A Computational Method for Classifying Different Human Tissues with Quantitatively Tissue-Specific Expressed Genes

JiaRui Li [1,†], **Lei Chen** [2,3,†] , **Yu-Hang Zhang** [4], **XiangYin Kong** [4,*], **Tao Huang** [4,*] **and Yu-Dong Cai** [1,*]

1   School of Life Sciences, Shanghai University, Shanghai 200444, China; jiaruili@shu.edu.cn
2   College of Information Engineering, Shanghai Maritime University, Shanghai 201306, China; chen_lei1@163.com
3   Shanghai Key Laboratory of PMMP, East China Normal University, Shanghai 200241, China
4   Institute of Health Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China; zhangyh825@163.com
*   Correspondence: xykong@sibs.ac.cn (X.K.); tohuangtao@126.com (T.H.); cai_yud@126.com (Y.-D.C.); Tel.: +86-021-6613-6132 (Y.-D.C.)
†   These authors contributed to work equally.

Abstract: Tissue-specific gene expression has long been recognized as a crucial key for understanding tissue development and function. Efforts have been made in the past decade to identify tissue-specific expression profiles, such as the Human Proteome Atlas and FANTOM5. However, these studies mainly focused on "qualitatively tissue-specific expressed genes" which are highly enriched in one or a group of tissues but paid less attention to "quantitatively tissue-specific expressed genes", which are expressed in all or most tissues but with differential expression levels. In this study, we applied machine learning algorithms to build a computational method for identifying "quantitatively tissue-specific expressed genes" capable of distinguishing 25 human tissues from their expression patterns. Our results uncovered the expression of 432 genes as optimal features for tissue classification, which were obtained with a Matthews Correlation Coefficient (*MCC*) of more than 0.99 yielded by a support vector machine (SVM). This constructed model was superior to the SVM model using tissue enriched genes and yielded *MCC* of 0.985 on an independent test dataset, indicating its good generalization ability. These 432 genes were proven to be widely expressed in multiple tissues and a literature review of the top 23 genes found that most of them support their discriminating powers. As a complement to previous studies, our discovery of these quantitatively tissue-specific genes provides insights into the detailed understanding of tissue development and function.

Keywords: tissue-specific expressed genes; transcriptome; tissue classification; support vector machine; feature selection

1. Introduction

A biological tissue is an ensemble of similar cells residing in the same location and performing specific biological functions in multicellular organisms. As the bridge between single cells and functional organs, tissues are elementary units with both phenotypical and functional contributions to biological identity [1]. All biological functions are regulated and manipulated directly or indirectly by proteins, which can be further attributed to gene expression patterns measured by messenger RNA (mRNA) expression [2]. Therefore, different tissues and cell types could have their own unique

expression patterns and a full picture of how genes are expressed in different tissues will help to unveil the molecular mechanisms involved in tissue development and function.

Ten years ago, two milestones for identifying tissue-specific gene expression were conducted and completed right after the human genome project, which built tissue-specific gene expression profiles at the protein and RNA levels [3,4]. At the protein level, the protein distribution in human tissues was explored using 718 antibodies corresponding to 650 human protein-coding genes in the Human Protein Atlas [3]. While at the RNA level, the FANTOM consortium initiated the creation of a gene atlas by integrating mouse and human expression data from multiple tissues using microarrays [4] and built the widely used BioGPS portal, which has expression data from numerous resources [5]. These projects have been extended and updated by incorporating more genes and taking advantage of advanced technologies, such as next-generation sequencing. For example, the number of genes included in the Human Protein Atlas project has been increased to 10,118, which is more than half of the protein-coding genes in humans [6]. Meanwhile, the application of RNA-Seq technologies has created an RNA-Seq Atlas in different human tissues, providing a more comprehensive and unbiased view of gene expression compared with the initial efforts using microarrays [7]. A recent study presented a map of the human tissue proteome by integrating multiple-omics approaches, including RNA-Seq and tissue microarray-based immunohistochemistry, which detected more than 90% of putative protein-coding genes and found that 2355 genes are significantly enriched in a single tissue, 3478 are enhanced in a single tissue and 1109 are enriched in a group of tissues [8]. These tissue-specific genes not only help understand human biology but can also be applied in medical research, such as pharmaceutical drug development and biomarker discovery in the field of translational medicine. However, these studies have mainly focused on "qualitatively tissue-specific expressed genes", which are enriched in a single or subgroup of tissues. The genes expressed in all or almost all tissues could also have divergent expression patterns among different cell types, which we termed as "quantitatively tissue-specific expressed genes". Although these quantitatively tissue-specific expressed genes have less expression enrichment compared with qualitatively expressed genes, they might also play important roles in tissue function and development.

In this study, we took advantage of recently published transcriptomic data in multiple tissues from the Genotype-Tissue Expression (GTEx) project [9] and present a new computational method that integrates machine learning algorithms to identify genes that are widely expressed in the human body but with different expression signatures across 25 human tissues and are capable of distinguishing different tissue types. According to our results, the 25 tissues have been taken into account and subtyped by 432 key genes using a prediction engine based on a support vector machine (SVM) [10,11] with an Matthews Correlation Coefficient (*MCC*) value more than 0.99, revealing the detailed expression characteristics of different tissue subtypes. The constructed classification model also had good generalization ability because it yielded *MCC* of 0.985 on an independent test dataset. In addition, the superiority of this model was proved by comparing it to the SVM model using tissue enriched genes. A detailed analysis was also performed on the 23 most important genes among 432 key genes. As a complement to previous studies, our results demonstrate the ability to classify tissues through a series of quantitatively tissue-specific expressed genes, suggesting that these genes could also play important roles in tissue development and function.

2. Materials and Methods

2.1. Dataset

The expression profiles from different tissue samples obtained by RNA-Seq were downloaded from GTEx V6p (http://gtexportal.org/home/datasets) [9]. Tissues with sample sizes smaller than 80 were excluded, resulting in a total of 8436 samples from 25 tissues. These samples comprised a training dataset. For an easy description, each tissue was denoted as T_i (*I* = 1, 2, . . . , 25). For the computational analysis, we extracted expression levels of 18,365 genes that

are expressed (i.e., expression level was not zero) in at least one of 8436 samples from the file "GTEx_Analysis_v6p_RNA-seq_RNA-SeQCv1.1.8_gene_rpkm.gct"; that is, each sample was represented by 18,365 features.

Besides, an independent test dataset was constructed using the new samples added in GTEx v7 after GTEx v6p. This dataset included 3367 samples from the same 25 tissues. These 25 tissues and their sample sizes in training and independent test datasets are listed in Table 1.

Table 1. The 25 tissue samples.

