeq transcripts from the opposite strand does not affect the conclusion.

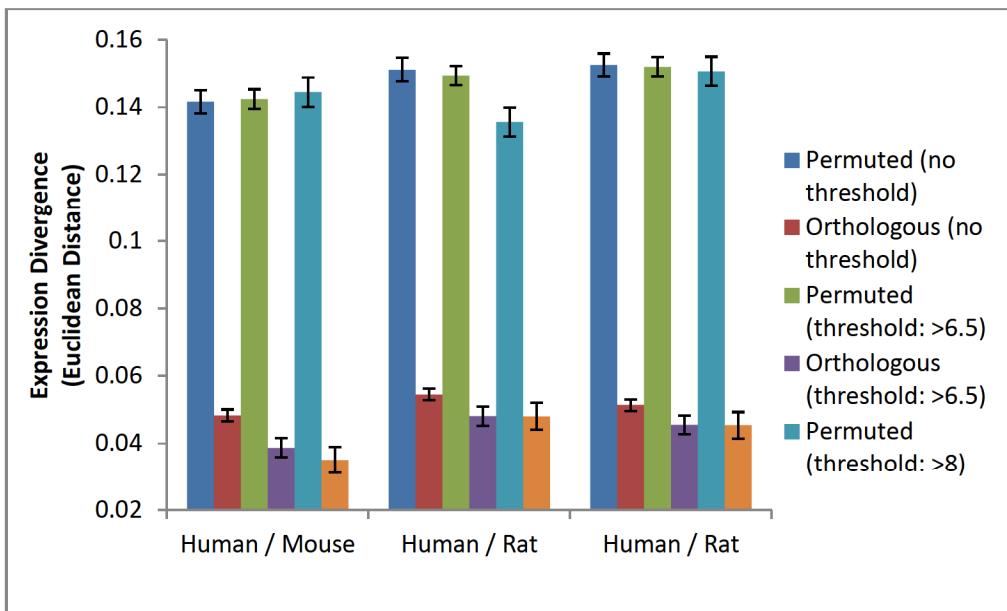
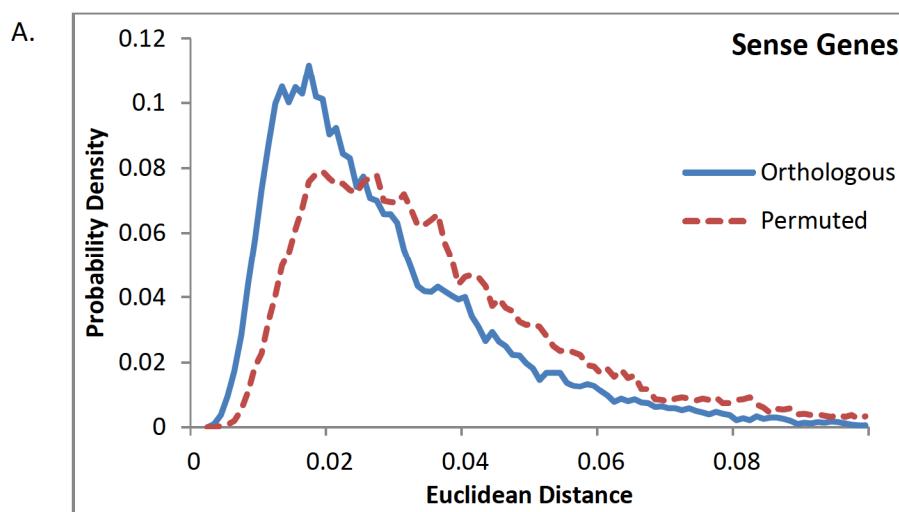


Figure S34. Average expression divergence between orthologous antisense transcript probesets at exon level and permuted pairs using Piasecka et al. (2012) randomization procedure to generate permuted pairs (after removal of overlapping RefSeq transcripts as antisense transcripts). “No threshold” refers to using full set of orthologous genes. “Threshold: >6.5” and “Threshold: >8” requires at least one of the orthologous probesets to have an expression intensity of more than 6.5 and 8 respectively. Error bars denote standard error.

Main Findings: Orthologous pairs show expression divergence which are significantly lower than that of permuted pairs regardless of threshold or measures, suggesting that different threshold or measures do not affect the conclusion. Comparing to Figure S26, the presence of interfering RefSeq transcripts from the opposite strand does not affect the conclusion.



