BioData Mining

METHODOLOGY

Open Access

Robust and rigorous identification of tissue-specific genes by statistically extending tau score

Hatice Büşra Lüleci[1] and Alper Yılmaz[2*]

*Correspondence:
alyilmaz@yildiz.edu.tr

[1] Department of Bioengineering, Gebze Technical University, Kocaeli, Turkey
[2] Department of Bioengineering, Yildiz Technical University, Istanbul, Turkey

Abstract

Objectives: In this study, we aimed to identify tissue-specific genes for various human tissues/organs more robustly and rigorously by extending the tau score algorithm.

Introduction: Tissue-specific genes are a class of genes whose functions and expressions are preferred in one or several tissues restrictedly. Identification of tissue-specific genes is essential for discovering multi-cellular biological processes such as tissue-specific molecular regulations, tissue development, physiology, and the pathogenesis of tissue-associated diseases.

Materials and Methods: Gene expression data derived from five large RNA sequencing (RNA-seq) projects, spanning 96 different human tissues, were retrieved from ArrayExpress and ExpressionAtlas. The first step is categorizing genes using significant filters and tau score as a specificity index. After calculating tau for each gene in all datasets separately, statistical distance from the maximum expression level was estimated using a new meaningful procedure. Specific expression of a gene in one or several tissues was calculated after the integration of tau and statistical distance estimation, which is called as extended tau approach. Obtained tissue-specific genes for 96 different human tissues were functionally annotated, and some comparisons were carried out to show the effectiveness of the extended tau method.

Results and Discussion: Categorization of genes based on expression level and identification of tissue-specific genes for a large number of tissues/organs were executed. Genes were successfully assigned to multiple tissues by generating the extended tau approach as opposed to the original tau score, which can assign tissue specificity to single tissue only.

Keywords: Tissue-specific genes, RNA-Seq, Tau score

Introduction

Protein-coding genes in the human genome demonstrate dramatic diversity in terms of expression levels and patterns [1]. Different transcripts are expressed in diverse organs, tissues, or cell types and in different developmental stages. An interesting subset of genes are observed which are strictly expressed in one, or several tissues/organs hence called tissue-specific genes [2, 3]. Identification and analysis of tissue specificity as a

dynamic and complex phenomenon, in combination with other biomedical data, provide crucial insights into molecular mechanisms, developmental processes [3, 4], expression-quantitative trait loci [5] and evolution of tissues/organs [6]. Moreover, tissue-specific genes are associated with prognosis, etiology of diseases, and discovery of novel specific drug targets as significant biomarkers for many complex diseases such as solid tumors, neurodegenerative and cardiovascular diseases [7–12]. There have been many studies that examine and determine restricted expression of genes in particular tissues and their relationships with diseases [8, 13–16].

Su et al. [17] and Liang et al. [18] independently generated tissue-specific mRNA expression profiles using a large number of healthy tissue types through microarray. VeryGene tool, which shows relationship between tissue-specific genes with diseases and drugs, was enhanced by Yang et al. [15]. Even though tissue specificity is often used in various researches [19], there is no gold standard method to identify it. Several databases were developed to establish a knowledge base of tissue-specific gene expression in a variety of human tissues. However, consensus among databases is weak due to diverse assumptions, methods, experimental procedures and data types used by those databases [19].

Calculation methods can be divided into two major groups [19] based on their outputs. The first group, including Tau specificity index [20], Gini [19], Tissue Similarity Index (TSI) [21], and Shannon entropy (Hg) [22], produce a single specificity score per gene indicating whether a gene has specific or wide-spread expression. Since genes can be specifically expressed in more than one tissue, producing a single score and pointing out only one tissue is a crucial deficiency of these methods. The second group of specificity calculation methods, which includes Z-score [23], Specificity Measure (SPM) [2], Expression Enrichment (EE) [1], and Preferential Expression Measure (PEM) [24], produce scores as many as the number of tissues for each particular gene and tissue specificity of a gene have to be decided according to threshold values. However, varying thresholds can cause incorrect and inconsistent results.

Previous research has attempted to identify tissue-specific genes with various approaches. Shannon entropy, [22] similar to TSI [19] was used in ROKU, [25] a tool for selection of tissue-specific patterns from microarray data. Tau specificity index was defined as a gene characterization score, and it is a quantitative, graded scalar measurement of specificity of gene expression [20]. Gene Expression and Regulation (TiGER) database [26] was established based on EE score [1], but using obsolete data type and containing data for the low number of tissues renders the database insufficient for extensive research. Z-score [23] approach considers absolute distance from the mean, thus favoring mostly over-expressed genes and occasionally under-expressed genes as tissue-specific genes [19, 27]. In other words, a gene showing housekeeping gene expression with high expression in a tissue would be considered a tissue-specific expression by Z-score calculation. Genotype-Tissue Expression (GTEx) [28] identified tissue-specific genes via Z-score for 53 different tissue types. Both TiSGeD [2] and A Pattern Gene Database (PaGenBase) [29] use SPM to calculate tissue specificity. However, they have a weak correlation in specificity results. PEM calculation proposed by Huminiecki et al. using EST and microarray data from SAGEmap [30], Gene Expression Atlas [31], and TissueInfo [32] databases is a simple form of the EE score [24]. SPM, PEM, and EE are

normalized by either maximum expression of a gene or by the sum of gene expressions. Hence they are not sensitive to absolute expression level [19]. Besides,there are some marker gene detection approaches such as CellMapper [33] and Marker Gene Finder in Microarray (MGFM) [34]. However, they are limited to several tissues and/or microarray and EST data omitting RNA-Seq data. Despite presence of multiple methods for calculating tissue-specific expression of a gene, these methods suffer from serious shortcomings. Thus, developing a more robust and rigorous method using more datasets is an important requirement for identifying tissue-specific genes.

