Journal of proteome research

ARTICLE

pubs.acs.org/jpr

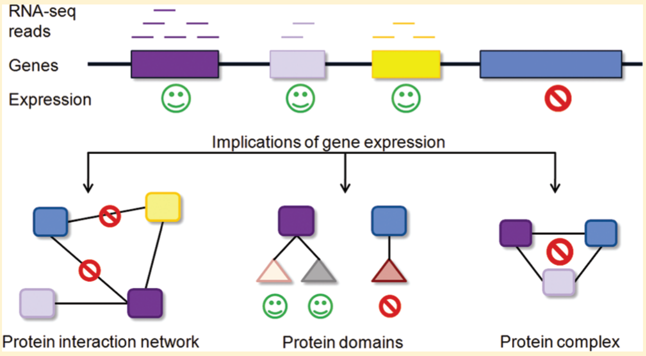
Tissue-Specific Proteins and Functional Implications

Dorothea Emig and Mario Albrecht*

Department of Computational Biology and Applied Algorithmics, Max Planck Institute for Informatics, Campus E1.4, 66123 Saarbrücken, Germany

S Supporting Information

ABSTRACT: Tissue-specific gene expression can result in the presence or absence of certain protein interactions and complexes, leading to profound functional differences of biological processes between the tissues. In this study, we integrate human gene expression data based on RNA-sequencing with protein interactions, domains and complexes to analyze the functional implications of their tissue specificity. This reveals that tissue specificity is characterized by much fewer proteins, domains, interactions, and complexes than previously thought. In contrast to previous microarray studies, our analysis based on RNA-sequencing suggests that tissue-specific protein interactions are less common and mainly involved with transmembrane transport and receptor activation. Additionally, tissue-specific protein domains show enrichments in DNA-related functions. This confirms that receptor-activated signaling processes and transcriptional regulation are two key factors for tissue specificity. Furthermore, many protein complexes are widely expressed regardless of their size, and their formation is frequently controlled by very few tissue-specific proteins. Interestingly, the number of alternative transcripts is increased for widely expressed genes. This suggests that alternative splicing plays a prominent role in generating specific functional characteristics of tissues.



KEYWORDS: tissue specificity, protein interaction, protein complex, protein domain, alternative splicing, RNA sequencing

INTRODUCTION

Proteins are involved in almost all biological processes, and their interactions are essential for the survival of cells. In the last years, many research projects were performed to identify universally expressed and tissue-specific genes and their products. This included the quest for specific molecular characteristics involved with the generation of tissue diversity.

Lehner and Fraser discovered tissue-specific mammalian protein domains and their cellular functions in comparison with universally expressed protein domains.[1] Another work by Bossi and Lehner focused on the tissue-specific occurrence of human protein interactions.[2] On the basis of gene expression data from microarray experiments, they showed that many protein interactions are tissue-specific. They also found that universally expressed proteins frequently interact with tissue-specific ones. Other studies concentrated on the identification and analysis of disease-related genes and protein complexes by examining the tissue specificity of known disease genes and their features such as evolutionary conservation and functional annotation.[3,4] Eisenberg and Levanon analyzed the genomic structure of universally expressed genes (housekeeping genes). They revealed that it is more compact than the structure of tissue-specific genes, that is, the number of exons, introns and untranslated regions is lower and the regions are shorter in length.[5] The results of follow-up research by She and colleagues suggested that CpG islands are enriched in transcription start sites of universally expressed genes.[6] In addition, Farré and colleagues observed that the sequence conservation of promoters of universally expressed genes is significantly lower than of tissue-specific ones.[7]

Most of the above-mentioned analyses on tissue-specific gene expression and structural and functional properties of genes and their products were based on microarray experiments. One of the most extensive experiments resulted in the Novartis Gene Atlas,[8] which provides gene expression profiles for 79 human tissues and cell lines. However, statistical methods for analyzing microarray data are limited in their ability of distinguishing between low gene expression and experimental noise. Thus, expression profiles are error-prone especially for genes expressed at low levels, giving rise to the misclassification of genes regarding their tissue specificity. This was confirmed in a recent study by Zhu and colleagues who compared various gene expression data sets.[9] They concluded that our current knowledge about universally expressed genes is very deficient and that their number is considerably underestimated. Recently, these findings were further supported by gene expression analyses based on next-generation RNA-sequencing data.[10,11] Here, the authors discovered that the number of universally expressed genes is by far higher than estimated in previous microarray-based studies.

Received: November 10, 2010
Published: February 22, 2011

Since accurate knowledge of tissue-specific gene and protein information is of great importance for understanding biological functions and determining biomarkers and drug targets,[12] we examined the functional implications of the gene expression data generated by RNA-sequencing for 15 human tissues and cell lines. In the following, we present our analysis of the tissue occurrence of protein interactions, domains, and complexes as well as of transcript isoforms. We identify tissue-specific elements and biological functions and investigate to which extent former findings based on microarrays are still in agreement with new results based on RNA-sequencing.

■ MATERIALS AND METHODS

Data Sources

We used the Ensembl database release 55 for all analyses.[13] For protein domains, we used Pfam annotations as provided by Ensembl together with GO annotations of the Pfam release 24.[14] We obtained gene expression estimates (RPKM values) for 10 human tissues and 5 human cell lines from the supplementary data of the alternative splicing study by Wang et al.[15] The RPKM value is a measure of the number of RNA-sequencing reads mapped to the constitutive exons of a certain gene and reflects whether a gene is transcribed in the tissue or not.[16] For each of the tissues and cell lines, the data contained about 20 million reads, of which about 60% could be uniquely mapped to the genome. We defined a gene to be expressed in a certain tissue or cell line if the respective RPKM value was at least 0.3. This is a reasonable gene expression threshold as has been shown by Ramskold et al., who examined different RPKM gene expression thresholds in an extensive gene expression study using the RNA-sequencing data.[11] Drugs and drug targets were taken from the supplementary data provided in the study by Yildirim et al.[17]

Universal and Tissue-Specific Proteins

We used two alternative definitions of universal and tissue-specific expression of genes encoding the respective proteins. The first definition is based on the mRNA presence and absence in the tissues and cell lines (P/A definition, PAD). We defined a protein to be universal if the corresponding gene was expressed in at least 14 of the 15 tissues and cell lines, accounting for potentially inaccurate read-to-genome mappings due to noisy and incomplete data. Accordingly, tissue-specific proteins are the products of genes expressed in at most two tissues and cell lines.

The second definition combines mRNA presence and absence with tissue-specific overexpression (Peak definition, PKD). As for PAD, we regarded a protein as tissue-specific if gene expression was detected in at most two tissues and cell lines. For genes expressed in at least three tissues and cell lines, we applied an overexpression analysis adapted from the study by Winter and colleagues[18] and computed the multinomial distribution of the expression values. For each gene, we checked whether the expression levels were equal in the 15 tissues and cell lines or whether the gene was overexpressed in one or two of them. To this end, we computed the tissue specificity value $\mathrm{TS}(g)_t$ of a gene g expressed in tissue t using the following formula:

$$\mathrm{TS}(g)_t = \frac{\mathrm{RPKM}_t}{\sum_{\hat{t}\,\in\,\mathrm{tissues}} \mathrm{RPKM}_{\hat{t}}}$$

$\mathrm{TS}(g)_t$ describes the contribution of the gene expression level (RPKM_t) of gene g in tissue t to the sum of the expression levels of gene g in all 15 tissues and cell lines. As defined by Winter et al.,[18] we regarded a gene to be tissue-specific if both its maximum tissue specificity value $\max(\mathrm{TS}(g)_t)$ was above 0.4 and the corresponding RPKM_t value was above the mean expression level in the respective tissue or cell line. The second requirement ensures that genes expressed at low levels in most tissues and cell lines and at a slightly higher level in another tissue are not identified as outliers due to only moderate differences in mRNA levels. Genes expressed in at least 14 tissues and cell lines with $\max(\mathrm{TS}(g)_t)$ below 0.4 (i.e., without a clear overexpression in at least one tissue or cell line) or with an RPKM_t expression value below the mean were defined as universal.