Tag	Tissue	Number of Samples		Tag	Tissue	Number of Samples	
		Training Dataset	Test Dataset			Training Dataset	Test Dataset
T_1	Adipose tissue	577	237	T_2	Adrenal gland	145	50
T_3	Blood	511	54	T_4	Blood vessel	689	242
T_5	Brain	1259	455	T_6	Breast	214	84
T_7	Colon	345	169	T_8	Esophagus	686	348
T_9	Heart	412	201	T_{10}	Liver	119	58
T_{11}	Lung	320	123	T_{12}	Muscle	430	155
T_{13}	Nerve	304	122	T_{14}	Ovary	97	39
T_{15}	Pancreas	171	82	T_{16}	Pituitary	103	82
T_{17}	Prostate	106	48	T_{18}	Skin	890	342
T_{19}	Small intestine	88	52	T_{20}	Spleen	104	60
T_{21}	Stomach	192	75	T_{22}	Testis	172	91
T_{23}	Thyroid	323	139	T_{24}	Uterus	83	32
T_{25}	Vagina	96	27	Total	-	8436	3367

2.2. Feature Ranking and Selection

The purpose of this study is to extract a group of genes that can contribute to tissue classification based on gene expression levels. To do that, the minimum redundancy maximum relevance (mRMR) method [12] proposed by Peng et al. was employed to analyze all the features and yield a feature list, named the mRMR feature list. This feature selection method has been applied to tackle various biological problems [13–25]. In the method, the discriminated power of a feature *f* is reflected by the relevance between it and a target class *c*, which is measured from their mutual information (*MI*). The *MI* can be evaluated according to the following equation (*x* and *y* represent two variables):

$$\iint p(x,y) log \frac{p(x,y)}{p(x)p(y)} dxdy \tag{1}$$

where $p(x)$ and $p(y)$ are the marginal probabilistic density of *x* and *y* and $p(x, y)$ is their joint probabilistic density. Additionally, the redundancy between two features f_1 and f_2 is also evaluated by their *MI*. To produce the mRMR feature list, the mRMR method repeatedly selects a non-selected feature that has a maximum relevance to the target and a minimum redundancy to the already selected features. The mRMR feature list ranks all the features according to their selection order. For the formulation, this list was denoted as follows:

$$F = [f_1, f_2, \ldots f_N] \tag{2}$$

where *N* is the total number of features (*N* = 18,365 in this study). It is clear that the top features in this list are more important for identifying tissue types.

The mRMR method only provides a feature list, in which important features receive high ranks. However, we still do not know which features should be selected for classification. In view of this, the incremental feature selection (IFS) method was adopted in this study to select a group of features that can effectively classify the human tissues. In this method, a series of feature sets, denoted as F_1, $F_2, \ldots, F_i, \ldots, F_N$, were constructed according to the mRMR feature list *F*, where the subscript *i* in F_i represents the top *i* features in *F* comprising the feature set F_i, that is,

$$F_i = [f_1, f_2, \ldots f_i] \tag{3}$$

Then, a classification model would be built on each feature set trained by a machine learning algorithm (e.g., SVM). Accordingly, the classification model yielding the best performance can be found and termed the optimal classification model. Features used in this optimal classification model comprise the optimal feature set. In this study, there were 18,365 features in total, meaning we could construct 18,365 classification models. However, due to our limited computational power, this would be time-consuming. Thus, we only examined feature sets that contained 4–500 features to construct the classification models and selected the model yielding the highest Matthews Correlation Coefficient, a powerful measurement to evaluate the performance of different models. For convenience, the obtained model was called optimal classification model.

2.3. Classification Algorithm

In the IFS method, a machine learning algorithm was used to build a classification model based on each feature set. Here, we selected one of the most classic machine learning algorithms, SVM [10,11]. The principle of SVM is to find a balance between the learning error and the minimum statistical risk. To date, several types of SVMs have been proposed to address different kinds of problems. In this study, we adopted an SVM that is trained by sequential minimum optimization (SMO) [26], which always breaks the quadratic programming (QP) problem into a series of the smallest possible sub-QP problems. Then, these sub-QP problems are solved analytically, thereby avoiding matrix storage and using numeric QP optimization steps. To quickly implement this type of SVM, a tool named "SMO" was employed in Weka (https://www.cs.waikato.ac.nz/ml/weka/) [27] and it was executed using its default settings, where the kernel was set as polynomial function and the tolerance parameter was set to 0.001.

2.4. Measurements

The prediction abilities of the constructed classification models were evaluated by a 10-fold cross-validation (10-CV) [16,28–30], which yields similar results to the stricter test called the Jackknife cross-validation (J-CV) [31,32] but saves significant computational resources.

In this study, 25 tissues were considered. Thus, based on the predicted results derived from the SVM on different feature sets and evaluated by 10-CV, we can compute the prediction accuracy for j-th tissue T_j, which was defined as follows:

$$ACC_j = \frac{n_j}{N_j}, \ j = 1, 2, \ldots, 25 \tag{4}$$

where n_j represents the number of correctly predicted samples in T_j and N_j represents the total number of samples in T_j. We can further compute the overall accuracy (*TACC*) with the following formula:

$$TACC = \frac{\sum_{j=1}^{25} n_j}{\sum_{j=1}^{25} N_j} \tag{5}$$

We can see from Table 1 that some tissues were large (e.g., brain), while some tissues had much fewer samples (e.g., uterus). The overall accuracy may strongly rely on the prediction accuracy of the tissues with large sizes. Thus, it is not proper to evaluate the performance of models using *TACC*. In binary classification, *MCC* [33] is deemed as a balanced measurement even if the class sizes are of great differences. Later, the multi-class version—called *MCC* in multiclass—was developed by Gorodkin [34], which inherits the merits of original *MCC*. Here, we give a brief description of the *MCC* in multiclass. A more detailed description can be found in Gorodkin's study [34].

Given a classification problem involving n samples, denoted by $s_1, s_2, \ldots, s_n$ and N classes, represented by 1, 2, $\ldots$, N. The true classes of all the samples can be formulated by the matrix $Y = (y_{ij})_{n \times N}$, where $y_{ij} = 1$ if s_i belongs to class j; otherwise, it is set to 0. For the predicted classes

of all the samples, the classes can be used to a define another matrix, $X = (x_{ij})_{n \times N}$, which can be defined as follows: $x_{ij} = 1$, if s_i is predicted to be in class j, otherwise, $x_{ij} = 0$. Then, the *MCC* in the multiclass is defined as follows:

$$MCC = \frac{cov(X, Y)}{\sqrt{cov(X, X)cov(Y, Y)}} \tag{6}$$

where *cov(X,Y)* is the covariance function of *X* and *Y*, which can be computed with the following formula:

$$cov(X, Y) = \frac{1}{N} \sum_{k=1}^{N} cov(x_k, y_k) = \frac{1}{N} \sum_{k=1}^{N} \sum_{i=1}^{n} (x_{ik} - \overline{x_k})(y_{ik} - \overline{y_k}) \tag{7}$$

where x_k and y_k denote the *k*-th column of *X* and *Y*, respectively and $\overline{x_k}$ and $\overline{y_k}$ denote the mean values of the numbers in x_k and y_k, respectively.

Like the original *MCC* value proposed by Matthews [33], the range of the *MCC* in the multiclass is between -1 and 1, where the higher the *MCC* value obtained, the better a classifier is (1 means the given classifier yields a perfect classification, 0 indicates a classification is no better than a random prediction and -1 represents a total misclassification). In this study, we used the *MCC* values in the multiclass as the key measurements and only called *MCC* in the following text for convenience.

3. Results

3.1. Results of Feature Ranking

To avoid the unbalanced data sizes among the different tissues, we collected the transcriptome data from 8436 samples originating from 25 tissues in the GTEx project [9]; tissues with sample sizes smaller than 80 were excluded from this study. To perform a comprehensive analysis, all 18,365 genes expressed in the 8436 samples, regardless of tissue specificity or expression, were used to construct the features. All the features were rigorously analyzed with the widely used mRMR method, inducing the mRMR feature list (Supplementary Material S1).