Tau is shown to be a more effective method for providing accurate and consistent results in different datasets [19]. It is calculated to determine tissue specificity or sharing of genes across each tissue [35]. However, the tau index is limited to identifying only one tissue in terms of specificity of a gene. In other words, tau can assign a gene to a single tissue, not multiple tissues. Since the definition of tissue-specific genes is considered to be "specifically expressed in one or several tissues", tau method needs to be improved by additional statistical procedures to assign genes to multiple tissues for specific expression. In this study, estimation of statistically significant interval from maximum expression was calculated to assign a gene to second and/or more tissues for the genes having high tau scores. Therefore, this study makes a major contribution to research on determining tissue-specificity by extending the already effective tau method allowing one-to-many mappings between genes and tissues. Throughout this paper, the term **extended tau** will refer to our novel and rigorous approach for assigning genes to multiple tissues for specific expression. More detailed and accurate tissue specificity of gene expression will enhance understanding evolution of tissues [36–39], relationship between expressions and main functions of genes [20, 40]; and others [41] in various organisms such as mouse [42], *Drosophila* [40] and *Arabidopsis thaliana* [43].

Methods

Data retrieval

RNA-seq data for gene expression profiles of 27 human tissues from Fagerberg et. al (EMTAB-1733) [44], 32 human tissues from Uhlen Lab (EMTAB-2836) [45], 53 human tissues from GTEx Project (EMTAB-5214) [28], 56 human tissues from FANTOM5 Project (EMTAB-3358) [46] and 13 human tissues from ENCODE Project (EMTAB-4344) [47, 48] were downloaded via Expression Atlas [49] and ArrayExpress [50]. Detailed information about the raw expression data, number of genes, and tissues are explained in Supplementary Tables 1 and 2, respectively. All calculations were performed using protein-coding genes and tissue types not cell types from the datasets. All tissue types were investigated and grouped according to localization determined via Brenda Tissue Ontology (BTO) [51].

Categorization of genes based on expression level

Genes were categorized according to their expression level patterns and tau scores. Genes expressed ≤ 1.0 FPKM or TPM in all tissues were designated as "Null expression" and were excluded from subsequent analysis. Then, expression levels were transformed based on log(2), and the tau score, ranging from 0 to 1, was calculated

for each gene [19] using the formula below where x_i is expression of a gene in tissue i and n is number of tissues.

$$\tau = \frac{\sum_{i=1}^{n} \left(1 - \hat{x}_i\right)}{n-1}$$

$$\hat{x}_i = \frac{x_i}{\max_{1 \le i \le n} x_i}$$

If tau score is $(\tau) \geq 0.85$ for a given gene, that gene is marked to have *Specific expression*. The genes having $\tau < 0.85$ were classified as *Wide-spread expression*. Genes which have expression values < 10 in all tissues was denoted as *Weak expression* [8]. Tau score was calculated for weakly expressed genes but their scores were ignored during tissue specificity assessment. Log transformation was used only during tau calculation; after that, all other calculations were performed using raw expression values. Rigorous tissue-specificity classification was proceeded with the genes with $\tau \geq 0.85$ and expression value > 10 in all tissues.

Estimation of statistically significant interval

F-test [52] was used to verify the equality of variance between datasets. Statistically significant distance from the maximum expression value was calculated in order to assign genes to multiple tissues in the context of specificity. For this purpose, the lower and upper bounds of raw expression data were calculated via Fuzzy c-means clustering [53]. Ratio of upper cluster was calculated for each tissue with the following formula where n_{up} is the number of elements in upper cluster and n_{total} is the number of total non-zero elements.

$$ratio = \frac{n_{up}}{n_{total}}$$

Regression analysis [54] was used to calculate an optimized threshold value of *ratio* and then converted it to Z-value using the inverse function of normal distribution for each dataset. Assessment of normality of datasets was performed by Kolmogorov-Smirnov (K-S) test [55] and Q-Q plots. The equation below was used to calculate statistically significant distance ($dist_{ss}$) where x_{max} is the maximum expression value of a gene among all tissues, σ is the standard deviation of non-zero expression of a gene among all tissues and z_{val} is optimized threshold as Z-value.

$$dist_{ss} = x_{max} - \sigma \times z_{val}$$

All calculations were performed for all datasets separately, and threshold ratios are available in Supplementary Table 3. Integration of tau score with the statistically significant interval from maximum expression was described as **extended tau** for robust and rigorous identification of tissue-specific genes via assignment of genes to possible multiple tissues. Extended tau approach is illustrated in Fig. 1.

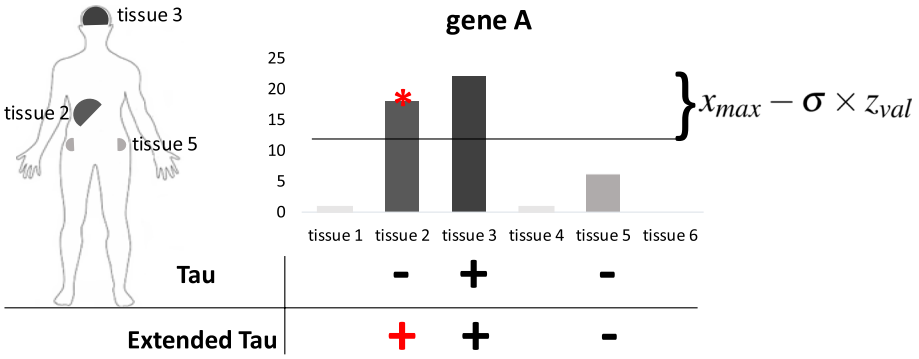

Fig. 1 Illustration of extended tau approach. Gene can specifically expressed one or more different tissues. Here, gene A is specifically expressed in tissue 2 and tissue 3. Extended tau can determine both of two compared to only tau calculation

Functional annotation of tissue-specific genes

Database for Annotation, Visualization and Integrated Discovery (DAVID) [56] was used to identify the roles of specific genes in biological processes, potential functions, related tissues, and diseases. Results were compared to GeneCards [57] which already integrated expression data from GTEx [28], Illumina Body Map [58], BioGPS [59], and CGAP SAGE [60]. Moreover, The Human Protein Atlas [45] was also used to compare transcript levels of tissue-specific genes with their protein levels.

Results and Discussion

Tau score is a robust method to identify tissue-specific genes [19] but is limited in its capacity to match genes with multiple tissues. To overcome this, we developed a new extensive procedure where the specific expression of a gene in one or several tissues was calculated by integrating tau score with statistical distance as a new rigorous approach described as extended tau calculation.

RNA-Seq data from five studies aimed to determine gene expression in multiple tissues were retrieved from publicly available databases. When all datasets are combined, expressions from the total of 96 different tissues are represented. The tissues had parent-child relationships with various depths in the hierarchy. After assigning parent-child mappings using Brenda Tissue Ontology (BTO), 96 different tissue types, referred to as *child tissue*, were mapped to 36 top-level tissues, referred to as *parent tissue*. All tissue types and their BTO accession IDs are listed in Supplementary Table 4.