PAD and PKD

In the following, we use both alternative definitions of tissue-specific gene expression, PAD and PKD, and compare potential differences in the biological results. However, concerning PKD based on gene overexpression, it is of note that gene expression levels do not necessarily reflect protein abundances.[19,20] Furthermore, the alternative definitions naturally lead to different classifications of genes. Genes identified as tissue-specific according to PAD are a subset of those detected with PKD. Similarly, genes classified as universally expressed with PKD are a subset of those found using PAD. This means that genes with strongly varying expression levels in all tissues and cell lines will be classified as universally expressed by PAD, but some of them might be regarded as tissue-specific based on PKD.

Protein Interaction Data

We mapped the gene expression data from RNA-sequencing results onto the large-scale data set of 80 923 physical protein—protein interactions compiled by Bossi and Lehner for the human interactome.[2] We retained only those interactions where both interacting proteins had expression values available. This reduced the data set to a protein interaction network of 63 815 interactions involving 8805 human proteins.

Universal and Tissue-Specific Protein Interactions

We defined a protein interaction to be universal (or universally expressed) if the interacting proteins were co-expressed in at least 14 of the 15 tissues and cell lines. According to this definition, an interaction of two universal proteins is not a universal protein interaction if the universal proteins are not co-expressed in the required number of tissues and cell lines. Additionally, we defined a tissue-specific interaction to occur in at most two tissues and cell lines. This means that the interacting proteins do not need to be tissue-specific themselves, but they have to co-occur in at most two tissues and cell lines.

Protein Domain Data

We obtained the Pfam domain annotations from Ensembl with 17 289 genes encoding at least one protein domain. For 3908 of these genes, we did not have expression data available and thus excluded them from further analyses. This provided 13 381 genes encoding a total of 3330 different Pfam domains. For each of these Pfam domains, we computed the percentage of genes encoding the respective Pfam domain that had expression values available. To increase the reliability of the analysis, we excluded Pfam domains that had expression values assigned for fewer than 75% of the encoding genes. This reduced the number of Pfam domains to 2840.

Universal and Tissue-Specific Domains

Protein domain expression was determined by the genes encoding the respective Pfam domain. For each domain, we averaged the number of tissues in which the genes encoding the domain are expressed. We defined a domain to be universal if the average number of tissues and cell lines was greater than 13, and to be tissue-specific if it was less than 3.

Protein Complex Data

Protein complexes were obtained from CORUM and PDB, downloaded in August 2009.[21,22] CORUM does usually not provide information on the stoichiometry of the complexes and reports only the names of the co-complexed proteins. Therefore, we disregarded the stoichiometry of the complexes for both CORUM and PDB and required at least three different proteins to be contained in a complex. We excluded binary complexes because they represent the physical interaction of two proteins, which we regard as a protein−protein interaction as in the protein interaction network described above.

The complexes were given by UniProt identifiers and mapped to Ensembl gene identifiers using the available Ensembl annotations in BioMart.[23] We retained only those complexes if all of the co-complexed proteins could be mapped to Ensembl gene identifiers. Furthermore, we required the availability of expression values for all proteins in a complex. This yielded 648 complexes.

Completeness and Tissue Specificity of Protein Complexes

In case of protein complexes, we combined the tissue specificity with the completeness of a complex. We first classified a protein complex according to its completeness in a specific tissue or cell line, that is, the fraction of expressed co-complexed proteins. If all co-complexed proteins were expressed, the complex was regarded as *fully expressed* in this tissue or cell line. If at least two of the co-complexed proteins, but not all of them, were expressed, we defined the complex to be *partially expressed* in this tissue or cell line. Protein complexes with less than two expressed proteins were called *absent* in this tissue or cell line.

From the completeness of a complex in each of the 15 tissues and cell lines, we inferred the tissue specificity by counting the number of tissues and cell lines in which the complex was fully expressed. We defined a universal protein complex to be fully expressed in at least 14 tissues and cell lines, and tissue-specific complexes to be fully expressed in at most two tissues and cell lines.

Computation of Pairwise Tissue Similarities

We performed pairwise comparisons of the tissues and cell lines based on their gene expression profiles. Let t_1 and t_2 be two tissues or cell lines, and let TSG_1 and TSG_2 be the sets of tissue-specific genes expressed in the respective tissues and cell lines. Then, the pairwise similarity simG is computed by normalizing the number of shared tissue-specific genes by the size of the smaller set of TSG_1 and TSG_2:

$$\mathrm{simG}(t_1, t_2) = \frac{|\mathrm{TSG}_1 \cap \mathrm{TSG}_2|}{\min(|\mathrm{TSG}_1|, |\mathrm{TSG}_2|)}$$

Furthermore, we computed pairwise tissue similarities based on their protein interaction profiles. Let TSPPI_1 and TSPPI_2 be the sets of tissue-specific protein interactions that occur in the respective tissues and cell lines. Then, the pairwise tissue similarity simPPI can be computed by normalizing the number of shared tissue-specific protein interactions by the size of the smaller set of TSG_1 and TSG_2:

$$\mathrm{simPPI}(t_1, t_2) = \frac{|\mathrm{TSPPI}_1 \cap \mathrm{TSPPI}_2|}{\min(|\mathrm{TSPPI}_1|, |\mathrm{TSPPI}_2|)}$$

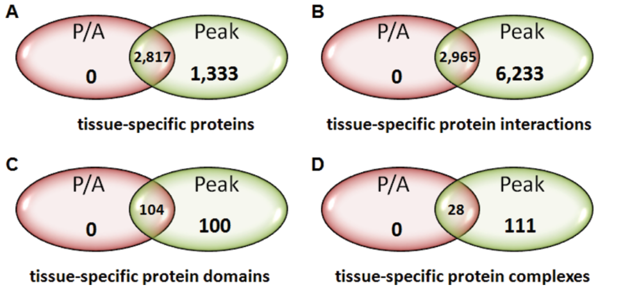

A
P/A
Peak
0
2,817
1,333
tissue-specific proteins
B
P/A
Peak
0
2,965
6,233
tissue-specific protein interactions
C
P/A
Peak
0
104
100
tissue-specific protein domains
D
P/A
Peak
0
28
111
tissue-specific protein complexes

Figure 1. Venn diagrams of tissue-specific proteins (A), protein interactions (B), protein domains (C), and protein complexes (D) according to the alternative P/A and Peak definitions.

GO Enrichments

We computed the GO term enrichments of proteins for molecular function and cellular component using the Web-based tool GOrilla[24] with a p-value threshold of 10^{-5} and default parameters otherwise. The GO term enrichments for Pfam protein domains were computed with the R package topGO using default parameters.[25]

RESULTS AND DISCUSSION

Tissue Specificity According to PAD and PKD

As can be seen in Figure 1, PAD and PKD result in the same 2817 tissue-specific proteins. The 1333 tissue-specific proteins additionally identified by PKD lead to 6233 tissue-specific protein interactions in addition to 2965 interactions detected by both PAD and PKD. For protein domains, we discovered very few highly tissue-specific domains, but still the number of tissue-specific domains nearly doubles from 104 to 204 using PKD. For protein complexes, we found only 28 to be tissue-specific according to PAD, but classified more than 100 additional complexes as tissue-specific with PKD.

Part I: Protein Interactions

In the first part of this work, we analyzed the protein interaction network in order to identify universal and tissue-specific protein interactions. Furthermore, we investigated several functional characteristics of proteins forming tissue-specific and universal interactions.

Most Protein Interactions Are Universal. We first analyzed the tissue specificity of 63 815 human protein interactions. In contrast to the previous study by Bossi and Lehner based on the Novartis Gene Atlas,[2] we find many protein interactions (∼73% for PAD and ∼69% for PKD) to occur universally (Supplementary Figure S1), and the number of tissue-specific protein interactions is surprisingly low (less than 5% for PAD and ∼14% for PKD, Table 1 top). However, in agreement with this study by Bossi and Lehner,[2] we see that, for both PAD and PKD, the tissue specificity of the protein interactions is often determined by one tissue-specific protein interacting with a universal one, while few interactions are formed by two tissue-specific proteins (Table 1, middle). Yet again contrary to the study by Bossi and Lehner,[2] we observe that the majority of universally expressed proteins (∼ 83% for PAD and ∼57% for PKD) in the protein interaction network do not interact with tissue-specific proteins at all (Table 1, bottom). In summary, our results indicate that, by far, fewer protein interactions are tissue-specific than previously thought,[2] and tissue diversity seems to involve only few tissue-specific protein interactions.