3.2. Results of Feature Selection

The obtained mRMR feature list was used with the IFS method to discover the optimal feature set for the SVM. However, because of our limited computational power, we only tried feature sets F_4, F_5, ..., F_{500}. For each feature set, an SVM-based classification model was built, in which each sample was represented by the features in the set. Then, 10-CV was adopted to evaluate the performance of each classification model; the predicted results included the prediction accuracies of the 25 tissues, *TACC* and *MCC* (Supplementary Material S2). To clearly show the relationship between the number of used features and the performance of the corresponding classification model, the IFS curve was plotted and is shown in Figure 1 with the feature number as the X-axis and the *MCC* value as the Y-axis. The highest *MCC* value (0.994) reached almost 1.00 and was obtained when the top 432 features in the mRMR feature list were used, suggesting that these 432 features could be optimal features for the SVM for distinguishing the 25 tissues. Accordingly, the optimal classification model using 432 optimal features and SVM can be built. The detailed performance of this model, including accuracy on each tissue and *TACC*, is shown in Figure 2, from which we can see that 17 of 25 tissues received a perfect classification ($ACC_j = 1$) and all accuracies are higher than 0.930, suggesting that the performance of the optimal classification model is quite stable on different tissues.

To examine whether these 432 genes are widely expressed or tissue-specific, we searched their annotations in the Human Protein Atlas [8]. Among the 432 genes, fourteen genes were annotated as tissue-enriched, tissue-enhanced or group-enriched, which suggests that these fourteen genes might be tissue-specific or group-specific (Supplementary Material S3). However, further examination revealed

that only one gene, *ASAP2*, is annotated as tissue-enriched (in testis) in the GTEx data and that this gene shows no enriched or enhanced expression in the FANTOM5 database [35], another expression atlas for RNA levels. Both the GTEx and FANTOM5 annotations were obtained through the Human Protein Atlas database to keep the criteria consistent. Therefore, these 432 genes showed the ability to distinguish different tissues that are widely expressed across multiple tissues.
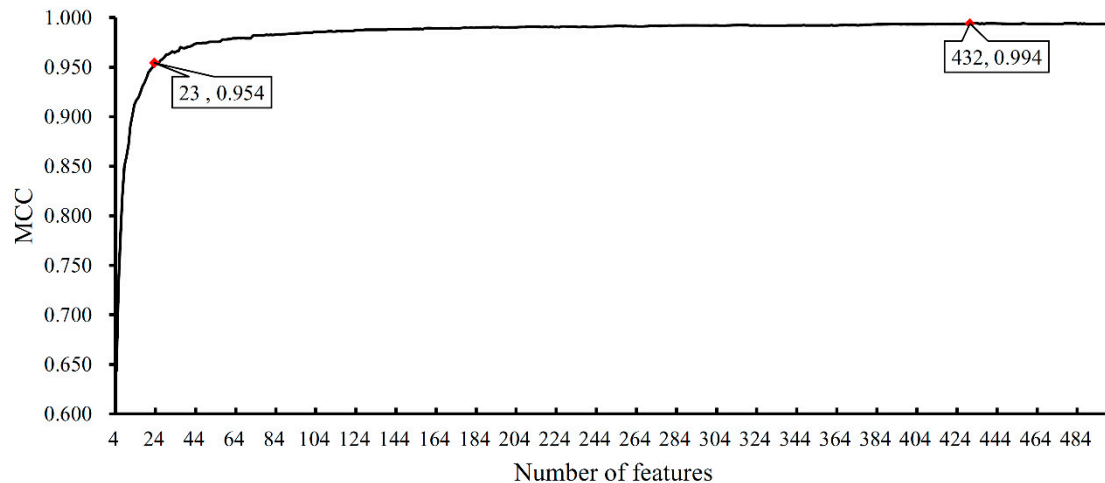

Figure 1. The incremental feature selection (IFS) curve illustrating the performance of the classification models using different numbers of features. Red diamonds represent the performance when the top 23 genes and 432 features were used for building the classification models.
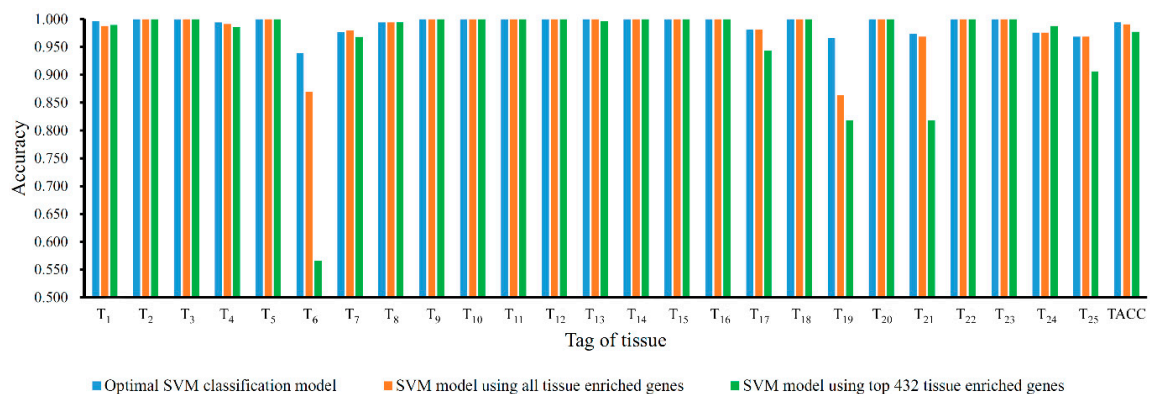
Figure 2. The performance of the optimal support vector machine (SVM) classification model, SVM model using all tissue enriched genes and SVM model using top 432 tissue enriched genes, including accuracy on each tissue and overall accuracy. The optimal SVM classification model gave better performance.

3.3. Comparison of SVM Model with Tissue Enriched Genes

As mentioned in Section 3.2, an SVM classification model was built to classify samples into 25 tissues, which yielded the *MCC* of 0.994. To indicate its effectiveness, we mapped the tissue enriched proteins retrieved from Human Protein Atlas onto the filtered GTEx expression dataset which excluded the non-expressed genes, resulting in 1981 genes (see Supplementary Material S4). The expression levels of these genes were used to represent each sample in the training dataset. Then, the SVM was executed on this data with its performance evaluated by 10-CV. The predicted results were counted as *MCC* of 0.990, which was lower than that obtained by the optimal SVM classification model. For detailed comparison, the detailed performance of these two models, including accuracy on each tissue and *TACC*, is illustrated in Figure 2. It can be observed that the optimal SVM classification

model gave better or equal performance on almost all tissues (expect tissue T_7, "Colon") than SVM model using tissue enriched genes.

Besides, considering that the above two models using different number of genes, to give a fairer comparison, we also evaluated the importance of 1981 tissue enriched genes via mRMR method and extracted the top 432 genes to construct an SVM classification model and evaluate it on the training dataset, yielding *MCC* of 0.976, which was much lower than that obtained by the optimal SVM classification model. The detailed performance on each tissue and overall accuracy is shown in Figure 2, from which we can see that the optimal SVM classification model gave higher or equal accuracies on almost all tissues expect one tissue (T_{24}, "Uterus").

All these suggested that important genes for classification of samples into different tissues can be extracted via advanced computational methods used in this study and they can be adopted to build a better classification model.

3.4. Performance of the Optimal SVM Classification Model on Test Dataset

We constructed an optimal SVM classification model in Section 3.2 via mRMR and IFS methods. To test its generalization, we performed this model on the independent test dataset. The obtained *TACC* and *MCC* were 0.986 and 0.985, respectively, indicating that the prediction ability of the constructed model was quite strong. We also calculated the *MCC* of tissue enriched genes on the independent test dataset and it was 0.972, smaller than the *MCC* (0.985) obtained by the optimal SVM classification model.