A summary of the workflow for categorizing protein-coding genes and calculating tissue-specific genes using extended tau is presented in Fig. 2, and a detailed workflow is shown in Supplementary Fig. 1. F-test was used to demonstrate whether there was a significant difference among datasets. According to F-test results, expression values in datasets were found to have equal variances (Supplementary Table 5). The higher F-score value for the dataset EMTAB-3358 is due to the fact that the dataset has units of TPM compared to FPKM in other datasets. Therefore, EMTAB-3358 is distinct from other datasets; still, there is no significant difference among the datasets. Boxplots and violin plots showing the distribution of expression levels for each

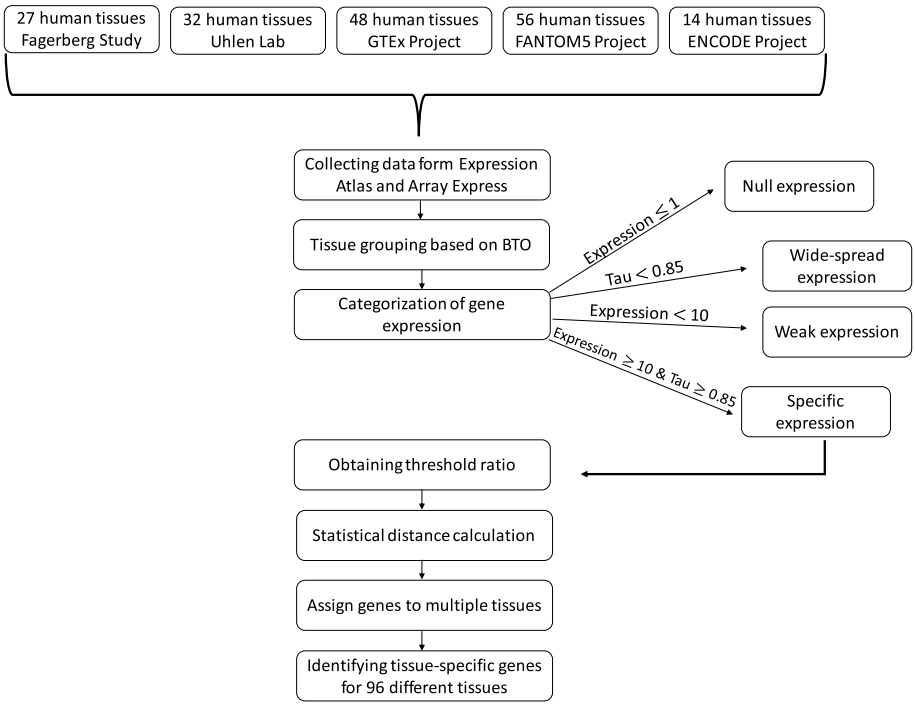

Fig. 2 Workflow for the identification of tissue-specific genes. Firstly, genes were categorized based on some significant filters and tau calculation. After that, statistical distance estimation was used to determine specific genes in a rigorous manner

Table 1 Number of genes in each category for all datasets

Gene profile	Fagerberg Study	Uhlen Lab	GTEx Project	FANTOM5 Project	ENCODE Project
Null expression	1,260	2,427	2,672	1,808	3,394
Weak expression	1,808	2,533	2,788	3,869	2,976
Wide-spread expression	13,126	11,733	11,434	8,698	11,013
Specific expression	2,669	2,983	2,782	2,063	2,293
Total number of genes	18,863	19,676	19,676	16,438	19,676

dataset are shown in Supplementary Fig. 2. Kolmogorov-Smirnov test shows that datasets have a normal distribution, and Q-Q plots for each dataset are presented in Supplementary Fig. 3.

According to criteria described in Fig. 2, the genes were categorized based on their expression level profiles and tau scores. Table 1 summarizes number of genes in each category for each dataset. Please note that total number of genes in the Table 1 matches the number of filtered genes in Supplementary Table 2.

According to Table 1, number of genes showing specific expression is comparable among all samples. After this categorization, statistically significant interval from maximum expression was applied to genes showing specific expression to assign them to multiple tissues. As expected, the extended tau approach reveals more gene-tissue pairs when compared to original the tau calculation as listed in Table 2.

Table 2 Number of specific gene-tissue pairs based on tau and extended tau calculations

Datasets	Tau	Extended Tau
Fagerberg Study	2,669	3,370
Uhlen Lab	2,983	4,257
GTEx Project	2,782	4,680
FANTOM5 Project	2,063	3,982
ENCODE Project	2,293	3,097