Tissues Differ in Proportion of Tissue-Specific and Absent Interactions. The number of tissue-specific protein interactions is very low throughout all tissues and cell lines (Figure 2). Although the PAD and PKD results vary in their numbers of tissue-specific protein interactions, testis always contains a comparatively high amount of tissue-specific interactions, with 977 interactions for PAD and 1040 interactions for PKD. Furthermore, both PAD and PKD find a relatively high number of tissue-specific interactions in brain, cerebellum and lymph node. The highest number of interactions that are absent in a tissue or cell line due to missing gene expression are found in skeletal muscle for both PAD and PKD (11 979 using PAD and 15 456 using PKD), followed by the breast cancer cell line BT474 and liver tissue. Overall, our results suggest that, in addition to tissue-specific interactions, the absence of certain interactions might be important as well to achieve tissue diversity.

Table 1. Tissue Specificity of Protein Interactions[a]

tissue specificity of protein interactions	number of protein interactions
universal	46291 (PAD); 44006 (PKD)
tissue-specific	2965 (PAD); 9198 (PKD)
other	14559 (PAD); 10611 (PKD)
tissue specificity of interacting proteins	**number of protein interactions**
both universal	46393 (PAD); 44091 (PKD)
universal with tissue-specific	1574 (PAD); 5971 (PKD)
both tissue-specific	158 (PAD); 974 (PKD)
other	15690 (PAD); 12779 (PKD)
number of tissue-specific interaction partners	**number of universal proteins**
0	4517 (PAD); 2976 (PKD)
1	650 (PAD); 1154 (PKD)
2	170 (PAD); 443 (PKD)
3	62 (PAD); 210 (PKD)
4−16	68 (PAD); 441 (PKD)
17−39	0 (PAD); 24 (PKD)

[a] The top part shows the number of protein interactions classified by their tissue specificity. The middle part classifies protein interactions according to the tissue specificity of the involved proteins. Many tissue-specific protein interactions are formed by a tissue-specific protein interacting with a universal protein, while interactions of two tissue-specific proteins are very rare. The bottom part gives the number of tissue-specific proteins with which a universal protein interacts.

Tissues and Cell Lines Show Similarities According to Gene Expression and Protein Interaction Profiles. We compared the tissues and cell lines regarding tissue-specific gene expression and protein interactions by computing pairwise similarities based on the number of shared tissue-specific genes and protein interactions. The pairwise similarities change depending on whether gene expression or protein interactions are analyzed. Regarding tissue-specific gene expression profiles, cerebellum and brain show the highest similarity for both PAD and PKD, while the similarities between all other tissues are equally low (Figure 3A,B). Interestingly, the investigation of tissue-specific protein interactions reveals high similarities between testis and MB435 (a cancer cell line under current discussion[26]) in addition to brain and cerebellum

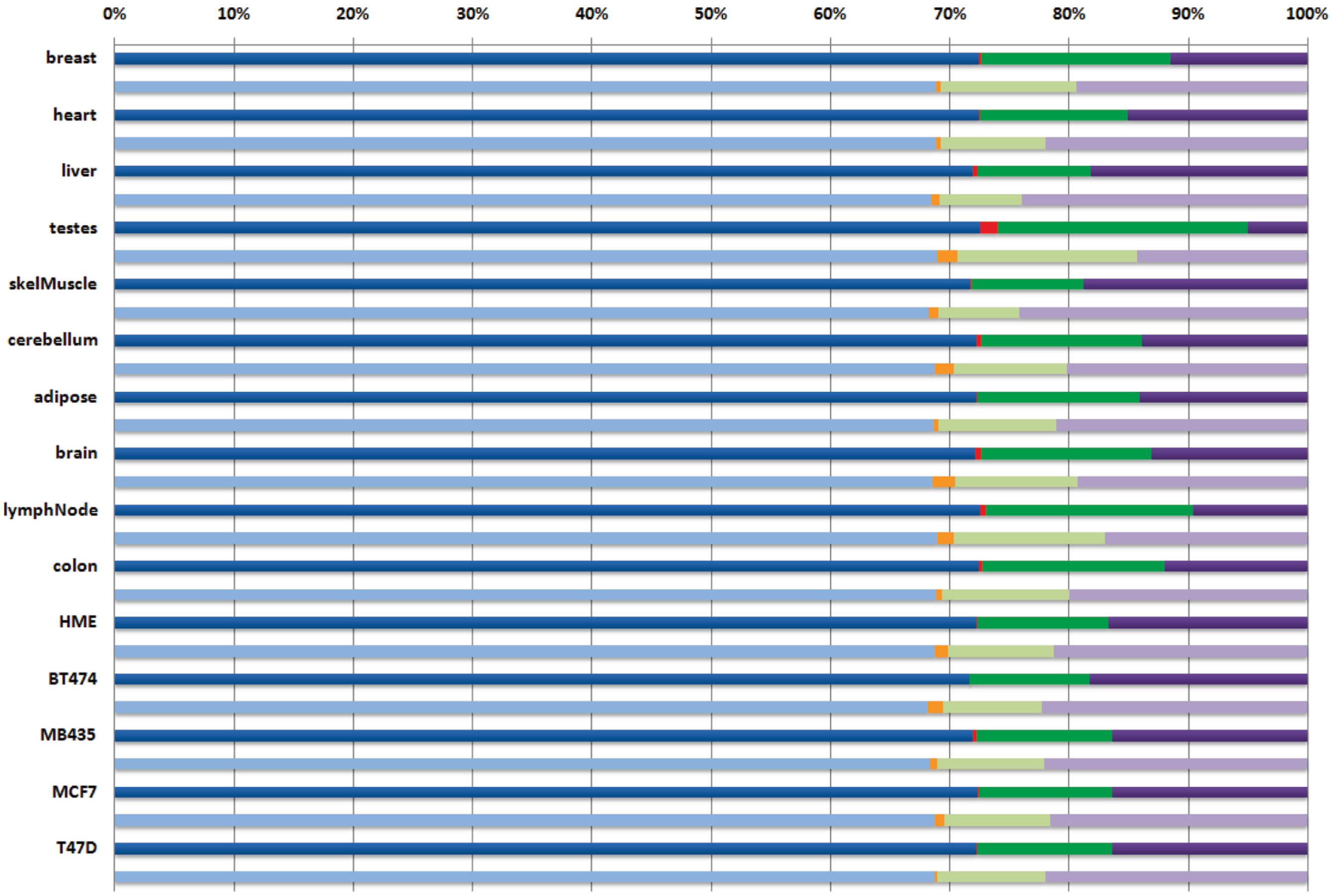


Figure 2. Tissue specificity of protein interactions. For each of the 15 tissues and cell lines, the respective percentage of universal protein interactions is shown in blue, of specific ones in red, of absent ones in purple, and of the remaining ones in green. The results obtained by using PAD and PKD are shown in dark and light colors, respectively.