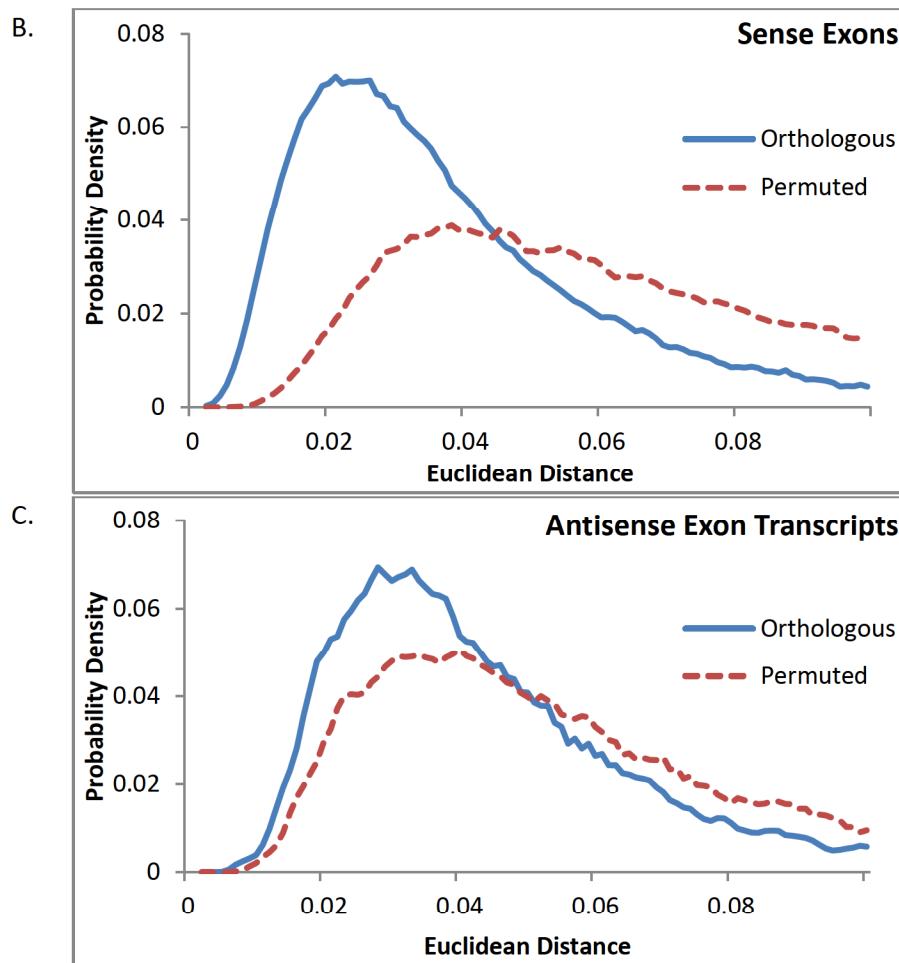


Figure S35. Distribution of mouse/rat orthologous transcripts at gene and exon level again permuted pairs. Vertical axis represents cumulative frequency. Horizontal axis represents expression divergence, calculated using Euclidean distance. Panel A shows the expression divergence distribution of protein-coding genes as defined by NCBI HomoloGene build 65 and its permuted pairs. Panel B shows the orthologous sense exons. Panel C shows antisense exon transcripts are the exon level which excluded exons with antisense RefSeq transcripts.