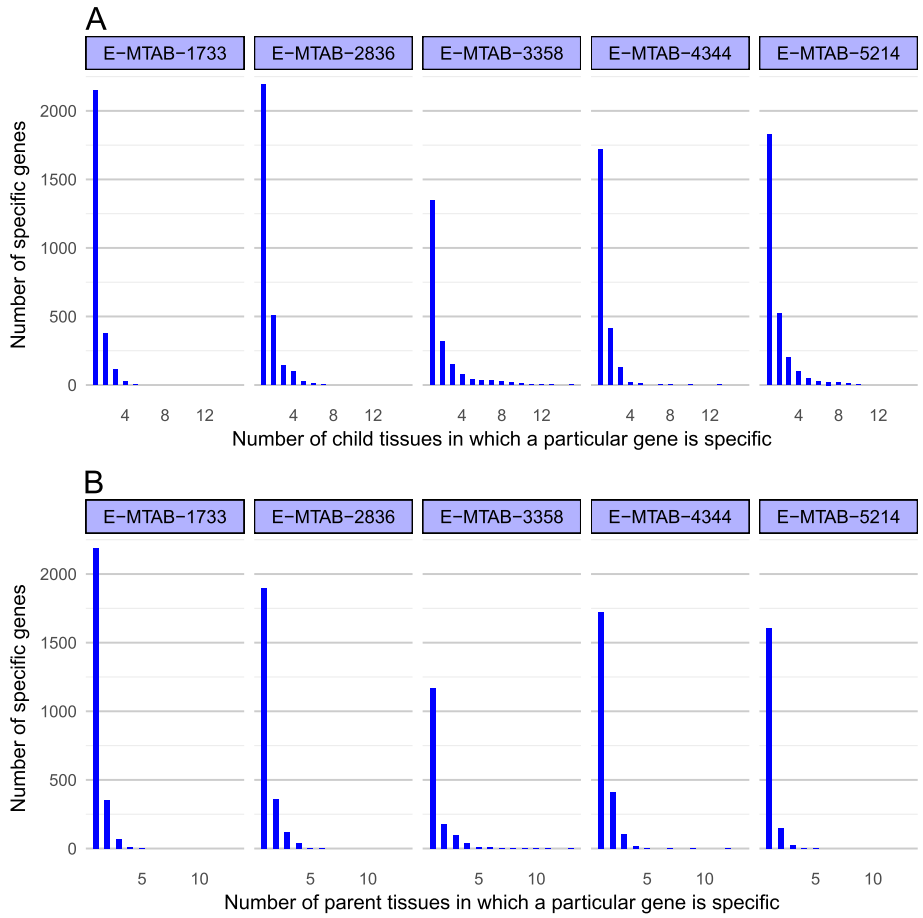

Fig. 3 Distribution of tissue-specific genes. Extensive graph shows how many genes are specific in how many tissues per dataset for all child (A) and parent tissues (B). In each plot, the leftmost bar represents number of tissue-specific genes that are specific to single tissue. Remaining bars show number of genes that are specific to multiple tissues which was calculated by multiple assignment of genes to tissues based on extensive tau

Although the tau scores of genes and the number of tissue-specific genes are same between two approaches, the extended tau approach provides more gene-tissue mappings for specific expression. The extended tau approach has ability to list gene-tissue mappings for genes specific to one or several tissues [2, 3]. The number of specific genes per tissue is provided in Supplementary Table 6, and Fig. 3 summarizes number of tissue-specific genes distributed by number of gene-tissue mappings for both parent and

child tissues in each dataset. Majority of genes were expressed specifically to only one single tissue and some genes are specifically expressed in two or more different tissues regardless of tissue hierarchy, parent or child.

We used DAVID and other resources to show coherences between our results and the functions of several genes. Alpha fetoprotein (AFP) is a liver-specific gene [61]; however, it is defined not only liver-specific but also kidney-specific gene based on extended tau. One of its related pathways is the glucocorticoid receptor regulatory network, and the level of AFP in amniotic fluid is used to measure renal loss of protein. Intestinal Alkaline Phosphatase (ALPI), which encodes a digestive brush-border enzyme, has a specific expression in the small intestine based on the tau calculation, although extended tau demonstrates that ALPI is specifically expressed in both small intestine and duodenum. Another example is MAP7 Domain Containing 2 (MAP7D2) which contributes to the structural integrity of a complex is specifically expressed in the brain based on tau, even though it is also specifically expressed in the testis. D-amino acid oxidase (DAO) has specific expression in kidney according to tau score. On the other hand, it was noticed that DAO is specific to brain and liver after calculation of extended tau. DAO may act as a detoxifying agent which removes D-amino acids that accumulate during aging. It is generally related to some neurological diseases. Shortly, extended tau is a more comprehensive approach to find several tissues for one specific gene.

The brain is a complex organ characterized by a high level of gene expression; at least 30-50% of approximately all protein-coding genes are expressed across all parts of the brain and it has a significant variety of functions [62, 63]. Significant differences in cell composition of the various anatomical brain regions result in cell-specific differences in gene expression. There are many specific genes all over the brain as parent tissue [64]. Comparison in terms of tissue-specific genes in child tissues of brain was not performed because, there are many different brain parts coming from different datasets. It can be stated that brain-specific genes are more often selectively expressed in either neurons or glial cells and vascular cells from the cerebral cortex [65]. The cerebral cortex has a higher number of specific genes shown in Supplementary Fig. 4 as child tissue.

The datasets used in this study did not examine the same set of tissues. Therefore, a gene is not necessarily specific in a particular tissue, according to five datasets. If five datasets have a certain consensus presented in Supplementary Fig. 5 for the specificity of a particular gene, we can be sure that it is absolutely specific to related tissue. For instance, 252 genes are specific to the testis according to all five datasets, and 262 genes are specific to the testis supported by 4 datasets shown in Supplementary Fig. 5. Agreement of datasets is important for accuracy. On the other hand, if tissues are included in only one single dataset, genes specific to that tissue will be supported by only one dataset, naturally. Gene expression in whole blood, tongue, epididymis, seminal vesicle, tibial nerve, Vas deferens, and spinal cord are examined in only one dataset. Therefore, genes specific to these tissues are supported by one dataset.

Figure 4 summarizes the comparison of tissue-specific gene lists across different datasets. Four hundred eighty-four genes were found to be tissue-specific according to all datasets. Although some genes are found to be specifically expressed in only one dataset, this result suggested that our approach is suitable and effective for determining tissue-specific genes. Comparison is very important for the correctness and reliability of the