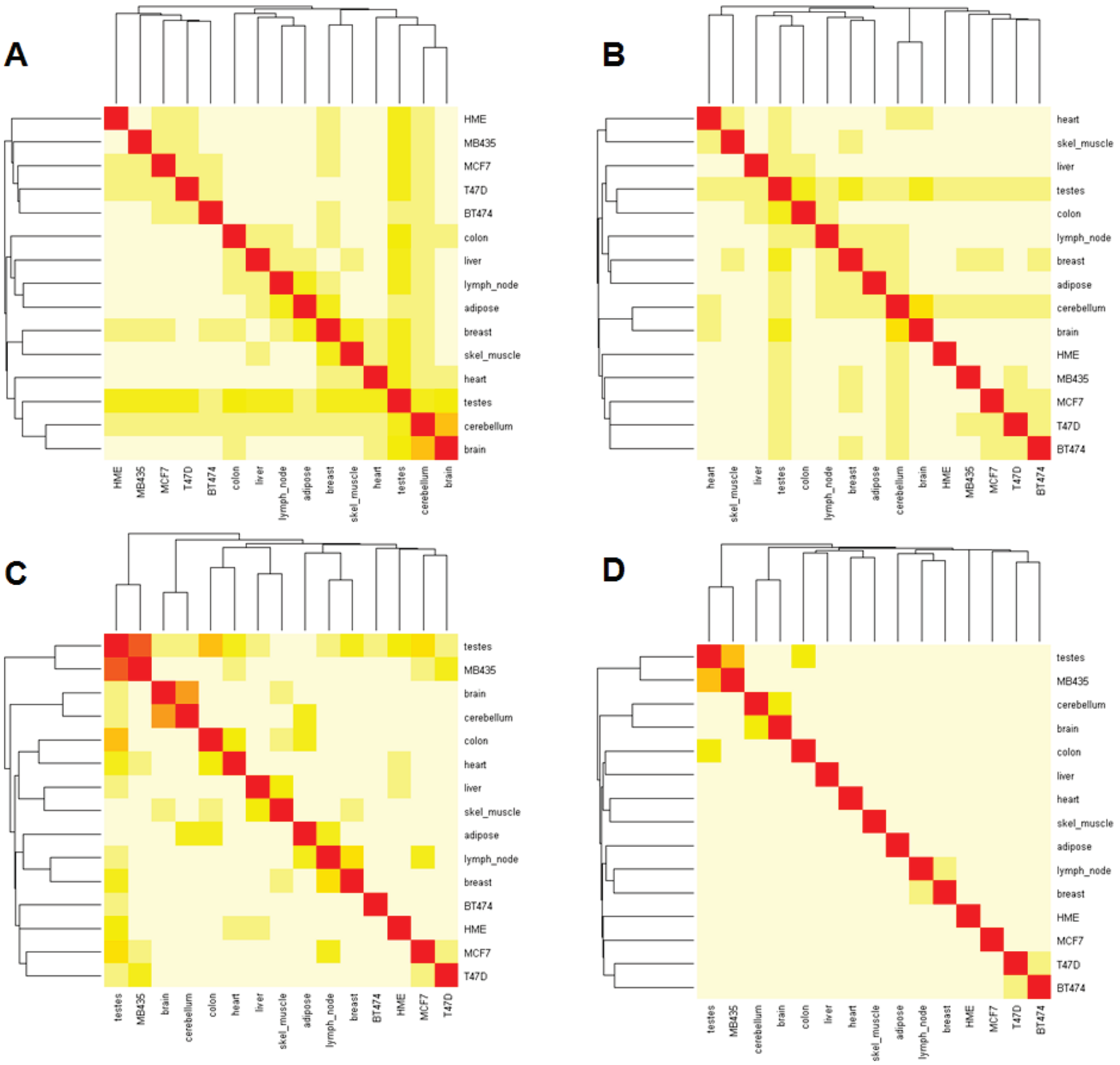


Figure 3. Pairwise tissue similarities. The heat maps show the pairwise similarities of tissues and cell lines according to their expression profiles of tissue-specific genes (A and B) or the presence of tissue-specific protein interactions (C and D). (A) and (C) are based on PAD, (B) and (D) on PKD. In each heat map, the color ranges from white (completely different) to red (identical).

(Figure 3C,D). The same trends are found for both PAD and PKD, although about 1400 more tissue-specific genes are identified with PKD than with PAD. These findings highlight that, even though the overall similarities between tissues and cell lines are low when considering all tissue-specific genes, only a fraction of the encoded proteins are actually involved in tissue-specific protein−protein interactions.

Tissue-Specific Protein Interactions Are Enriched in Transporter and Receptor Activities. We next analyzed the functions of tissue-specific proteins interacting with universal ones. Our GO term enrichment analysis regarding molecular function reveals that, according to both PAD and PKD, tissue-specific proteins are mainly involved in transporter and receptor activities, which is in agreement with previous observations.[1,18] Furthermore, tissue-specific proteins are active in the immune system and, according to PKD, involved in structural activities of the cytoskeleton (Table 2; the complete GO enrichment graphs are shown in Supplementary Figures S2 and S3). When computing enriched GO terms for cellular component, we primarily find extracellular and membrane regions for both PAD and PKD (Table 3; the complete GO enrichment graphs are shown in Supplementary Figures S4 and S5). These findings imply that many tissue-specific cellular processes are induced by the stimulation and activation of receptors, causing tissue-specific signaling cascades. Such signaling cascades can then result in tissue-specific gene expression and protein interactions.

Targets of FDA-Approved and Experimental Drugs Differ in Tissue Specificity. Drug targets need to be highly

Table 2. GO Term Enrichments for Tissue-Specific Proteins[a]

GO − Molecular Function (PAD)	*p*-value
G-protein coupled receptor activity	$1.34e^{-10}$
transmembrane receptor activity	$1.48e^{-10}$
receptor activity	$1.59e^{-10}$
cytokine activity	$1.75e^{-10}$
cytokine receptor binding	$1.19e^{-8}$
GO − Molecular Function (PKD)	***p*-value**
transporter activity	$6.07e^{-11}$
transmembrane transporter activity	$1.65e^{-10}$
substrate-specific transmembrane transporter activity	$5.07e^{-10}$
structural constituent of muscle	$7.16e^{-10}$
substrate-specific channel activity	$4.49e^{-9}$

[a] The most significant GO term enrichments for molecular function based on PAD and PKD are listed together with their *p*-values. The enrichments are computed using GOrilla for the tissue-specific proteins that interact with at least one universal protein.

Table 3. GO Term Enrichments for Tissue-Specific Proteins[a]

GO − Cellular Component (PAD)	*p*-value
extracellular region	$1.61e^{-22}$
intrinsic to plasma membrane	$1.99e^{-15}$
extracellular space	$2.45e^{-15}$
integral to plasma membrane	$4.38e^{-15}$
extracellular region part	$8.87e^{-14}$
GO − Cellular Component (PKD)	***p*-value**
plasma membrane part	$1.21e^{-27}$
plasma membrane	$1.14e^{-20}$
extracellular region	$1.30e^{-20}$
membrane part	$3.91e^{-17}$
intrinsic to membrane	$1.29e^{-16}$

[a] The most significant GO term enrichments for cellular component based on PAD and PKD are listed together with their *p*-values. The enrichments are computed using GOrilla for the tissue-specific proteins that interact with at least one universal protein.

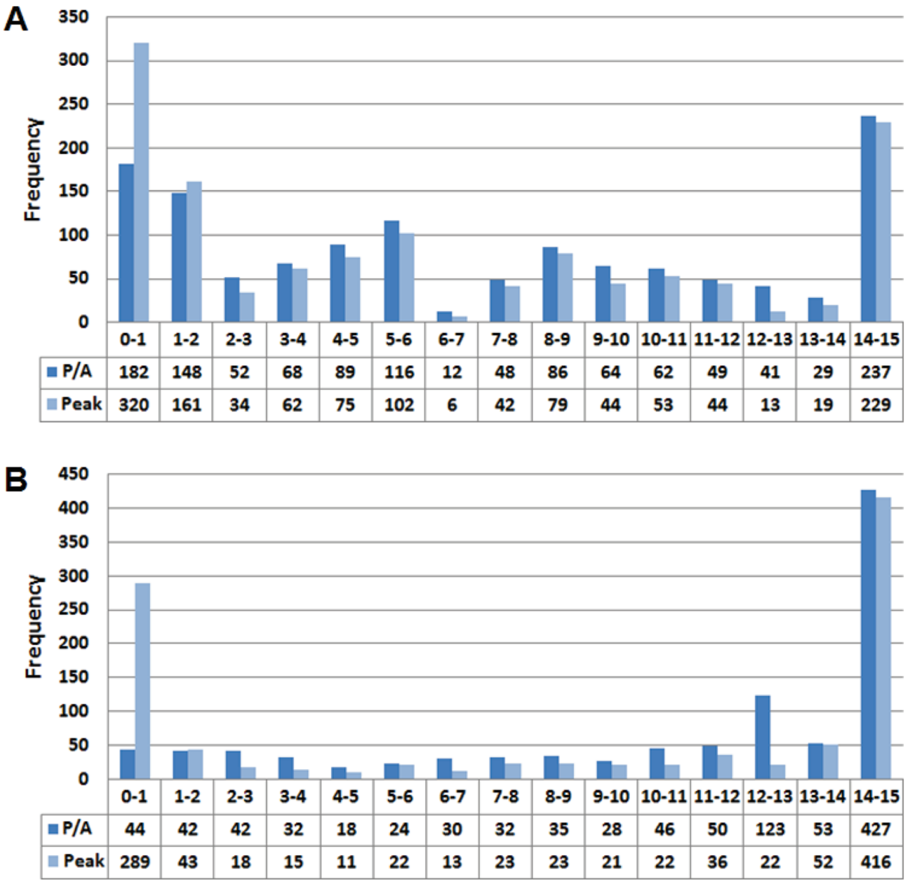
specific for a certain disease to avoid undesirable side effects. When comparing the tissue specificity of protein targets of FDA-approved and experimental drugs, we observed that many more experimental drugs target universal proteins than FDA-approved drugs (for PAD and PKD). The targets of FDA-approved drugs show a multimodal distribution regarding the number of tissues in which the targets are expressed. The two highest peaks are for tissue-specific and universal targets (Figure 4A). Targets of experimental drugs, however, are frequently universal proteins according to both PAD and PKD. PKD additionally identifies a large number of tissue-specific drug targets (Figure 4B, Supplementary Figure S6 demonstrates the difference of the two distributions). This observation is very likely due to the fact that universal proteins are often involved in certain diseases such as cancer that occur in different tissues, while other diseases are associated with very tissue-specific proteins and processes (see Supplementary Table T1).