3.5. Comparison of SVM Model with t-Test Genes

Beside the machine learning algorithms used in present study, there were other ways to identify the quantitatively tissue-specific expressed genes. Another more straightforward way was to identify the significantly highly expressed genes in each tissue by performing t test between one tissue and all the other tissues and then combine the highly expressed genes in each tissue to get the final quantitatively tissue-specific expressed genes. To make the t test result comparable with our result, we selected the top 19 significantly highly expressed genes from 25 tissues and obtained 422 unique *t*-test genes (see Supplementary Material S5), which had similar number of genes with the 432 optimal genes. We evaluated the performance of the 422 *t*-test genes on independent test dataset and its *MCC* was 0.984, slightly smaller than the *MCC* obtained by 432 optimal genes, 0.985. It was difficult to tell whether the increase of *MCC* from 0.984 to 0.985 was statistically significant, especially when the *MCC* was already so high and there was little space for improvement. But we strongly believe that the quantitative gene expression signatures identified either by *t*-test based method or mRMR and IFS based method, are better than traditional tissue specific gene lists that only consider the expression or not rather than the expression level differences. The results on the high-quality dataset GTEx had supported it. Although the GTEx project has ended, the Enhancing GTEx (eGTEx) project [36] carries on. We will work closely with the eGTEx Consortium to test the constructed model on larger new samples and update the model correspondingly.

4. Discussion

As we have analyzed above, RNA-Seq has been reported to be an effective classification tool for identifying cell types. Based on the transcriptome datasets from different human tissues presented in a recent study, we developed a new computational method and successfully identified 432 quantitatively tissue-specific expressed genes capable of classifying twenty-five human tissues with high accuracy ($MCC > 0.99$). To further examine the reliability of our results, we selected the top 23 genes in the mRMR feature list for detailed analyses (Table 2). The performance of the classification model built on these genes displayed an *MCC* value > 0.95 (Figure 1, Supplementary Material S2), suggesting the significant role of these genes among the 432 genes in the classification. As shown in Table 2, we searched the 23 genes in Human Protein Atlas [8] and Expression Atlas [37]. Based on Human

Protein Atlas, 16 genes were "Expressed in all"; six genes were "Mixed"; only one gene was "Tissue enhanced (thyroid gland)". Based on Expression Atlas, all genes were expressed in "Multiple tissues". Both databases suggested these genes were expressed in multiple tissues. But our method can find their expression level difference rather than whether they were expressed. For clearly displaying the expression level of these 23 genes across 25 tissues, we gave a box plot for each gene, which is illustrated in Supplementary Figure S1. It can be observed that for almost all genes, their expression levels on different tissues are quite different, indicating that they can be important biomarkers.

The first gene, *ARAF*, is a potential proto-oncogene that can be clustered into the *RAF* subfamily of Ser/Thr protein kinases and contributes to the regulation of cell proliferation and tissue development [38]. For its tissue-specific expression pattern, this gene has been reported to be steadily expressed in multiple tissues, including most of tissue/cell subtypes incorporated in this study, suggesting a wide expression signature [39]. More importantly, it has also been confirmed to have a unique expression pattern in the skin; thus, though this gene is expressed widely in multiple tissues, the expression levels could be different from one another [39].

Table 2. The top 23 genes selected for further investigation via a literature review.

Rank	Gene	Description	The Human Protein Atlas [8]	Expression Atlas of EMBL-EBI [37]
1	*ARAF*	A-Raf Proto-Oncogene, Serine/Threonine Kinase	Expressed in all	Multiple tissues
2	*ITGA3*	Integrin Subunit Alpha 3	Mixed	Multiple tissues
3	*SLAIN2*	SLAIN Motif Family Member 2	Expressed in all	Multiple tissues
4	*ZNF532*	Zinc Finger Protein 532	Mixed	Multiple tissues
5	*PPIC*	Peptidylprolyl Isomerase C	Mixed	Multiple tissues
6	*KDELR1*	KDEL Endoplasmic Reticulum Protein Retention Receptor 1	Expressed in all	Multiple tissues
7	*NBL1*	Neuroblastoma 1, DAN Family BMP Antagonist	Expressed in all	Multiple tissues
8	*PLP2*	Proteolipid Protein 2	Expressed in all	Multiple tissues
9	*STAT6*	Signal Transducer and Activator of Transcription 6	Expressed in all	Multiple tissues
10	*ARHGAP23*	Rho GTPase Activating Protein 23	Mixed	Multiple tissues
11	*LRIG3*	Leucine Rich Repeats And Immunoglobulin Like Domains 3	Tissue enhanced (thyroid gland)	Multiple tissues
12	*MANBAL*	Mannosidase Beta Like	Expressed in all	Multiple tissues
13	*PTPRA*	Protein Tyrosine Phosphatase, Receptor Type A	Expressed in all	Multiple tissues
14	*YAP1*	Yes Associated Protein 1	Mixed	Multiple tissues
15	*CLIC1*	Chloride Intracellular Channel 1	Expressed in all	Multiple tissues
16	*TMEM109*	Transmembrane Protein 109	Expressed in all	Multiple tissues
17	*MOCS2*	Molybdenum Cofactor Synthesis 2	Expressed in all	Multiple tissues
18	*PTPRF*	Protein Tyrosine Phosphatase, Receptor Type F	Mixed	Multiple tissues
19	*MYO1C*	Myosin IC	Expressed in all	Multiple tissues
20	*FAM127B*	Family with Sequence Similarity 127 Member B	Expressed in all	Multiple tissues
21	*TRIP10*	Thyroid Hormone Receptor Interactor 10	Expressed in all	Multiple tissues
22	*SERPING1*	Serpin Family G Member 1	Expressed in all	Multiple tissues
23	*TOM1L2*	Target of Myb1 Like 2 Membrane Trafficking Protein	Expressed in all	Multiple tissues

Another gene named *ITGA3* has been widely reported to function as a cell surface adhesion molecule and is involved in the malignant metastasis of certain tumor subtypes [40,41]. Therefore, it is expected that this gene has low expression in the blood, as many blood cells freely float [42], suggesting that this gene may have different expression patterns across tissues corresponding to their different cell adhesion requirements. Similarly, *SLAIN2*, a microtubule dynamics-associated regulator, has also been reported to be down-regulated in blood cells compared with other tissue subtypes in our candidates, indicating the distinctive roles of *SLAIN2* as well as microtubule dynamics in different tissues [43]. *ZNF532*, a nucleic acid binding-associated gene, has been confirmed to be involved in transcriptional regulation [44,45]. Based on recent publications, this gene has also been reported to be down-regulated in liver and whole blood cells compared with other candidate tissues, implying that this gene may be a potential marker for the identification of hepatocytes and blood cells [46].

PPIC encodes functional peptidylprolyl isomerase C (cyclophilin C), which has been confirmed to bind to the immunosuppressant cyclosporin A, implying that this gene is an immune-associated regulator [47,48]. This gene has been confirmed to have quite high expression levels in the tibial

nerve but lower levels in the cortex, suggesting its differential expression levels across various cell types and indicating potential roles in tissue classification [47]. In contrast to *PPIC*, another nerve associated-gene, *NBL1*, has been confirmed to be highly expressed in the central nervous system (brain) but lowly expressed in the peripheral nervous system (such as the spinal cord) [49,50].