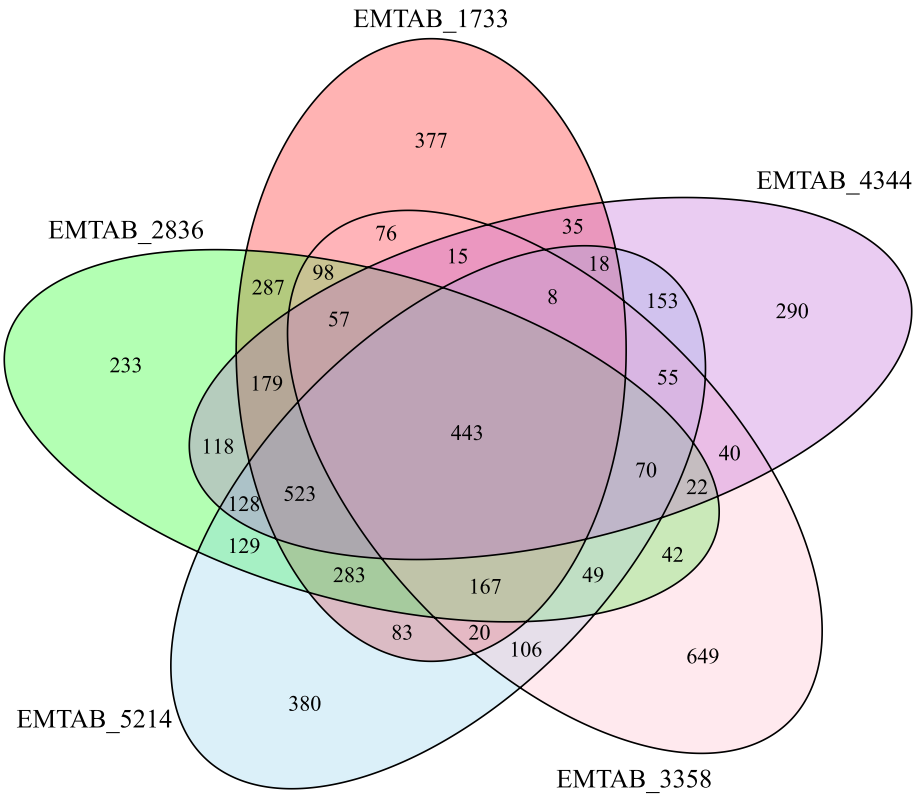

Fig. 4 Venn diagram of tissue-specific gene lists derived from five different datasets. Comparison of results depends on the number of tissue-specific genes was illustrated via Venn Diagram. Intersected genes among five datasets are absolutely specific to related tissues

results. Naturally, there are differences in the results as they are created with different experimental and laboratory conditions, samples, and also different normalization methods during obtaining RNA-seq data. If we examine Fig. 4, 692 genes are determined to be tissue-specific by only EMTAB-3358 (FANTOM5 Project) because expression values were normalized using a different method, despite the same experimental procedure.

The correlation of raw expressions of datasets is demonstrated in Supplementary Fig. 6, and similarity is defined as a dark color; the size of nodes shows the number of genes common between two datasets. It can be shown that the similarity of EMTAB-5214 (GTEx Project) and EMTAB-2836 (Uhlen Lab) is higher than any other dataset pairs, according to Supplementary Fig. 6. EMTAB-3358 (FANTOM5 Project) is quite different from all other datasets, as observed in the previous results. After examining the datasets for similarity of raw gene expression, the datasets were compared to each other to understand the correlation, reliability, and effectiveness of the extended tau method. Genes have a tau score greater than 0.85, and an accompanying correlation plot has been drawn in Supplementary Fig. 7. It was shown that EMTAB-4344, EMTAB-1733, and EMTAB-2836 give similar results after determining tissue specificity. As before, EMTAB-3358 is also far from the other datasets. Two data that provide the closer results are EMTAB-1733 and EMTAB-2836. On the other hand, the two data that give the most distant results are EMTAB-3358 and EMTAB-4344, according to Supplementary Fig. 7.

Tissue-specific expression profiles can be used for biomedical applications such as tissue-specific regulation [66] of genes, examining gene profiles for various disorders, enhancing the efficiency of therapies, discovering new biomarkers for diagnosis and also targeted treatment of diseases such as cancer [67] and malignancies [68]. Organogenesis is another important phenomenon related to biological processes [36, 37]. The progression of tissues, organs, and systems in living organisms can be understood by identifying tissue-specific genes and their roles. Besides, interpretation of relationship between tissues in the context of specific genes is a very crucial approach to find out not only tissue progression but also the discovery of mechanisms of diseases. Common tissue-specific genes between different tissues might give clues to unravel relationships between various tissues. Interestingly, the brain has connections to all of the tissues because brain expression may reflect developmental ontogeny, or developmental stages processes of the human body [62, 69].

Bone marrow, spleen, and lymph node are tightly connected in terms of specific genes and it is known that they are members of the lymph system [70]. A group of tissues is related to female reproductive system, including vagina, uterus, ovary, and oviduct [71] that express a list of common specific genes. Another case concerns human digestive system organs which are small intestine, colon, rectum, liver, and stomach which have some common tissue-specific genes. However, it has also been observed that some specific genes related to digestive system are associated with kidney. Different organs/tissues can have similar subsequent processes such as ammonia-urea conversion [72, 73] and the single gene can be specific to several organs. In addition, epididymis, Vas deferens, penis, and testis are tightly connected to each other as parts of male reproductive system. They have shared tissue-specific genes according to the extended tau results and our findings are consistent with both literature and Brenda Tissue Ontology (BTO).

Although all tissues carry out common processes in the human body, tissues can be distinguished by gene expression levels [74]. After filtering out transcripts with low-level and wide-spread expressions, protein-coding genes were assessed for specific expression in tissues using a robust and rigorous calculation, extended tau. Tissue-specific genes were successfully assigned to multiple tissues and identified with great care.

Conclusion

This study provides a large insight into tissue specificity. Tissue-specific genes for 96 different tissue samples from the human body via five RNA-seq datasets were calculated by the extended tau approach. Tau score is a more effective and accurate calculation method, and the statistical distance term is meaningful for assigning genes to several specific tissues. After the categorization of protein-coding genes and identification of tissue-specific genes in a broad sense, their functional properties were investigated. It can be suggested that tissue specificity results will benefit further studies to reveal molecular mechanisms of healthy tissues and diseases.

The extended tau approach can be used in other regulatory elements like transcription factor (TF) [74]. TFs may have higher tissue specificity, because tissue-specific processes are ultimately controlled by gene regulatory networks [74, 75]. Therefore computational analysis of tissue-specific TFs and other regulatory networks will provide a critical perspective.

RNA-Seq data of a single tissue can include genes which are specific to other tissues after calculation of tissue specificity. This condition may be related to change of expression level or migration of cells which are originated from other tissues/organs. When tissue heterogeneity, cell migration, change of expression level or behaviors of genes would like to be examined, the tau score will be incomplete. In this situation, the extended tau approach can give more robust and rigorous results for various research.

Abbreviations

EE	Expression Enrichment
FPKM	Fragments per kilobase per million reads
GTEx	Genotype-Tissue Expression
Hg	Shannon entropy
PaGenBase	Pattern Gene Database
PEM	Preferential Expression Measure
RNA-Seq	RNA sequencing
SPM	Specificity Measure
TiGER	Gene Expression and Regulation
TPM	transcript per million
TSI	Tissue Similarity Index

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13040-022-00315-9.

Additional file 1. Algorithm to determine tissue specific genes.

Additional file 2. Distribution of gene expression levels in all datasets - violin and boxplots.

Additional file 3. Distribution of gene expression levels in all datasets.