Tissue-Specific Genes Have Fewer Transcript Isoforms than Universal Ones. We investigated the number of transcript isoforms encoded by the different genes and additionally combined the results with gene expression data to find out whether the number of transcripts depends on the gene expression strength. When using PAD, tissue-specific genes encode fewer transcript isoforms on average than universally expressed genes, and the more tissues a gene is expressed in, the more transcript isoforms it encodes (Figure 5). We find this trend on a genome-wide basis as well as for the subset of genes encoding for proteins involved in the protein interaction network. Interestingly, for PKD, we find a high number of transcript isoforms from tissue-specific genes that are overexpressed in exactly one tissue. This results from genes that are classified as widely expressed by PAD, but are overexpressed in a single tissue and accordingly classified as tissue-specific by PKD. The comparatively high number of transcript isoforms for such tissue-specific genes suggests that specific isoforms of the overexpressed genes may be essential in the respective tissue. It may also be that such genes are actually expressed in multiple tissues as detected by PAD and encode a variety of transcript isoforms, which allow them to adapt to different tissue environments. Here, different transcript isoforms may contain different functional motifs, for instance, miRNA binding sites or protein interaction domains.

The correlation between the expression level of the genes and the transcript isoforms is low for both PAD and PKD, especially when considering only genes encoding proteins involved in the protein interaction network (Pearson correlation 0.07 (PAD) and 0.04 (PKD) in contrast to 0.51 (PAD) and 0.27 (PKD) for all genes; Supplementary Figure S7). This suggests that the detection of transcript isoforms does not depend on transcript abundance. We also suppose that the lack of correlation can be due to the fact that the used Ensembl database stores both experimentally verified and computationally derived transcript isoforms. Nevertheless, our results indicate that alternative splicing might be an important mechanism for protein isoforms to function in different environments and to enlarge the repertoire of available interaction partners.

Furthermore, we analyzed the interaction degrees in the protein network. We examined all proteins expressed in a certain number of tissues, computed their expressed-interaction degrees according to the number of expressed interaction partners and compared them to the maximal-interaction degrees as given by the static interaction network. With respect to both PAD and PKD, we see the same increase of the interaction degrees with of the number of transcript isoforms (Spearman correlation coefficients for expressed-interaction degree are 0.98 (PAD) and 0.84 (PKD), for maximal-interaction degree 0.98 (PAD) and 0.86 (PKD), Supplementary Figure S8), an observation that is in agreement with the study by Bossi and Lehner.[2] This suggests that alternative splicing is an important mechanism for producing a greater number of protein isoforms from universally expressed genes for functionally diverse interactions in different tissues.

Combining PAD and PKD Results. While PAD identifies tissue-specific proteins according to their RPKM values, PKD points to overexpression in certain tissues. This is exemplified further by means of the STAT1 protein, a well-known signal transducer and activator of transcription.[27] It is involved in 76 protein interactions in our protein interaction network. According to PAD, STAT1 is universally expressed with an average RPKM value of 19.7, while PKD classifies STAT1 as tissue-specific in the cancer cell line MCF7, since the expression in MCF7 is about 6-fold higher than in the other tissues and cell lines. This comparison



	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10-11	11-12	12-13	13-14	14-15
P/A	182	148	52	68	89	116	12	48	86	64	62	49	41	29	237
Peak	320	161	34	62	75	102	6	42	79	44	53	44	13	19	229

	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10-11	11-12	12-13	13-14	14-15
P/A	44	42	42	32	18	24	30	32	35	28	46	50	123	53	427
Peak	289	43	18	15	11	22	13	23	23	21	22	36	22	52	416

Figure 4. Tissue specificity of targets of FDA-approved and experimental drugs. (A) Number of tissues in which protein targets of FDA-approved drugs are expressed. (B) Number of tissues in which targets of experimental drugs are expressed. Dark blue bars show the results using PAD, light blue bars the results using PKD.

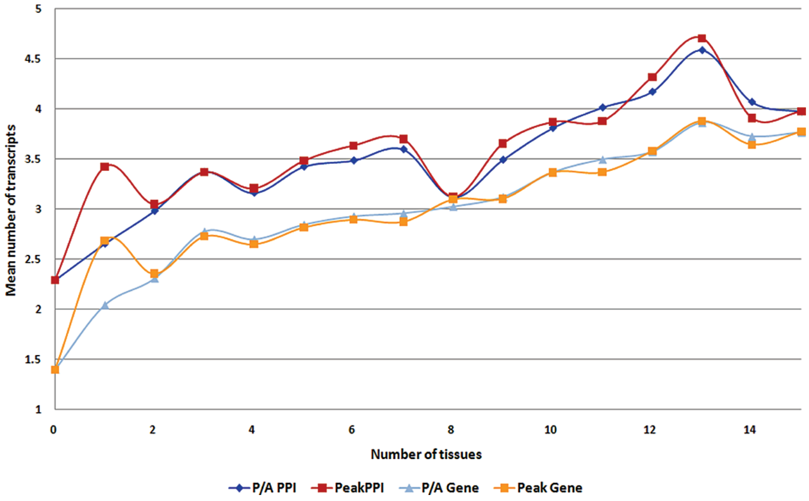


Figure 5. Tissue specificity and alternative splicing. The plot shows the average number of transcript isoforms produced in the respective number of tissues. The dark blue and dark red curves depict the average numbers for all genes involved in the protein interaction network, and the light blue and light red curves the average numbers for all genes. Blue curves represent gene classifications according to PAD and red curves according to PKD.

demonstrates that applying PAD and PKD in combination can provide additional insights. Here, it suggests that STAT1 is universally required for signal transduction and transcription activation and that the specific overexpression of STAT1 in MCF7 is characteristic of this cell line.

Part II: Protein Domains

In the second part of this work, we investigated functional characteristics of protein domains with respect to their tissue-specific occurrence. The following analyses are not restricted to domains occurring in the interacting proteins of the network

Table 4. GO Term Enrichments for Tissue-Specific Protein Domains[a]

GO − Molecular Function (PAD)	*p*-value
receptor binding	1.10e^{-9}
cytokine receptor binding	4.66e^{-8}
growth factor receptor binding	1.53e^{-7}
hormone activity	1.16e^{-5}
growth factor activity	1.33e^{-5}
GO − Molecular Function (PKD)	***p*-value**
receptor binding	2.00e^{-8}
growth factor receptor binding	1.29e^{-7}
cytokine receptor binding	3.22e^{-7}
hormone activity	7.66e^{-6}
growth factor activity	4.01e^{-4}

[a] The most significant GO term enrichments for molecular function based on PAD and PKD are listed together with their *p*-values. The enrichments are computed using topGO.

analyzed in the first part above. Instead, we considered all domains contained in human proteins to discover additional tissue-specific domain functions.