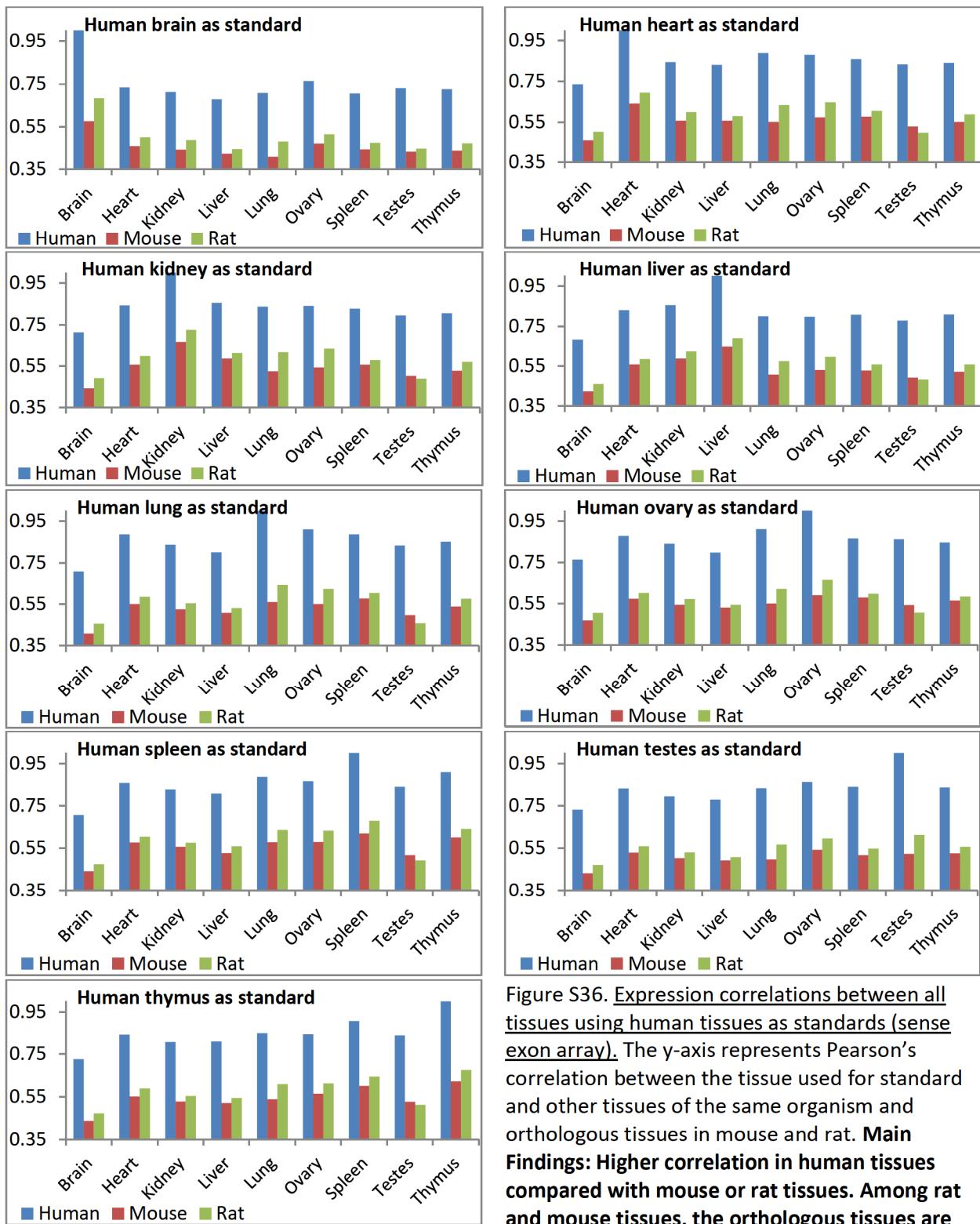


Figure S36. Expression correlations between all tissues using human tissues as standards (sense exon array). The y-axis represents Pearson's correlation between the tissue used for standard and other tissues of the same organism and orthologous tissues in mouse and rat. **Main Findings:** Higher correlation in human tissues compared with mouse or rat tissues. Among rat and mouse tissues, the orthologous tissues are more correlated than the rest.