KDELR1, a potential endoplasmic reticulum (ER)-associated gene, has no direct evidence reporting any unique expression patterns. However, this gene has been reported to affect ion transmembrane transporter activity in the nervous system and blood cells [51,52], so it is likely that this gene may have specific expression patterns corresponding to divergent ion transmembrane transporter activity requirements in different tissues.

PLP2, which has not been reported to display unique expression patterns, has been found to increase neuronal apoptosis when down-regulated and promotes cell proliferation in leukemia when up-regulated, indicating that different expression levels are required in tissues based on their needs for cell apoptosis and proliferation [53,54].

STAT6, a transcription factor in the STAT family, has been confirmed to contribute to the proliferation and differentiation of T helper 2 cells, indicating its specific role in the immune system and certain cell types [55–57].

ARHGAP23 is an effective component of the Rho GTPase family and has been widely reported to be involved in transmembrane receptor signal transduction [58]. With regards to its detailed expression, this gene has been confirmed to be down-regulated in blood and muscles, revealing a specific expression pattern [59]. Similarly, *LRIG3* also has quite low expression levels in the blood [60]. Moreover, *LRIG3* displays specific expression patterns in the nervous system as it regulates the normal functioning of the inner ear, implying that this gene may also contribute to the identification of nerve tissues [61,62].

As a gene encoding a functional component of the Hippo signaling pathway involved in development, growth, repair and homeostasis, *YAP1* has low expression in mature blood cells [63,64]. This gene, as well as another three genes *MANBAL*, *PTRPA* and *CLIC1* has also been reported to be differentially expressed across different human tissues [9].

Another membrane-associated gene, *TMEM109*, has also been predicted to be distinctively expressed in different tissues. Generally, this gene has been widely reported to mediate cellular responses to DNA damage, such as ultraviolet C-induced cell death [65,66]. Not present in the bone marrow or thymus, this gene may also participate in the identification of certain cell types [66].

MOCS2, a gene encoding the eukaryotic molybdoenzyme, has been widely reported to contribute to molybdenum cofactor synthesis [67]. For its expression pattern, this gene has been confirmed to be predominantly expressed in heart and skeletal muscle, thus contributing to the identification of muscle and heart cells [67–69].

PTPRF, a member of the protein tyrosine phosphatase (PTP) family, has been reported to participate in various core survival-associated biological processes, such as cell growth, differentiation and mitosis [70,71], which suggests that differential expression patterns could be found in different tissues due to divergent levels of cell growth and differentiation. For example, the expression of this gene has been reported to be down-regulated in tissues with low proliferation potential, such as the heart and brain [72,73].

Encoding a common protein in muscle, *MYO1C* could be up-regulated in tissues that contain smooth muscle, skeletal muscle or heart muscle, which would distinguish these tissues from other tissues such as blood and cartilage [74,75].

The *TRIP10* gene has been confirmed to have quite a low expression level in the blood and nervous system, implying its potential in tissue classification [76,77]. The abnormal expression of this gene usually indicates specific pathological processes such as cancer and leukemia [76].

The last two genes *SERPING1* and *TOM1L2* both have been found to be expressed in all tissues [8]. *SERPING1* encodes the plasma protease C1 inhibitor involved in regulating important physiological pathways, including complement activation, blood coagulation, fibrinolysis and the

generation of kinins [78], while *TOM1L2* may regulate growth factor-induced mitogenic signaling [79]. These discovered functional roles suggest potential requirements for differential expression patterns across different tissues.

We have discussed 22 of the 23 genes (*FAM127B* has little information in terms of its function) and shown their potential for tissue classification based on previous studies. Some genes, such as *ARAF*, *PTPRF*, *ITGA3* and *SLAIN2*, are involved in crucial cell processes, including cell proliferation, cell adhesion, cell invasion and the regulation of microtubule dynamics, which have diverse roles in different tissues, indicating that these genes might have differential expression patterns across different tissues. Other genes were confirmed to have specific functions in particular tissues or cell types, suggesting their specific roles in these tissues or cell types. By comparing the GO annotations between the 432 quantitatively-specific expressed genes identified in this study and the 2355 tissue-enriched genes annotated previously by Uhlen et al. [3] through the online tool DAVID [80], we found that these two sets of genes showed diverse clustering on biological processes. The former ones are enriched in translational, transcriptional and posttranscriptional regulation processes, while the latter ones are enriched in those biological processes specific to cell-types or tissues, such as spermatogenesis, muscle filament sliding, cell differentiation and so forth (see Supplementary Material S6). This finding indicates those quantitatively tissue-specific genes possibly contribute to tissue-specific features through transcriptional, translational, posttranscriptional and posttranslational regulations.

5. Conclusions

In summary, this study identified quantitatively tissue-specific expressed genes with discriminating power for classifying different tissue types based on their expression patterns via machine-learning algorithms. As a complement to previous studies that have identified genes enriched in a single or a few tissues, our findings of genes that are commonly but also differentially expressed in multiple tissues will provide insights into a detailed understanding of tissue development and function. In addition, we provided tissue samples represented by expression profiles of 432 optimal genes in Supplementary Material S7, in which the instructions of how to use this file to build the classification model and further make predictions were also given.

Author Contributions: X.K., T.H. and Y.-D.C. conceived and designed the experiments; L.C. performed the experiments; J.L. analyzed the data; J.L. and Y.-H.Z. contributed reagents/materials/analysis tools; J.L. and L.C. wrote the paper.

References

1. Singh, S.R. Stem cell niche in tissue homeostasis, aging and cancer. *Curr. Med. Chem.* **2012**, *19*, 5965–5974. [CrossRef] [PubMed]

2. Lipscombe, D.; Andrade, A. Calcium channel cavα_1 splice isoforms—Tissue specificity and drug action. *Curr. Mol. Pharmacol.* **2015**, *8*, 22–31. [CrossRef] [PubMed]

3. Uhlen, M.; Bjorling, E.; Agaton, C.; Szigyarto, C.A.; Amini, B.; Andersen, E.; Andersson, A.C.; Angelidou, P.; Asplund, A.; Asplund, C.; et al. A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol. Cell. Proteom. MCP* **2005**, *4*, 1920–1932. [CrossRef] [PubMed]

4. Su, A.I.; Wiltshire, T.; Batalov, S.; Lapp, H.; Ching, K.A.; Block, D.; Zhang, J.; Soden, R.; Hayakawa, M.; Kreiman, G.; et al. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 6062–6067. [CrossRef] [PubMed]

5. Wu, C.; Orozco, C.; Boyer, J.; Leglise, M.; Goodale, J.; Batalov, S.; Hodge, C.L.; Haase, J.; Janes, J.; Huss, J.W., III; et al. BioGPS: An extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol.* **2009**, *10*, R130. [CrossRef] [PubMed]

6. Uhlen, M.; Oksvold, P.; Fagerberg, L.; Lundberg, E.; Jonasson, K.; Forsberg, M.; Zwahlen, M.; Kampf, C.; Wester, K.; Hober, S.; et al. Towards a knowledge-based human protein atlas. *Nat. Biotechnol.* **2010**, *28*, 1248–1250. [CrossRef] [PubMed]

7. Krupp, M.; Marquardt, J.U.; Sahin, U.; Galle, P.R.; Castle, J.; Teufel, A. RNA-seq atlas—A reference database for gene expression profiling in normal tissue by next-generation sequencing. *Bioinformatics* **2012**, *28*, 1184–1185. [CrossRef] [PubMed]