Many Protein Domains Are Neither Tissue-Specific nor Universal. We performed a proteome-wide analysis of Pfam domains to identify universal and tissue-specific domain families. We find many domains to be universally expressed according to both PAD and PKD (1527 (∼54%) for PAD; 1428 (∼50%) for PKD). However, in case of both PAD and PKD, our results also identify a remarkably large number of domains that are neither universal nor tissue-specific (1209 (∼43%) for PAD; 1204, (∼42%) for PKD). Only 104 (PAD) and 204 (PKD) domains are tissue-specific.

64 (∼62%) of the 104 domains that are tissue-specific according to PAD occur in proteins contained in the interaction network. The remaining 40 domains (∼38%), however, do not occur in any of the network proteins. In comparison, 147 (∼72%) of the 204 tissue-specific domains found with PKD are contained in proteins of the interaction network, while only 57 (∼28%) are not.

The numbers of universal domains significantly differ from these proportions. In case of PAD, 1265 domains (∼83%) are contained in interacting proteins and only 262 (∼17%) not (two-tailed *p*-value 0.001, Fisher's exact test). Similarly, in case of PKD, 1175 domains (∼82%) are included in interacting proteins, while 253 (∼18%) are not (two-tailed *p*-value 0.129, Fisher's exact test). To sum up, the results from both PAD and PKD suggest that tissue-specific domains are not necessarily involved in protein−protein interactions, but might also fulfill other biological functions such as DNA-binding.

Tissue-Specific Domains Are Often Receptor-Binding or DNA-Related. We examined the GO functions of all Pfam domains classified as tissue-specific. The GO term enrichment analysis shows a very similar outcome for both PAD and PKD. The results for molecular function reveal that tissue-specific domains are highly enriched in receptor binding functions (Table 4). Our analysis also confirms the previous observation that growth factor binding domains are very tissue-specific.[1] Furthermore, we find several other less pronounced enrichments, depending on whether PAD and PKD is used, such as symporters, DNA binding, amino acid binding, lipid binding, and enzymatic activities (see Supplementary Figures S9 and S10 for complete GO enrichment graphs; Supplementary Table T2 describes all tissue-specific domains in detail). When computing the GO enrichment for cellular component, we always find DNA- and chromosome-related terms to be enriched in addition to extracellular region (Table 5, see Supplementary Figures S11 and S12 for complete GO enrichment graphs), which is in agreement with previous studies.[1] Apparently, many tissue-specific domains play an important role in the nucleus and are probably responsible for transcriptional control. In brief, many tissue-specific proteins form either protein−protein interactions or protein−DNA interactions, the latter of which are not represented by the protein interaction network used in the first part of this work.

Table 5. GO Term Enrichments for Tissue-Specific Protein Domains[a]

GO − Cellular Component (PAD)	*p*-value
extracellular region	3.28e^{-17}
protein−DNA complex	2.58e^{-5}
chromatin	2.58e^{-5}
chromosomal part	9.58e^{-4}
chromosome	1.44e^{-3}
GO − Cellular Component (PKD)	***p*-value**
extracellular region	<1e-20
protein−DNA complex	0.000183
chromatin	0.000183
chromosomal part	0.008503
chromosome	0.016816

[a] The most significant GO term enrichments for cellular component based on PAD and PKD are listed together with their *p*-values. The enrichments are computed using topGO.

Part III: Protein Complexes

The last part of this work concentrates on multimeric protein complexes as well as tissue-specific proteins that might control the formation of a complex in a given tissue. In particular, we study the tissue specificity of protein complexes and their assembly.

Most Protein Complexes Are Universal. To investigate the tissue occurrence of protein complexes, we mapped the gene expression data from RNA-sequencing results onto 648 known protein complexes. Surprisingly, we find a large number of universal complexes (∼58% and ∼51% according to PAD and PKD, respectively), that is, complexes that are fully expressed in more than 13 tissues and cell lines (Supplementary Figure S13). Comparatively few of them are highly tissue-specific (∼4% and ∼21% according to PAD and PKD, respectively), and the remaining complexes are fully expressed in a medium number (3−13) of tissues and cell lines. The complexes included in our study have a minimum size of 3 and a maximum size of 18, but the size is not correlated to the completeness of the expressed complex (Pearson correlation coefficient −0.02 for PAD and −0.19 for PKD; Supplementary Figure S14). For example, the largest complex in our data set, the HCF-1 complex,[28] is fully expressed in all 15 tissues and cell lines according to both PAD and PKD. HCF-1 acts as a transcriptional regulator, and our data suggest that this complex is universally required.

According to both PAD and PKD, the highest number of fully expressed complexes is found in testis, while the lowest number

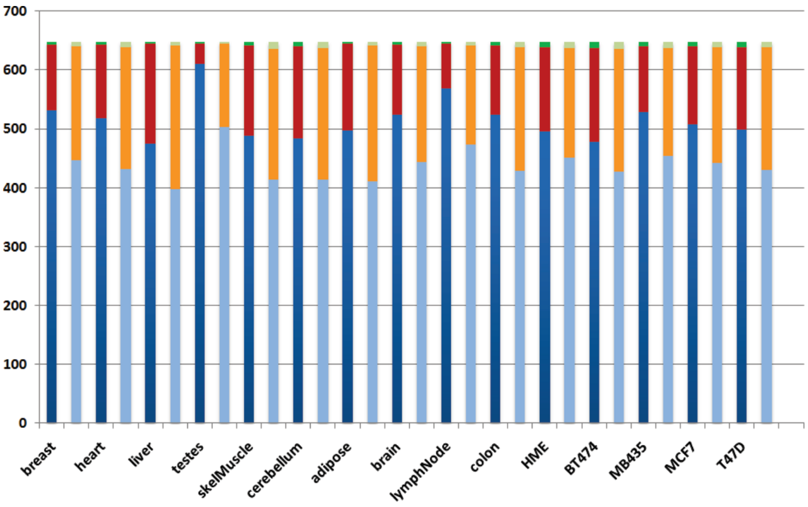


Figure 6. Tissue specificity of protein complexes. Distribution of fully expressed, partially expressed, and absent complexes in the respective tissues. The dark-colored bars represent the results according to PAD, the light-colored bars according to PKD. The number of fully expressed protein complexes in the respective tissue is depicted in blue, the number of partially expressed complexes in red, and the absent complexes in green.

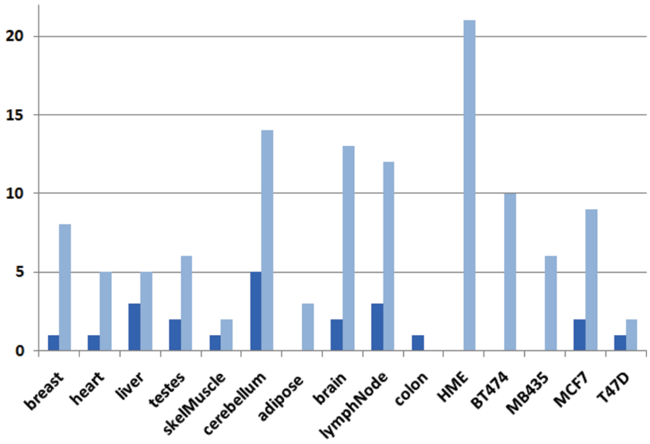


Figure 7. Occurrence of tissue-specific complexes in different tissues and cell lines. The histogram presents the number of protein complexes occurring specifically in the respective tissues and cell lines. The dark blue bars show the results according to PAD, the light blue bars according to PKD.

occurs in liver (Figure 6). In particular, our results also indicate that the complete absence of a complex is very rare among all tissues and cell lines because at least some parts of a protein complex are usually expressed. This supports the hypothesis that cells always maintain partial complexes, which are activated by expressing the missing proteins at appropriate time points.[29] It also supports the notion of core complexes and attachment proteins, where tissue-specific attachment proteins can alter the function of a complex.[30] Another explanation might be that proteins observed in partially expressed complexes perform multiple biological functions in cells and are needed even in the absence of the complete complex.

Few Members Often Determine the Tissue Specificity of Protein Complexes. We identified 28 and 139 tissue-specific protein complexes in our data when applying PAD and PKD, respectively (Figure 7; Supplementary Table T3). Interestingly, the results vary considerably for the two definitions. Using PAD yields the most tissue-specific complexes in cerebellum and none or very few in the cell lines, while PKD gives the highest number of tissue-specific complexes for the HME cell line and slightly less for cerebellum. In contrast to PAD, PKD also identifies a high number of tissue-specific complexes in most of the other cell lines. This discrepancy in the results suggests that the cell lines, which are all cancer cell lines except for HME, contain up-regulated genes compared to normal tissues, which is indeed characteristic of cancer cells. PAD does not allow for the detection of overexpressed genes, and in this case, PKD can thus help to discover abnormal overexpression of otherwise universal complexes.