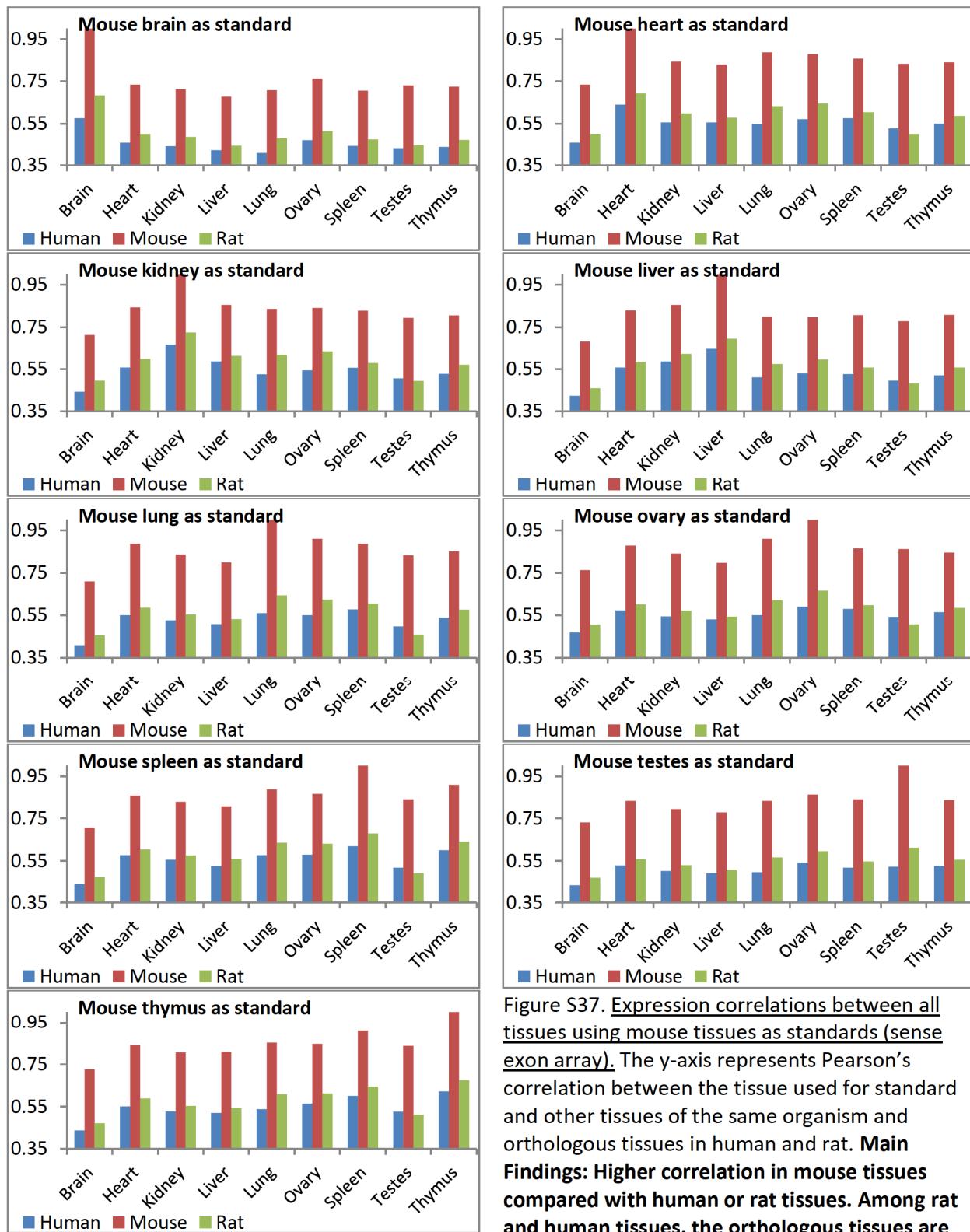


Figure S37. Expression correlations between all tissues using mouse tissues as standards (sense exon array). The y-axis represents Pearson's correlation between the tissue used for standard and other tissues of the same organism and orthologous tissues in human and rat. **Main Findings:** Higher correlation in mouse tissues compared with human or rat tissues. Among rat and human tissues, the orthologous tissues are more correlated than the rest.