8. Uhlen, M.; Fagerberg, L.; Hallstrom, B.M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, A.; Kampf, C.; Sjostedt, E.; Asplund, A.; et al. Tissue-based map of the human proteome. *Science* **2015**, *347*, 1260419. [CrossRef] [PubMed]

9. The GTEx Consortium; Human genomics. The genotype-tissue expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **2015**, *348*, 648–660. [CrossRef] [PubMed]

10. Meyer, D.; Leisch, F.; Hornik, K. The support vector machine under test. *Neurocomputing* **2003**, *55*, 169–186. [CrossRef]

11. Corinna, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]

12. Peng, H.; Long, F.; Ding, C. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226–1238. [CrossRef] [PubMed]

13. Li, B.Q.; Cai, Y.D.; Feng, K.Y.; Zhao, G.J. Prediction of protein cleavage site with feature selection by random forest. *PLoS ONE* **2012**, *7*, e45854. [CrossRef] [PubMed]

14. Chen, L.; Zhang, Y.-H.; Lu, G.; Huang, T.; Cai, Y.-D. Analysis of cancer-related lncRNAs using gene ontology and kegg pathways. *Artif. Intell. Med.* **2017**, *76*, 27–36. [CrossRef] [PubMed]

15. Cai, Y.; He, J.; Lu, L. Predicting sumoylation site by feature selection method. *J. Biomol. Struct. Dyn.* **2011**, *28*, 797–804. [CrossRef] [PubMed]

16. Chen, L.; Zhang, Y.-H.; Huang, G.; Pan, X.; Wang, S.; Huang, T.; Cai, Y.-D. Discriminating cirRNAs from other lncRNAs using a hierarchical extreme learning machine (H-ELM) algorithm with feature selection. *Mol. Genet. Genom.* **2018**, *293*, 137–149. [CrossRef] [PubMed]

17. Lu, J.; Wang, S.; Cai, Y.D.; Zhang, Q. Analysis and prediction of nitrated tyrosine sites with mRMR method and support vector machine algorithm. *Curr. Bioinform.* **2017**, *13*, 3–13.

18. Liu, L.; Chen, L.; Zhang, Y.H.; Wei, L.; Cheng, S.; Kong, X.; Zheng, M.; Huang, T.; Cai, Y.D. Analysis and prediction of drug-drug interaction by minimum redundancy maximum relevance and incremental feature selection. *J. Biomol. Struct. Dyn.* **2017**, *35*, 312–329. [CrossRef] [PubMed]

19. Chen, L.; Zhang, Y.H.; Huang, T.; Cai, Y.D. Gene expression profiling gut microbiota in different races of humans. *Sci. Rep.* **2016**, *6*, 23075. [CrossRef] [PubMed]

20. Ni, Q.; Chen, L. A feature and algorithm selection method for improving the prediction of protein structural classes. *Comb. Chem. High Throughput Screen.* **2017**, *20*, 612–621. [CrossRef] [PubMed]

21. Chen, L.; Zhang, Y.H.; Zheng, M.; Huang, T.; Cai, Y.D. Identification of compound-protein interactions through the analysis of gene ontology, kegg enrichment for proteins and molecular fragments of compounds. *Mol. Genet. Genom.* **2016**, *291*, 2065–2079. [CrossRef] [PubMed]

22. Wang, S.; Zhang, Y.-H.; Huang, G.; Chen, L.; Cai, Y.-D. Analysis and prediction of myristoylation sites using the mRMR method, the ifs method and an extreme learning machine algorithm. *Comb. Chem. High Throughput Screen.* **2017**, *20*, 96–106. [CrossRef] [PubMed]

23. Chen, L.; Wang, S.; Zhang, Y.-H.; Wei, L.; Xu, X.; Huang, T.; Cai, Y.-D. Prediction of nitrated tyrosine residues in protein sequences by extreme learning machine and feature selection methods. *Comb. Chem. High Throughput Screen.* **2018**, *21*, 393–402. [CrossRef] [PubMed]

24. Li, B.Q.; Zheng, L.L.; Hu, L.L.; Feng, K.Y.; Huang, G.; Chen, L. Prediction of linear B-ceel epitopes with mRMR feature selection and analysis. *Curr. Bioinform.* **2016**, *11*, 22–31. [CrossRef]

25. Chen, L.; Pan, X.; Hu, X.; Zhang, Y.-H.; Wang, S.; Huang, T.; Cai, Y.-D. Gene expression differences among different MSI statuses in colorectal cancer. *Int. J. Cancer* **2018**. [CrossRef] [PubMed]

26. Platt, J. *Sequential Minimal Optimizaton: A Fast Algorithm for Training Support Vector Machines*; Technical Report MSR-TR-98-14; Microsoft Res: Redmon, WA, USA, 1998.

27. Frank, E.; Hall, M.; Trigg, L.; Holmes, G.; Witten, I.H. Data mining in bioinformatics using weka. *Bioinformatics* **2004**, *20*, 2479–2481. [CrossRef] [PubMed]

28. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the International Joint Conference on Artificial Intelligence, Montreal, QC, Canada, 20–25 August 1995; Lawrence Erlbaum Associates Ltd.: Mahwah, NJ, USA, 1995; pp. 1137–1145.

29. Chen, L.; Wang, S.; Zhang, Y.-H.; Li, J.; Xing, Z.-H.; Yang, J.; Huang, T.; Cai, Y.-D. Identify key sequence features to improve CRISPR sgRNA efficacy. *IEEE Access* **2017**, *5*, 26582–26590. [CrossRef]

30. Wang, D.; Li, J.-R.; Zhang, Y.-H.; Chen, L.; Huang, T.; Cai, Y.-D. Identification of differentially expressed genes between original breast cancer and xenograft using machine learning algorithms. *Genes* **2018**, *9*, 155. [CrossRef] [PubMed]

31. Chen, L.; Chu, C.; Zhang, Y.-H.; Zheng, M.-Y.; Zhu, L.; Kong, X.; Huang, T. Identification of drug-drug interactions using chemical interactions. *Curr. Bioinform.* **2017**, *12*, 526–534. [CrossRef]

32. Chen, L.; Zeng, W.M.; Cai, Y.D.; Feng, K.Y.; Chou, K.C. Predicting anatomical therapeutic chemical (ATC) classification of drugs by integrating chemical-chemical interactions and similarities. *PLoS ONE* **2012**, *7*, e35254. [CrossRef] [PubMed]

33. Matthews, B.W. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta* **1975**, *405*, 442–451. [CrossRef]

34. Gorodkin, J. Comparing two K-category assignments by a K-category correlation coefficient. *Comput. Biol. Chem.* **2004**, *28*, 367–374. [CrossRef] [PubMed]

35. Lizio, M.; Harshbarger, J.; Shimoji, H.; Severin, J.; Kasukawa, T.; Sahin, S.; Abugessaisa, I.; Fukuda, S.; Hori, F.; Ishikawa-Kato, S.; et al. Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol.* **2015**, *16*, 22. [CrossRef] [PubMed]

36. eGTEx Project; Stranger, B.E.; Brigham, L.E.; Hasz, R.; Hunter, M.; Johns, C.; Johnson, M.; Kopen, G.; Leinweber, W.F.; Lonsdale, J.T.; et al. Enhancing gtex by bridging the gaps between genotype, gene expression, and disease. *Nat. Genet.* **2017**, *49*, 1664. [CrossRef] [PubMed]