Only 17 of the 28 tissue-specific complexes identified by PAD are fully expressed in at least one of the tissues and cell lines. According to PKD, 113 of the 139 tissue-specific complexes are fully expressed in some tissue or cell line. One interesting observation is that 17 of the 28 complexes identified using PAD (12 of the 17 fully expressed ones) and 123 of the 139 complexes identified using PKD (106 of the 113 fully expressed ones) include universal proteins. On the basis of PAD, 25 of the complexes (14 of the 17 fully expressed ones) involve one or more tissue-specific protein, which may be important for regulating the assembly and functioning of the complex. Similar results are found by PKD with 138 of the complexes (112 of the 113 expressed ones) containing tissue-specific proteins. Interestingly, though many of the co-complexed proteins are universal, the formation of the complexes appears to be controlled by very few tissue-specific proteins or even only a single protein.

Tissue-Specificity of a SNARE Complex. SNARE complexes are essential for the exocytosis of transport vesicles by mediating the fusion of vesicles and the membrane. Many variations of SNARE complexes exist, and one particular SNARE complex (CORUM identifier 1137) in our study is known to be involved in synaptic transport.[31] This complex consists of 7 proteins. According to the PAD results, 3 of them are universally expressed, 3 are neither tissue-specific nor universal, and 1 protein, Complexin-4, is expressed in cerebellum, but no other tissues and cell lines in our study. When applying PKD, we find 3 of the proteins to be universal and 1 to be neither universal nor tissue-specific. In this case, the three proteins Complexin-4, Complexin-3, and SNAP-25 are tissue-specific. Depending on the used definition of tissue specificity, it appears that either

Complexin-4 is the sole protein that controls the tissue-specific assembly of this complex or Complexin-3 and SNAP-25 work together. Strikingly, Complexin-3 and SNAP-25 are found to be overexpressed in cerebellum only. Functional annotations of the proteins suggest that SNAP-25 is contained in the SNARE core complex, while Complexin-3 and Complexin-4 regulate late steps of the vesicle exocytosis. The overexpression of Complexin-3 and Complexin-4 according to PKD could thus be an indicator for a temporal aspect in the tissue-specific complex formation as suggested in general by de Lichtenberg et al.[29] This example demonstrates that combining results obtained from both definitions of tissue specificity helps identifying tissue-specific as well as time-specific proteins.

Transcriptional Regulation Achieved by a Single Protein. The NCOR-SIN3-HDAC-HESX1 complex (CORUM identifier 3167) functions as transcriptional regulator. According to both PAD and PKD, 5 of the 6 co-complexed proteins are universally expressed, while HESX1 is the only protein expressed in a highly tissue-specific manner in pituitary organogenesis.[32] Correspondingly, HESX1 is not found expressed in any of the samples studied by us. HESX1 directs its transcriptional regulation complex to promoter regions that have to be enhanced or silenced at a particular developmental stage. Strikingly, we detected several subcomplexes of NCOR-SIN3-HDAC-HESX1 that are widely expressed in all tissues and cell lines. This suggests that these subcomplexes act as universal transcriptional regulators, and the presence of one additional protein transforms the universal complex into a complex with very specific function. This example points to the possibility that several transcriptional complexes, both the tissue-specific one containing the HESX1 protein and the universally expressed subcomplexes, are all present and required in pituitary organogenesis. Alternatively, it is possible that the complex is specifically formed to associate with HESX1 during pituitary organogenesis. This would suggest that HESX1 is a highly tissue-specific attachment protein, which alters the function of the protein complex, while the rest constitutes the complex core.

CONCLUSIONS

Next-generation RNA-sequencing is able to measure transcript expression even at low levels. Our results based on these expression data of high quality indicate that substantially fewer protein interactions, protein domains, and protein complexes are responsible for tissue specificity than estimated in previous microarray-based studies. Some tissue-specific functions identified by us agree well with former findings, and our results are more informative and accurate due to the drastically increased detection sensitivity of RNA-sequencing. In particular, we found a remarkably low number of protein interactions to be tissue-specific, many of which are involved in transporter or receptor-activated signaling processes. As with previous studies, our findings rely on the currently available biological data sets, and many protein interactions in human cells are still unknown.[33]

Furthermore, we observed a considerably increased number of transcript isoforms for universally expressed genes. This suggests that the encoded protein isoforms are necessary for different environments and increase the number of possible interactions. We also found universal domains to form protein−protein interactions more frequently than tissue-specific domains, the latter of which are often involved in binding functions such as interactions with receptors or DNA. Therefore, receptor activation and transcriptional regulation seem to be two important factors, besides alternative splicing, for tissue specificity. Moreover, our results suggest that many known protein complexes are widely expressed regardless of their size, and their tissue-specific assembly is often controlled by few tissue-specific proteins.

ASSOCIATED CONTENT

Supporting Information

Supplementary Figure S1, histogram of the number of tissues in which the protein interactions occur. Supplementary Figures S2−S5, GO term enrichment graphs for tissue-specific protein interactions. Supplementary Figure S6, Q-Q plot comparing the distributions of FDA-approved and experimental drugs. Supplementary Figure S7, plot of the average gene expression levels in the respective number of tissues. Supplementary Figure S8, plot of the average interaction degrees in the respective number of tissues. Supplementary Figures S9−S12, GO term enrichment graphs for tissue-specific Pfam domains. Supplementary Figure S13, histogram of the number of protein complexes fully expressed in the respective number of tissues. Supplementary Figure S14, plot of the protein complex size vs. the average number of tissues. Supplementary Table T1, OMIM diseases and the average number of tissues for the respective disease genes. Supplementary Table T2, tissue-specific protein domains. Supplementary Table T3, most tissue-specific protein complexes. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: mario.albrecht@mpi-inf.mpg.de. Phone: +49-681-9325-327. Fax: +49-681-9325-399.

ACKNOWLEDGMENT

Part of this study was financially supported by the German National Genome Research Network (NGFN) and by the German Research Foundation (DFG), contract number KFO 129/1-2. The work was conducted in the context of the DFG-funded Cluster of Excellence for Multimodal Computing and Interaction.

REFERENCES

(1) Lehner, B.; Fraser, A. G. Protein domains enriched in mammalian tissue-specific or widely expressed genes. *Trends Genet.* **2004**, *20*, 468–472.

(2) Bossi, A.; Lehner, B. Tissue specificity and the human protein interaction network. *Mol. Syst. Biol.* **2009**, *5*, 260.

(3) Lage, K.; Hansen, N. T.; Karlberg, E. O.; Eklund, A. C.; Roque, F. S.; Donahoe, P. K.; Szallasi, Z.; Jensen, T. S.; Brunak, S. A large-scale analysis of tissue-specific pathology and gene expression of human disease genes and complexes. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 20870–20875.

(4) Tu, Z.; Wang, L.; Xu, M.; Zhou, X.; Chen, T.; Sun, F. Further understanding human disease genes by comparing with housekeeping genes and other genes. *BMC Genomics* **2006**, *7*, 31.

(5) Eisenberg, E.; Levanon, E. Y. Human housekeeping genes are compact. *Trends Genet.* **2003**, *19*, 362–365.

(6) She, X.; Rohl, C. A.; Castle, J. C.; Kulkarni, A. V.; Johnson, J. M.; Chen, R. Definition, conservation and epigenetics of housekeeping and tissue-enriched genes. *BMC Genomics* **2009**, *10*, 269.

(7) Farre, D.; Bellora, N.; Mularoni, L.; Messeguer, X.; Alba, M. M. Housekeeping genes tend to show reduced upstream sequence conservation. *Genome Biol.* **2007**, *8*, R140.

(8) Su, A. I.; Wiltshire, T.; Batalov, S.; Lapp, H.; Ching, K. A.; Block, D.; Zhang, J.; Soden, R.; Hayakawa, M.; Kreiman, G.; Cooke, M. P.; Walker, J. R.; Hogenesch, J. B. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6062–6067.