Additional file 4. Number of genes for child and parent tissues.

Additional file 5. Number of genes per tissue in all the datasets.

Additional file 6. Correlation of raw expressions of datasets.

Additional file 7. Correlation of gene expressions for genes which have tau score greater than 0.85.

Additional file 8. Descriptive summary of all datasets.

Additional file 9. Number of genes and tissues in each dataset.

Additional file 10. Optimized thresholds and Z-scores for each dataset.

Additional file 11. Determining parent-child tissue relationship.

Additional file 12. F-test results for each pair of datasets.

Additional file 13. Number of tissue specific genes for each tissue.

Acknowledgements
We are thankful to Muhammed Raşit Cesur for estimation of significance intervals.

Authors' contributions
Alper Yilmaz contributed to analysis of data and edited the manuscript. Hatice Büşra Lüleci gathered the data and did the analysis and wrote the manuscript. The author(s) read and approved the final manuscript.

Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.

Availability of data and materials
Codes are available at https://gitlab.com/busra/modified_tau.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Received: 12 August 2022 Accepted: 11 November 2022
Published online: 09 December 2022

References

1. Yu X, Lin J, Zack DJ, Qian J. Computational analysis of tissue-specific combinatorial gene regulation: predicting interaction between transcription factorsin human tissues. Nucleic Acids Res. 2006;34(17):4925–36.
2. Xiao SJ, Zhang C, Zou Q, Ji ZL. TiSGeD: a database for tissue-specific genes. Bioinformatics. 2010;26(9):1273–5. https://doi.org/10.1093/bioinformatics/btq109.
3. Kim P, Park A, Han G, Sun H, Jia P, Zhao Z. TissGDB: tissue-specific gene database in cancer. Nucleic Acids Res. 2017;46(D1):D1031–8. https://doi.org/10.1093/nar/gkx850.
4. Jiang W, Chen L. Tissue Specificity of Gene Expression Evolves Across Mammal Species. J Comput Biol. 2022;29(8):880–91. https://doi.org/10.1089/cmb.2021.0592.
5. Petretto E, Mangion J, Dickens NJ, Cook SA, Kumaran MK, Lu H, et al. Heritability and Tissue Specificity of Expression Quantitative Trait Loci. PLoS Genet. 2006;2(10): e172. https://doi.org/10.1371/journal.pgen.0020172.
6. Nagaraj SH, Ingham A, Reverter A. The interplay between evolution, regulation and tissue specificity in the Human Hereditary Diseasome. BMC Genomics. 2010;11(Suppl 4):S23. https://doi.org/10.1186/1471-2164-11-s4-s23.
7. Lage K, Hansen NT, Karlberg EO, Eklund AC, Roque FS, Zoltan Szallasi PKD, et al. A large-scale analysis of tissue-specific pathology and gene expression of human disease genes and complexes. Proc Natl Acad Sci USA. 2008;105(52):20870–5. https://doi.org/10.1073/pnas.0810772105.
8. Dezső Z, Nikolsky Y, Sviridov E, Shi W, Serebriyskaya T, Dosymbekov D, et al. A comprehensive functional analysis of tissue specificity of human gene expression. BMC Biol. 2008;6(1):49. https://doi.org/10.1186/1741-7007-6-49.
9. Song Y, Ahn J, Suh Y, Davis ME, Lee K. Identification of Novel Tissue-Specific Genes by Analysis of Microarray Databases: A Human and Mouse Model. PLoS ONE. 2013;8(5): e64483. https://doi.org/10.1371/journal.pone.0064483.
10. Nguyen TT, Almon RR, DuBois DC, Sukumaran S, Jusko WJ, Ioannis P. Androulakis: Tissue-Specific Gene Expression and Regulation in Liver and Muscle following Chronic Corticosteroid Administration. Gene Regul Syst Biol. 2014;8:75–87.
11. Rodemoyer A, Kibiryeva N, Bair A, Marshall J, O'Brien JE, Bittel DC. A tissue-specific gene expression template portrays heart development and pathology. Hum Genomics. 2014;8(1). https://doi.org/10.1186/1479-7364-8-6.
12. Kitsak M, Sharma A, Menche J, Guney E, Ghiassian SD, Loscalzo J, et al. Tissue Specificity of Human Disease Module. Sci Rep. 2016;6(1). https://doi.org/10.1038/srep35241.
13. Greco D, Somervuo P, Lieto AD, Raitila T, Nitsch L, Castrén E, et al. Physiology, Pathology and Relatedness of Human Tissues from Gene Expression Meta-Analysis. PLoS ONE. 2008;3(4):e1880.
14. Reverter A, Ingham A, Dalrymple BP. Mining tissue specificity, gene connectivity and disease association to reveal a set of genes that modify the action of disease causing genes. BioData Min. 2008;1(1). https://doi.org/10.1186/1756-0381-1-8.
15. Yang X, Ye Y, Wang G, Huang H, Yu D, Liang S. VeryGene: linking tissue-specific genes to diseases, drugs, and beyond for knowledge discovery. Physiol Genomics. 2011;43(8):457–460. https://doi.org/10.1152/physiolgenomics.00178.2010.
16. Greene CS, Krishnan A, Wong AK, Ricciotti E, Zelaya RA, Himmelstein DS, et al. Understanding multicellular function and disease with human tissue-specific networks. Nat Genet. 2015;47(6):569–76. https://doi.org/10.1038/ng.3259.
17. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci USA. 2004;101(16):6062–7. https://doi.org/10.1073/pnas.0400782101.
18. Liang S, Li Y, Be X, Howes S, Liu W. Detecting and profiling tissue-selective genes. Physiol Genomics. 2006;26(2):158–162. https://doi.org/10.1152/physiolgenomics.00313.2005.
19. Kryuchkova-Mostacci N, Robinson-Rechavi M. A benchmark of gene expression tissue-specificity metrics. Brief Bioinform. 2016;bbw008. https://doi.org/10.1093/bib/bbw008.
20. Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, et al. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics. 2004;21(5):650–659. https://doi.org/10.1093/bioinformatics/bti042.
21. Julien P, Brawand D, Soumillon M, Necsulea A, Liechti A, Schütz F, et al. Mechanisms and Evolutionary Patterns of Mammalian and Avian Dosage Compensation. PLoS Biol. 2012;10 (5):e1001328. https://doi.org/10.1371/journal.pbio.1001328.
22. Schug J, Schuller WP, Kappen C, Salbaum JM, Bucan M, Stoeckert CJ. Promoter features related to tissue specificity as measured by Shannon entropy. Genome Biol. 2005;6(4). https://doi.org/10.1186/gb-2005-6-4-r33.
23. Cheadle C, Vawter MP, Freed WJ, Becker KG. Analysis of Microarray Data Using Z Score Transformation. J Mol Diagn. 2003;5(2):73–81. https://doi.org/10.1016/s1525-1578(10)60455-2.
24. Huminiecki L, Lloyd AT, Wolfe KH. Congruence of tissue expression profiles from Gene Expression Atlas, SAGEmap and TissueInfo databases. BMC Genomics. 2003;4(1). https://doi.org/10.1186/1471-2164-4-31.
25. Kadota K, Ye J, Nakai Y, Terada T, Shimizu K. ROKU: a novel method for identification of tissue-specific genes. BMC Bioinformatics. 2006;7(1). https://doi.org/10.1186/1471-2105-7-294.
26. Liu X, Yu X, Zack DJ, Zhu H, Qian J. TiGER: A database for tissue-specific gene expression and regulation. BMC Bioinformatics. 2008;9(1). https://doi.org/10.1186/1471-2105-9-271.
27. Vandenbon A, Nakai K. Modeling tissue-specific structural patterns in human and mouse promoters. Nucleic Acids Res. 2009;38(1):17–25. https://doi.org/10.1093/nar/gkp866.
28. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science. 2015;348(6235):648–660. https://doi.org/10.1126/science.1262110.
29. Pan JB, Hu SC, Shi D, Cai MC, Li YB, Zou Q, et al. PaGenBase: A Pattern Gene Database for the Global and Dynamic Understanding of Gene Function. PLoS ONE. 2013;8(12): e80747. https://doi.org/10.1371/journal.pone.0080747.