37. Papatheodorou, I.; Fonseca, N.A.; Keays, M.; Tang, Y.A.; Barrera, E.; Bazant, W.; Burke, M.; Fullgrabe, A.; Fuentes, A.M.; George, N.; et al. Expression atlas: Gene and protein expression across multiple studies and organisms. *Nucleic Acids Res.* **2018**, *46*, D246–D251. [CrossRef] [PubMed]

38. Lee, A.W. The role of atypical protein kinase C in CSF-1-dependent ERK activation and proliferation in myeloid progenitors and macrophages. *PLoS ONE* **2011**, *6*, e25580. [CrossRef] [PubMed]

39. Kang, Z.H.; Xu, F.; Zhang, Q.A.; Wu, Z.Y.; Zhang, X.J.; Xu, J.H.; Luo, Y.; Guan, M. Oncogenic mutations in extramammary Paget's disease and their clinical relevance. *Int. J. Cancer* **2013**, *132*, 824–831. [CrossRef] [PubMed]

40. Li, D.; Lu, Z.Y.; Jia, J.Y.; Zheng, Z.F.; Lin, S. Changes in microRNAs associated with podocytic adhesion damage under mechanical stress. *J. Renin-Angiotensin Aldosterone Syst.* **2013**, *14*, 97–102. [CrossRef] [PubMed]

41. Pinatel, E.M.; Orso, F.; Penna, E.; Cimino, D.; Elia, A.R.; Circosta, P.; Dentelli, P.; Brizzi, M.F.; Provero, P.; Taverna, D. miR-223 is a coordinator of breast cancer progression as revealed by bioinformatics predictions. *PLoS ONE* **2014**, *9*, e84859. [CrossRef] [PubMed]

42. O'Connell, G.C.; Treadway, M.B.; Petrone, A.B.; Tennant, C.S.; Lucke-Wold, N.; Chantler, P.D.; Barr, T.L. Peripheral blood AKAP7 expression as an early marker for lymphocyte-mediated post-stroke blood brain barrier disruption. *Sci. Rep.* **2017**, *7*, 1172. [CrossRef] [PubMed]

43. Van der Vaart, B.; Franker, M.A.M.; Kuijpers, M.; Hua, S.S.; Bouchet, B.P.; Jiang, K.; Grigoriev, I.; Hoogenraad, C.C.; Akhmanova, A. Microtubule plus-end tracking proteins SLAIN1/2 and ch-TOG promote axonal development. *J. Neurosci.* **2012**, *32*, 14722–14729. [CrossRef] [PubMed]

44. Suchy-Dicey, A.; Heckbert, S.R.; Smith, N.L.; McKnight, B.; Rotter, J.I.; Chen, Y.I.; Psaty, B.M.; Enquobahrie, D.A. Gene expression in thiazide diuretic or statin users in relation to incident type 2 diabetes. *Int. J. Mol. Epidemiol. Genet.* **2014**, *5*, 22–30. [PubMed]
45. Cowell, J.K.; Lo, K.C.; Luce, J.; Hawthorn, L. Interpreting aCGH-defined karyotypic changes in gliomas using copy number status, loss of heterozygosity and allelic ratios. *Exp. Mol. Pathol.* **2010**, *88*, 82–89. [CrossRef] [PubMed]
46. Zhou, M.; Ye, Z.; Gu, Y.; Tian, B.; Wu, B.; Li, J. Genomic analysis of drug resistant pancreatic cancer cell line by combining long non-coding RNA and mRNA expression profling. *Int. J. Clin. Exp. Pathol.* **2015**, *8*, 38–52. [PubMed]
47. Gao, Y.F.; Zhu, T.; Mao, C.X.; Liu, Z.X.; Wang, Z.B.; Mao, X.Y.; Li, L.; Yi, J.Y.; Zhou, H.H.; Liu, Z.Q. PPIC, EMP3 and CHI3L1 are novel prognostic markers for high grade glioma. *Int. J. Mol. Sci.* **2016**, *17*, 1808. [CrossRef] [PubMed]
48. Romero-Saavedra, F.; Laverde, D.; Wobser, D.; Michaux, C.; Budin-Verneuil, A.; Bernay, B.; Benachour, A.; Hartke, A.; Huebner, J. Identification of peptidoglycan-associated proteins as vaccine candidates for enterococcal infections. *PLoS ONE* **2014**, *9*, e111880. [CrossRef] [PubMed]
49. Krizhanovsky, V.; Ben-Arie, N. A novel role for the choroid plexus in BMP-mediated inhibition of differentiation of cerebellar neural progenitors. *Mech. Dev.* **2006**, *123*, 67–75. [CrossRef] [PubMed]
50. Ohtori, S.; Yamamoto, T.; Ino, H.; Hanaoka, E.; Shinbo, J.; Ozaki, T.; Takada, N.; Nakamura, Y.; Chiba, T.; Nakagawara, A.; et al. Differential screening-selected gene aberrative in neuroblastoma protein modulates inflammatory pain in the spinal dorsal horn. *Neuroscience* **2002**, *110*, 579–586. [CrossRef]
51. Yi, C.H.; Zheng, T.Z.; Leaderer, D.; Hoffman, A.; Zhu, Y. Cancer-related transcriptional targets of the circadian gene NPAS2 identified by genome-wide ChIP-on-chip analysis. *Cancer Lett.* **2009**, *284*, 149–156. [CrossRef] [PubMed]
52. Siggs, O.M.; Popkin, D.L.; Krebs, P.; Li, X.H.; Tang, M.; Zhan, X.M.; Zeng, M.; Lin, P.; Xia, Y.; Oldstone, M.B.A.; et al. Mutation of the er retention receptor kdelr1 leads to cell-intrinsic lymphopenia and a failure to control chronic viral infection. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E5706–E5714. [CrossRef] [PubMed]
53. Zhang, L.; Wang, T.; Valle, D. Reduced PLP2 expression increases ER-stress-induced neuronal apoptosis and risk for adverse neurological outcomes after hypoxia ischemia injury. *Hum. Mol. Genet.* **2015**, *24*, 7221–7226. [CrossRef] [PubMed]
54. Zhu, H.; Miao, M.H.; Ji, X.Q.; Xue, J.; Shao, X.J. miR-664 negatively regulates PLP2 and promotes cell proliferation and invasion in T-cell acute lymphoblastic leukaemia. *Biochem. Biophys. Res. Commun.* **2015**, *459*, 340–345. [CrossRef] [PubMed]
55. Dorsey, N.J.; Chapoval, S.P.; Smith, E.P.; Skupsky, J.; Scott, D.W.; Keegan, A.D. STAT6 controls the number of regulatory T cells in vivo, thereby regulating allergic lung inflammation. *J. Immunol.* **2013**, *191*, 1517–1528. [CrossRef] [PubMed]
56. Myklebust, J.H.; Irish, J.M.; Brody, J.; Czerwinski, D.K.; Houot, R.; Kohrt, H.E.; Timmerman, J.; Said, J.; Green, M.R.; Delabie, J.; et al. High PD-1 expression and suppressed cytokine signaling distinguish T cells infiltrating follicular lymphoma tumors from peripheral T cells. *Blood* **2013**, *121*, 1367–1376. [CrossRef] [PubMed]
57. Weber, M.S.; Prod'homme, T.; Youssef, S.; Dunn, S.E.; Steinman, L.; Zamvil, S.S. Neither T-helper type 2 nor Foxp3+ regulatory T cells are necessary for therapeutic benefit of atorvastatin in treatment of central nervous system autoimmunity. *J. Neuroinflamm.* **2014**, *11*, 29. [CrossRef] [PubMed]
58. Martin-Vilchez, S.; Whitmore, L.; Asmussen, H.; Zareno, J.; Horwitz, R.; Newell-Litwa, K. RhoGTPase regulators orchestrate distinct stages of synaptic development. *PLoS ONE* **2017**, *12*, e0170464. [CrossRef] [PubMed]
59. Katoh, M.; Katoh, M. Characterization of human ARHGAP10 gene in silico. *Int. J. Oncol.* **2004**, *25*, 1201–1206. [PubMed]
60. Hellstrom, M.; Ericsson, M.; Johansson, B.; Faraz, M.; Anderson, F.; Henriksson, R.; Nilsson, S.K.; Hedman, H. Cardiac hypertrophy and decreased high-density lipoprotein cholesterol in Lrig3-deficient mice. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **2016**, *310*, R1045–R1052. [CrossRef] [PubMed]