(9) Zhu, J.; He, F.; Song, S.; Wang, J.; Yu, J. How many human genes can be defined as housekeeping with current expression data?. *BMC Genomics* **2008**, *9*, 172.

(10) Emig, D.; Kacprowski, T.; Albrecht, M. Measuring and analyzing tissue specificity of human genes and protein complexes. In *Proceedings of the 7th International Workshop on Computational Systems Biology (WCSB)*, Luxembourg, June 16−18, 2010; Vol. 51, pp 27−30.

(11) Ramskold, D.; Wang, E. T.; Burge, C. B.; Sandberg, R. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Comput. Biol.* **2009**, *5*, e1000598.

(12) Vasmatzis, G.; Klee, E. W.; Kube, D. M.; Therneau, T. M.; Kosari, F. Quantitating tissue specificity of human genes to facilitate biomarker discovery. *Bioinformatics* **2007**, *23*, 1348–1355.

(13) Hubbard, T. J.; Aken, B. L.; Ayling, S.; Ballester, B.; Beal, K.; Bragin, E.; Brent, S.; Chen, Y.; Clapham, P.; Clarke, L.; Coates, G.; Fairley, S.; Fitzgerald, S.; Fernandez-Banet, J.; Gordon, L.; Graf, S.; Haider, S.; Hammond, M.; Holland, R.; Howe, K.; Jenkinson, A.; Johnson, N.; Kahari, A.; Keefe, D.; Keenan, S.; Kinsella, R.; Kokocinski, F.; Kulesha, E.; Lawson, D.; Longden, I.; Megy, K.; Meidl, P.; Overduin, B.; Parker, A.; Pritchard, B.; Rios, D.; Schuster, M.; Slater, G.; Smedley, D.; Spooner, W.; Spudich, G.; Trevanion, S.; Vilella, A.; Vogel, J.; White, S.; Wilder, S.; Zadissa, A.; Birney, E.; Cunningham, F.; Curwen, V.; Durbin, R.; Fernandez-Suarez, X. M.; Herrero, J.; Kasprzyk, A.; Proctor, G.; Smith, J.; Searle, S.; Flicek, P. Ensembl 2009. *Nucleic Acids Res.* **2009**, *37*, D690–697.

(14) Finn, R. D.; Tate, J.; Mistry, J.; Coggill, P. C.; Sammut, S. J.; Hotz, H. R.; Ceric, G.; Forslund, K.; Eddy, S. R.; Sonnhammer, E. L.; Bateman, A. The Pfam protein families database. *Nucleic Acids Res.* **2008**, *36*, D281–288.

(15) Wang, E. T.; Sandberg, R.; Luo, S.; Khrebtukova, I.; Zhang, L.; Mayr, C.; Kingsmore, S. F.; Schroth, G. P.; Burge, C. B. Alternative isoform regulation in human tissue transcriptomes. *Nature* **2008**, *456*, 470–476.

(16) Mortazavi, A.; Williams, B. A.; McCue, K.; Schaeffer, L.; Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **2008**, *5*, 621–628.

(17) Yildirim, M. A.; Goh, K. I.; Cusick, M. E.; Barabasi, A. L.; Vidal, M. Drug-target network. *Nat. Biotechnol.* **2007**, *25*, 1119–1126.

(18) Winter, E. E.; Goodstadt, L.; Ponting, C. P. Elevated rates of protein secretion, evolution, and disease among tissue-specific genes. *Genome Res.* **2004**, *14*, 54–61.

(19) Maier, T.; Guell, M.; Serrano, L. Correlation of mRNA and protein in complex biological samples. *FEBS Lett.* **2009**, *583*, 3966–3973.

(20) Vogel, C.; Abreu Rde, S.; Ko, D.; Le, S. Y.; Shapiro, B. A.; Burns, S. C.; Sandhu, D.; Boutz, D. R.; Marcotte, E. M.; Penalva, L. O. Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. *Mol. Syst. Biol.* **2010**, *6*, 400.

(21) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(22) Ruepp, A.; Brauner, B.; Dunger-Kaltenbach, I.; Frishman, G.; Montrone, C.; Stransky, M.; Waegele, B.; Schmidt, T.; Doudieu, O. N.; Stumpflen, V.; Mewes, H. W. CORUM: the comprehensive resource of mammalian protein complexes. *Nucleic Acids Res.* **2008**, *36*, D646–650.

(23) Smedley, D.; Haider, S.; Ballester, B.; Holland, R.; London, D.; Thorisson, G.; Kasprzyk, A. BioMart—biological queries made easy. *BMC Genomics* **2009**, *10*, 22.

(24) Eden, E.; Navon, R.; Steinfeld, I.; Lipson, D.; Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinf.* **2009**, *10*, 48.

(25) Alexa, A.; Rahnenfuhrer, J.; Lengauer, T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **2006**, *22*, 1600–1607.

(26) Chambers, A. F. MDA-MB-435 and M14 cell lines: identical but not M14 melanoma?. *Cancer Res.* **2009**, *69*, 5292–5293.

(27) Liu, B.; Liao, J.; Rao, X.; Kushner, S. A.; Chung, C. D.; Chang, D. D.; Shuai, K. Inhibition of Stat1-mediated gene activation by PIAS1. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 10626–10631.

(28) Wysocka, J.; Myers, M. P.; Laherty, C. D.; Eisenman, R. N.; Herr, W. Human Sin3 deacetylase and trithorax-related Set1/Ash2 histone H3-K4 methyltransferase are tethered together selectively by the cell-proliferation factor HCF-1. *Genes Dev.* **2003**, *17*, 896–911.

(29) de Lichtenberg, U.; Jensen, L. J.; Brunak, S.; Bork, P. Dynamic complex formation during the yeast cell cycle. *Science* **2005**, *307*, 724–727.

(30) Gavin, A. C.; Aloy, P.; Grandi, P.; Krause, R.; Boesche, M.; Marzioch, M.; Rau, C.; Jensen, L. J.; Bastuck, S.; Dumpelfeld, B.; Edelmann, A.; Heurtier, M. A.; Hoffman, V.; Hoefert, C.; Klein, K.; Hudak, M.; Michon, A. M.; Schelder, M.; Schirle, M.; Remor, M.; Rudi, T.; Hooper, S.; Bauer, A.; Bouwmeester, T.; Casari, G.; Drewes, G.; Neubauer, G.; Rick, J. M.; Kuster, B.; Bork, P.; Russell, R. B.; Superti-Furga, G. Proteome survey reveals modularity of the yeast cell machinery. *Nature* **2006**, *440*, 631–636.

(31) Reim, K.; Wegmeyer, H.; Brandstatter, J. H.; Xue, M.; Rosenmund, C.; Dresbach, T.; Hofmann, K.; Brose, N. Structurally and functionally unique complexins at retinal ribbon synapses. *J. Cell Biol.* **2005**, *169*, 669–680.

(32) Dasen, J. S.; Barbera, J. P.; Herman, T. S.; Connell, S. O.; Olson, L.; Ju, B.; Tollkuhn, J.; Baek, S. H.; Rose, D. W.; Rosenfeld, M. G. Temporal regulation of a paired-like homeodomain repressor/TLE corepressor complex and a related activator is required for pituitary organogenesis. *Genes Dev.* **2001**, *15*, 3193–3207.

(33) Venkatesan, K.; Rual, J. F.; Vazquez, A.; Stelzl, U.; Lemmens, I.; Hirozane-Kishikawa, T.; Hao, T.; Zenkner, M.; Xin, X.; Goh, K. I.; Yildirim, M. A.; Simonis, N.; Heinzmann, K.; Gebreab, F.; Sahalie, J. M.; Cevik, S.; Simon, C.; de Smet, A. S.; Dann, E.; Smolyar, A.; Vinayagam, A.; Yu, H.; Szeto, D.; Borick, H.; Dricot, A.; Klitgord, N.; Murray, R. R.; Lin, C.; Lalowski, M.; Timm, J.; Rau, K.; Boone, C.; Braun, P.; Cusick, M. E.; Roth, F. P.; Hill, D. E.; Tavernier, J.; Wanker, E. E.; Barabasi, A. L.; Vidal, M. An empirical framework for binary interactome mapping. *Nat. Methods* **2009**, *6*, 83–90.