30. Lash AE, Tolstoshev CM, Wagner L, Schuler GD, Strausberg RL, Riggins GJ, et al. SAGEmap: A Public Gene Expression Resource. Genome Res. 2000;10(7):1051–60. https://doi.org/10.1101/gr.10.7.1051.
31. Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, et al. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat Biotechnol. 1996;14(13):1675–80. https://doi.org/10.1038/nbt1296-1675.
32. Skrabanek L. TissueInfo: high-throughput identification of tissue expression profiles and specificity. Nucleic Acids Res. 2001;29(21):102e–102. https://doi.org/10.1093/nar/29.21.e102.
33. Nelms BD, Waldron L, Barrera LA, Weflen AW, Goettel JA, Guo G, et al. CellMapper: rapid and accurate inference of gene expression in difficult-to-isolate cell types. Genome Biol. 2016;17(1). https://doi.org/10.1186/s13059-016-1062-5.
34. Amrani KE, Stachelscheid H, Lekschas F, Kurtz A, Andrade-Navarro MA. MGFM: a novel tool for detection of tissue and cell specific marker genes from microarray gene expression data. BMC Genomics. 2015;16(1). https://doi.org/10.1186/s12864-015-1785-9.
35. Duffy Á, Verbanck M, Dobbyn A, Won HH, Rein JL, Forrest IS, et al. Tissue-specific genetic features inform prediction of drug side effects in clinical trials. Sci Adv. 2020;6(37). https://doi.org/10.1126/sciadv.abb6242.
36. Liao BY, Zhang J. Low Rates of Expression Profile Divergence in Highly Expressed Genes and Tissue-Specific Genes During Mammalian Evolution. Mol Biol Evol. 2006;23(6):1119–28. https://doi.org/10.1093/molbev/msj119.
37. Smeds L, Warmuth V, Bolivar P, Uebbing S, Burri R, Suh A, et al. Evolutionary analysis of the female-specific avian W chromosome. Nat Commun. 2015;6(1). https://doi.org/10.1038/ncomms8330.
38. Kryuchkova-Mostacci N, Robinson-Rechavi M. Tissue-Specific Evolution of Protein Coding Genes in Human and Mouse. PLoS ONE. 2015;10(6): e0131673. https://doi.org/10.1371/journal.pone.0131673.
39. Kryuchkova-Mostacci N, Robinson-Rechavi M. Tissue-Specificity of Gene Expression Diverges Slowly between Orthologs, and Rapidly between Paralogs. PLoS Comput Biol. 2016;1(12): e1005274. https://doi.org/10.1371/journal.pcbi.1005274.
40. Assis R, Bachtrog D. Neofunctionalization of young duplicate genes in Drosophila. Proc Natl Acad Sci USA. 2013;110(43):17409–14. https://doi.org/10.1073/pnas.1313759110.
41. Schuster EF, Blanc E, Partridge L, Thornton JM. Correcting for sequence biases in present/absent calls. Genome Biol. 2007;8(6):R125. https://doi.org/10.1186/gb-2007-8-6-r125.
42. Piasecka 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(14):1865–1872. https://doi.org/10.1093/bioinformatics/bts266.
43. Bush SJ, Kover PX, Urrutia AO. Lineage-specific sequence evolution and exon edge conservation partially explain the relationship between evolutionary rate and expression level in A. thaliana. Mol Ecol. 2015;24(12):3093–3106. https://doi.org/10.1111/mec.13221.
44. Fagerberg L, Hallström BM, Oksvold P, Kampf C, Djureinovic D, andMasato Habuka JO, et al. Analysis of the Human Tissue-specific Expression by Genome-wide Integration of Transcriptomics and Antibody-based Proteomics. Mol Cell Proteomics. 2014;13(2):397–406. https://doi.org/10.1074/mcp.m113.035600.
45. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. Science. 2015;347(6220). https://doi.org/10.1126/science.1260419.
46. Noguchi S, Arakawa T, Fukuda S, Furuno M, Hasegawa A, Hori F, et al. FANTOM5 CAGE profiles of human and mouse samples. Sci Data. 2017;4(1). https://doi.org/10.1038/sdata.2017.112.
47. ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 2007;447(7146):799–816. https://doi.org/10.1038/nature05874.
48. Lin S, Lin Y, Nery JR, Urich MA, Breschi A, Davis CA, et al. Comparison of the transcriptional landscapes between human and mouse tissues. Proc Natl Acad Sci USA. 2014;111(48):17224–9. https://doi.org/10.1073/pnas.1413624111.
49. Petryszak R, Burdett T, Fiorelli B, Fonseca NA, Gonzalez-Porta M, Huber EHW, et al. Expression Atlas update—a database of gene and transcript expression from microarray- and sequencing-based functional genomics experiments. Nucleic Acids Res. 2013;42(D1):D926–32. https://doi.org/10.1093/nar/gkt1270.
50. Parkinson H, Kapushesky M, Shojatalab M, Abeygunawardena N, Coulson R, Farne A, et al. ArrayExpress–a public database of microarray experiments and gene expression profiles. Nucleic Acids Res. 2007;35(Database):D747–D750. https://doi.org/10.1093/nar/gkl995.
51. Gremse M, Chang A, Schomburg I, Grote A, Scheer M, Ebeling C, et al. The BRENDA Tissue Ontology (BTO): the first all-integrating ontology of all organisms for enzyme sources. Nucleic Acids Res. 2010;39(Database):D507–D513. https://doi.org/10.1093/nar/gkq968.
52. Gibbons MR, Ross SA, Shanken J. A Test of the Efficiency of a Given Portfolio. Econometrica. 1989;57(5):1121. https://doi.org/10.2307/1913625.
53. Tari L, Baral C, Kim S. Fuzzy c-means clustering with prior biological knowledge. J Biomed Inform. 2009;42(1):74–81. https://doi.org/10.1016/j.jbi.2008.05.009.
54. Ranganathan P, Aggarwal R. Common pitfalls in statistical analysis: Linear regression analysis. Perspect Clin Res. 2017;8(2):100. https://doi.org/10.4103/2229-3485.203040.
55. Sachs L. Applied Statistics, A Handbook of Techniques. 2nd ed. New York: Springer; 1984. https://doi.org/10.1007/978-1-4612-5246-7.
56. Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: database for annotation, visualization, and integrated discovery. Genome Biol. 2003;4(9):R60.
57. Rappaport N, Fishilevich S, Nudel R, Twik M, Belinky F, Plaschkes I, et al. Rational confederation of genes and diseases: NGS interpretation via GeneCards, MalaCards and VarElect. BioMed Eng OnLine. 2017;16(S1). https://doi.org/10.1186/s12938-017-0359-2.
58. Fonseca NA, Marioni J, Brazma A. RNA-Seq Gene Profiling - A Systematic Empirical Comparison. PLoS ONE. 2014;9(9): e107026. https://doi.org/10.1371/journal.pone.0107026.