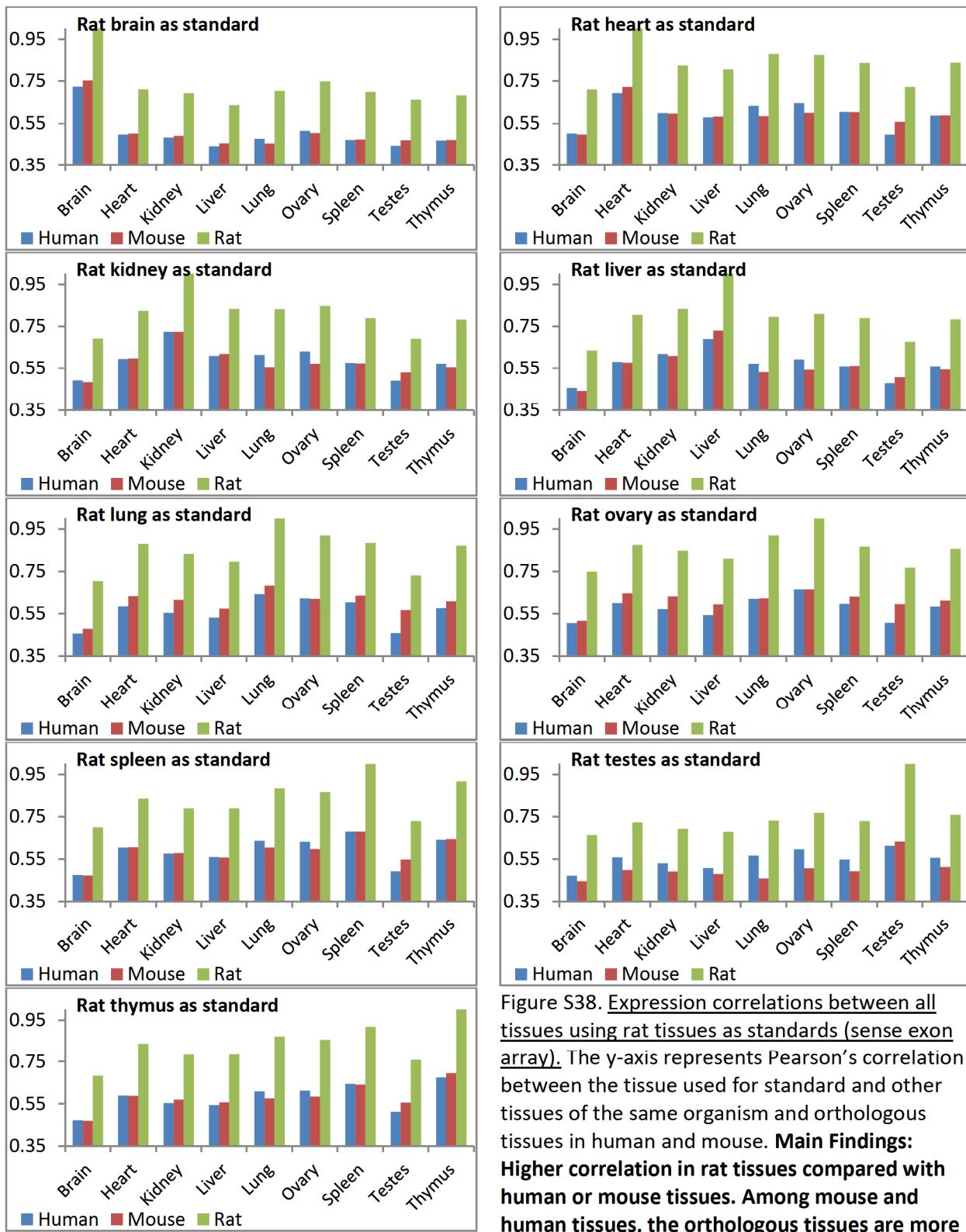


Figure S38. Expression correlations between all tissues using rat tissues as standards (sense exon array). The y-axis represents Pearson's correlation between the tissue used for standard and other tissues of the same organism and orthologous tissues in human and mouse. **Main Findings:** Higher correlation in rat tissues compared with human or mouse tissues. Among mouse and human tissues, the orthologous tissues are more correlated than the rest.