61. Abraira, V.E.; Satoh, T.; Fekete, D.M.; Goodrich, L.V. Vertebrate Lrig3-erbb interactions occur in vitro but are unlikely to play a role in Lrig3-dependent inner ear morphogenesis. *PLoS ONE* **2010**, *5*, e8981. [CrossRef] [PubMed]

62. Abraira, V.E.; del Rio, T.; Tucker, A.F.; Slonimsky, J.; Keirnes, H.L.; Goodrich, L.V. Cross-repressive interactions between Lrig3 and netrin 1 shape the architecture of the inner ear. *Development* **2008**, *135*, 4091–4099. [CrossRef] [PubMed]

63. Jansson, L.; Larsson, J. Normal hematopoietic stem cell function in mice with enforced expression of the hippo signaling effector YAP1. *PLoS ONE* **2012**, *7*, e32013. [CrossRef] [PubMed]

64. Hoshiba, T.; Otaki, T.; Nemoto, E.; Maruyama, H.; Tanaka, M. Blood-compatible polymer for hepatocyte culture with high hepatocyte-specific functions toward bioartificial liver development. *ACS Appl. Mater. Interfaces* **2015**, *7*, 18096–18103. [CrossRef] [PubMed]

65. Loke, S.Y.; Wong, P.T.; Ong, W.Y. Global gene expression changes in the prefrontal cortex of rabbits with hypercholesterolemia and/or hypertension. *Neurochem. Int.* **2017**, *102*, 33–56. [CrossRef] [PubMed]

66. Yamashita, A.; Taniwaki, T.; Kaikoi, Y.; Yamazaki, T. Protective role of the endoplasmic reticulum protein mitsugumin23 against ultraviolet C-induced cell death. *FEBS Lett.* **2013**, *587*, 1299–1303. [CrossRef] [PubMed]

67. Reiss, J.; Hahnewald, R. Molybdenum cofactor deficiency: Mutations in *GPHN*, *MOCS1*, and *MOCS2*. *Hum. Mutat.* **2011**, *32*, 10–18. [CrossRef] [PubMed]

68. Wang, J.; Krizowski, S.; Fischer-Schrader, K.; Niks, D.; Tejero, J.; Sparacino-Watkins, C.; Wang, L.; Ragireddy, V.; Frizzell, S.; Kelley, E.E.; et al. Sulfite oxidase catalyzes single-electron transfer at molybdenum domain to reduce nitrite to nitric oxide. *Antioxid. Redox Signal.* **2015**, *23*, 283–294. [CrossRef] [PubMed]

69. Ricketts, C.D.; Bates, W.R.; Reid, S.D. The effects of acute waterborne exposure to sublethal concentrations of molybdenum on the stress response in rainbow trout, oncorhynchus mykiss. *PLoS ONE* **2015**, *10*, e0115334. [CrossRef] [PubMed]

70. Stewart, K.; Uetani, N.; Hendriks, W.; Tremblay, M.L.; Bouchard, M. Inactivation of LAR family phosphatase genes Ptprs and Ptprf causes craniofacial malformations resembling pierre-robin sequence. *Development* **2013**, *140*, 3413–3422. [CrossRef] [PubMed]

71. Unoki, M.; Shen, J.C.; Zheng, Z.M.; Harris, C.C. Novel splice variants of ing4 and their possible roles in the regulation of cell growth and motility. *J. Biol. Chem.* **2006**, *281*, 34677–34686. [CrossRef] [PubMed]

72. Silver, D.J.; Siebzehnrubl, F.A.; Schildts, M.J.; Yachnis, A.T.; Smith, G.M.; Smith, A.A.; Scheffler, B.; Reynolds, B.A.; Silver, J.; Steindler, D.A. Chondroitin sulfate proteoglycans potently inhibit invasion and serve as a central organizer of the brain tumor microenvironment. *J. Neurosci.* **2013**, *33*, 15603–15617. [CrossRef] [PubMed]

73. Park, J.; Lee, J.; Choi, C. Evaluation of drug-targetable genes by defining modes of abnormality in gene expression. *Sci. Rep.* **2015**, *5*, 13576. [CrossRef] [PubMed]

74. Desh, H.; Gray, S.L.; Horton, M.J.; Raoul, G.; Rowlerson, A.M.; Ferri, J.; Vieira, A.R.; Sciote, J.J. Molecular motor MYO1C, acetyltransferase KAT6B and osteogenetic transcription factor RUNX2 expression in human masseter muscle contributes to development of malocclusion. *Arch. Oral Biol.* **2014**, *59*, 601–607. [CrossRef] [PubMed]

75. Toyoda, T.; An, D.; Witczak, C.A.; Koh, H.J.; Hirshman, M.F.; Fujii, N.; Goodyear, L.J. *Myo1c* regulates glucose uptake in mouse skeletal muscle. *J. Biol. Chem.* **2011**, *286*, 4133–4140. [CrossRef] [PubMed]

76. Akahane, K.; Inukai, T.; Zhang, X.; Hirose, K.; Kuroda, I.; Goi, K.; Honna, H.; Kagami, K.; Nakazawa, S.; Endo, K.; et al. Resistance of t-cell acute lymphoblastic leukemia to tumor necrosis factor–related apoptosis-inducing ligand-mediated apoptosis. *Exp. Hematol.* **2010**, *38*, 885–895. [CrossRef] [PubMed]

77. Yu, R.; Mao, J.; Yang, Y.; Zhang, Y.; Tian, Y.; Zhu, J. Protective effects of calcitriol on diabetic nephropathy are mediated by down regulation of TGF-β1 and CIP4 in diabetic nephropathy rat. *Int. J. Clin. Exp. Pathol.* **2015**, *8*, 3503–3512. [PubMed]

78. Aulak, K.S.; Davis, A.E., III; Donaldson, V.H.; Harrison, R.A. Chymotrypsin inhibitory activity of normal C1-inhibitor and a P1 arg to his mutant: Evidence for the presence of overlapping reactive centers. *Protein Sci. Publ. Protein Soc.* **1993**, *2*, 727–732. [CrossRef] [PubMed]

79. Katoh, Y.; Imakagura, H.; Futatsumori, M.; Nakayama, K. Recruitment of clathrin onto endosomes by the Tom1-Tollip complex. *Biochem. Biophys. Res. Commun.* **2006**, *341*, 143–149. [CrossRef] [PubMed]

80. Huang, D.W.; Sherman, B.T.; Lempicki, R.A. Systematic and integrative analysis of large gene lists using david bioinformatics resources. *Nat. Protoc.* **2009**, *4*, 44–57. [CrossRef] [PubMed]