59. Wu C, Orozco C, Boyer J, Leglise M, Goodale J, Batalov S, et al. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. Genome Biol. 2009;10(11). https://doi.org/10.1186/gb-2009-10-11-r130.
60. Liang P. SAGE Genie: A suite with panoramic view of gene expression. Proc Natl Acad Sci USA. 2002;99(18):11547–8. https://doi.org/10.1073/pnas.192436299.
61. Nikoozad Z, Ghorbanian MT, Rezaei A. Comparison of the liver function and hepatic specific genes expression in cultured mesenchymal stem cells and hepatocytes. Iran J Basic Med Sci. 2014;17(1):27.
62. Sjöstedt E, Fagerberg L, Hallström BM, Häggmark A, Mitsios N, Nilsson P, et al. Defining the Human Brain Proteome Using Transcriptomics and Antibody-Based Profiling with a Focus on the Cerebral Cortex. PLoS ONE. 2015;10(6):e0130028. https://doi.org/10.1371/journal.pone.0130028.
63. Myers AJ, Gibbs JR, Webster JA, Rohrer K, Zhao A, Marlowe L, et al. A survey of genetic human cortical gene expression. Nat Genet. 2007;39(12):1494–9. https://doi.org/10.1038/ng.2007.16.
64. Naumova OY, Lee M, Rychkov SY, Vlasova NV, Grigorenko EL. Gene Expression in the Human Brain: The Current State of the Study of Specificity and Spatiotemporal Dynamics. Child Dev. 2012;84(1):76–88. https://doi.org/10.1111/cdev.12014.
65. Zhang Y, Chen K, Sloan SA, Bennett ML, Scholze AR, O'Keeffe S, et al. An RNA-Sequencing Transcriptome and Splicing Database of Glia, Neurons, and Vascular Cells of the Cerebral Cortex. J Neurosci. 2014;34(36):11929–47. https://doi.org/10.1523/jneurosci.1860-14.2014.
66. Göring HHH. Tissue specificity of genetic regulation of gene expression. Nat Genet. 2012;44(10):1077–8. https://doi.org/10.1038/ng.2420.
67. Blighe K. Cancer mutations and their tissue-specific nature. J Cancer Sci Ther. 2014;6:009–11.
68. Maris JM, Knudson AG. Revisiting tissue specificity of germline cancer predisposing mutations. Nat Rev Cancer. 2015;15(2):65–6. https://doi.org/10.1038/nrc3894.
69. Ko Y, Ament SA, Eddy JA, Caballero J, Earls JC, Hood L, et al. Cell type-specific genes show striking and distinct patterns of spatial expression in the mouse brain. Proc Natl Acad Sci USA. 2013;110(8):3095–100. https://doi.org/10.1073/pnas.1222897110.
70. Willard-Mack CL. Normal Structure, Function, and Histology of Lymph Nodes. Toxicol Pathol. 2006;34(5):409–24. https://doi.org/10.1080/01926230600867727.
71. Waters S. The Female Reproductive System. New York: The Rosen Publishing Group; 2007.
72. Saladin KS, Miller L. Anatomy & physiology. New York: WCB/McGraw-Hill; 1998.
73. Vela CIB, Padilla FJB. Determination of ammonia concentrations in cirrhosis patients-still confusing after all these years? Ann Hepatol. 2011;10:S60–5. https://doi.org/10.1016/s1665-2681(19)31609-6.
74. Sonawane AR, Platig J, Fagny M, Chen CY, Paulson JN, Lopes-Ramos CM, et al. Understanding Tissue-Specific Gene Regulation. Cell Rep. 2017;21(4):1077–88. https://doi.org/10.1016/j.celrep.2017.10.001.
75. Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. A census of human transcription factors: function, expression and evolution. Nat Rev Genet. 2009;10(4):252–63. https://doi.org/10.1038/nrg2538.

Publisher's Note