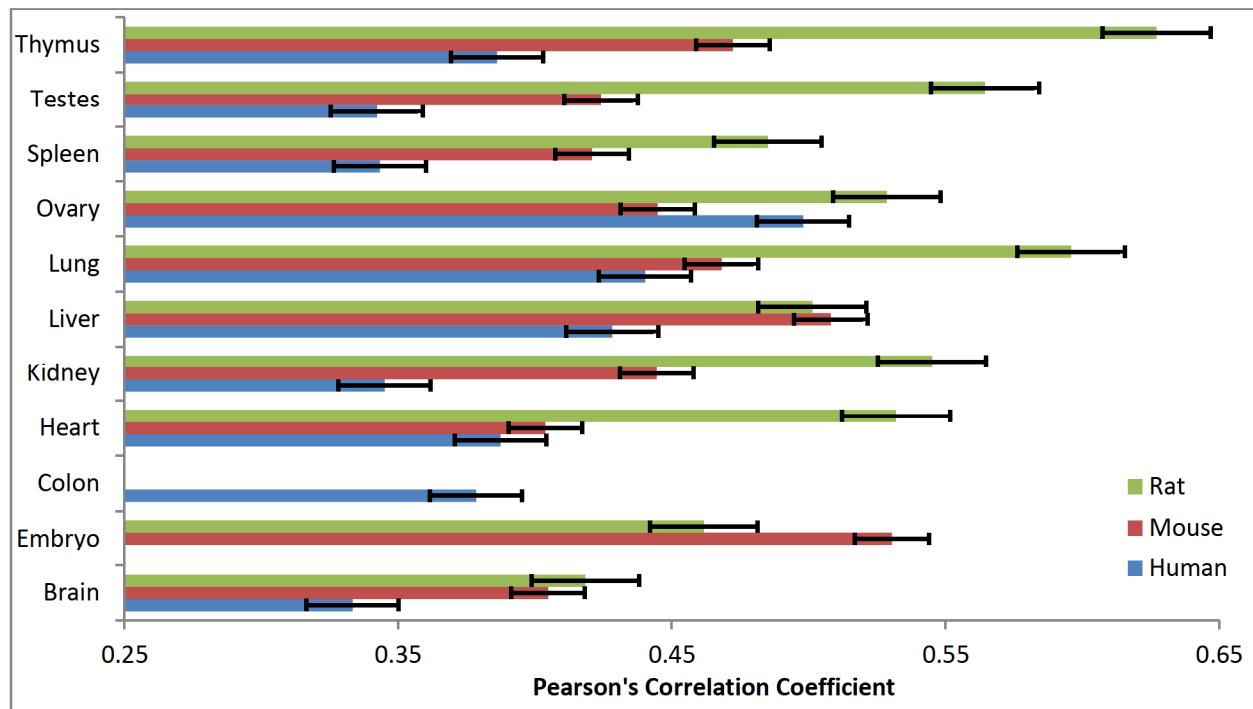


Figure S39. Correlation between sense and antisense expression levels. Error bars represent standard error.

Code S1: R code for Piasecka et al. (2012) randomization

```
#####
# randomization process according to Piasecka et al. (Bioinformatics 2012)#
#####

# PART I. Functions #
#####

# TAU calculation function
TAU_calc = function(DAT){
    X_head = t(apply(DAT,1,function(h)h/max(h)))
    N = ncol(DAT)
    apply(X_head, 1, function(h)sum(1-h)/(N-1))
}

# randomization function according to uniform TAU
Uniform_tau_rand = function(DAT, N){
    N_rand = runif(N, min=min(DAT), max=max(DAT))
    INDEX = rep(0, N)
    for(i in 1:N){
        INDEX[i] = which.min(abs(DAT-N_rand[i]))[1]
    }
    INDEX
}

# expression divergence. relative abundance
eura2 = function(data1, data2){
    Vec_d = rep(0,nrow(data1))
    for (i in 1:nrow(data1)){
        expr.1 = as.numeric(as.matrix(data1[i,]))
        expr.2 = as.numeric(as.matrix(data2[i,]))
        expr.1 = expr.1/sum(expr.1)
        expr.2 = expr.2/sum(expr.2)
        d = sqrt(sum((expr.1-expr.2)^2))
        if(is.na(d)){
            Vec_d[i]=NA
        }else{
            Vec_d[i]=d
        }
    }
    Vec_d
}

#####

# PART II. Main #
#####

main_function = function(N, sourcefile, h_outputfile, r_outputfile){
    # READ data
    DATA = read.csv(sourcefile)
```

```

# split data with labels "1" and "2" representing the 2 sets of tissues
DATA_s1 = DATA[,grep("1",names(DATA),value=T)]
DATA_s2 = DATA[,grep("2",names(DATA),value=T)]

# calculate TAU
TAU_s1 = TAU_calc(DATA_s1)
TAU_s2 = TAU_calc(DATA_s2)

# randomized uniform TAU with N, result is a vector of
# randomized indexs corresponding to gene expression data
Index_rand_s1 = Uniform_tau_rand(TAU_s1, N)
Index_rand_s2 = Uniform_tau_rand(TAU_s2, N)

# calculate expression divergence, result is a vector
# empirical ED
ExprDiv_empirical = eura2(DATA_s1, DATA_s2)

# ED calculated from randomized data
ExprDiv_rand = eura2(DATA_s1[Index_rand_s1, ], DATA_s2[Index_rand_s2, ])

# write homologous and randomized EDs into files
write.table(ExprDiv_rand, r_outputfile, row.names=F,col.names=F,quote=F,sep="\t")
write.table(ExprDiv_empirical, h_outputfile, row.names=F,col.names=F,quote=F,sep="\t")
}

# main_function(<number of randomized pairs>,
#               <Orthologous sample file (CSV)>
#               <Orthologous pairs ED filename (for output)>
#               <Randomized pairs ED filename (for output)>
main_function(30000, "Human-Mouse-Antisense-Exon.csv", "homologous.txt", "random.txt")

```