Nucleic Acids Research Advance Access published July 16, 2014

Nucleic Acids Research, 2014 **1**
doi: 10.1093/nar/gku532

Physiological state co-regulates thousands of mammalian mRNA splicing events at tandem splice sites and alternative exons

Karol Szafranski[1,*], Claudia Fritsch[1,2], Frank Schumann[1,2], Lisa Siebel[1], Rileen Sinha[1], Jochen Hampe[3], Michael Hiller[4], Christoph Englert[1], Klaus Huse[1] and Matthias Platzer[1]

[1]Fritz Lipmann Institute–Institute for Age Research, Beutenbergstr. 11, 07745 Jena, Germany, [2]Department of General Internal Medicine, Christian-Albrechts-University, Schittenhelmstrasse 12, 24105 Kiel, Germany, [3]Medical Department I, University Hospital, Technical University Dresden, Fetscherstr. 74, 01307 Dresden, Germany and [4]Max Planck Institute of Molecular Cell Biology and Genetics & Max Planck Institute for the Physics of Complex Systems, Pfotenhauerstr. 108, 01307 Dresden, Germany

Received August 1, 2013; Revised May 19, 2014; Accepted May 29, 2014

ABSTRACT

Thousands of tandem alternative splice sites (TASS) give rise to mRNA insertion/deletion variants with small size differences. Recent work has concentrated on the question of biological relevance in general, and the physiological regulation of TASS in particular. We have quantitatively studied 11 representative TASS cases in comparison to one mutually exclusive exon case and two cassette exons (CEs) using a panel of human and mouse tissues, as well as cultured cell lines. Tissues show small but significant differences in TASS isoform ratios, with a variance 4- to 20-fold lower than seen for CEs. Remarkably, in cultured cells, all studied alternative splicing (AS) cases showed a cell-density-dependent shift of isoform ratios with similar time series profiles. A respective genome-wide co-regulation of TASS splicing was shown by next-generation mRNA sequencing data. Moreover, data from human and mouse organs indicate that this co-regulation of TASS occurs *in vivo*, with brain showing the strongest difference to other organs. Together, the results indicate a physiological AS regulation mechanism that functions almost independently from the splice site context and sequence.

INTRODUCTION

Pre-mRNA splicing, the removal of intron sequences from pre-mRNA, is conserved in all eukaryotes. This process is non-deterministic since splice sites, the sequence motifs which mark the intron boundaries (5′SS and 3′SS, respectively), are multiple and their recognition is ambiguous. So, splicing often results in alternatively spliced mRNA species of the same gene, and as a result, most mammalian genes give rise to distinct splicing isoforms (1). Protein isoforms that originate from alternative splicing (AS) were shown to display structural and functional differences (2,3). They contribute to cell type-specific phenotypes, and their relative expression may be affected by external stimuli and diseases (2–5). In addition, AS represents an important regulatory step in the gene expression pathway through coupling to surveillance mechanisms like nonsense-mediated decay (6). The mechanism of AS regulation is closely linked to the activity of splicing factors, as documented by a huge body of case studies (reviewed by (2)). Splicing factors bind pre-mRNA in a sequence-specific manner and contact spliceosomal components by protein–protein interactions. It is believed that the combinatorial arrangement of splicing factors in the vicinity of splice sites promotes formation of an active spliceosomal complex which then initiates the splicing reaction (1,2). Apart from constitutive splicing factors which ubiquitously contribute to spliceosome function, the regulatory effects are typically mediated by accessory splicing factors which act in a cell type- or signal-specific manner, thereby affecting splicing of only a subset of introns, under defined circumstances. However, this model of splicing factor-mediated AS regulation may well be specific to particular structural classes of AS, in particular, cassette exons (CEs) which are, by far, the best studied group in this respect (1,2).

Alternative splice sites in close proximity (2–12 nt distance), also called ‘tandem alternative splice sites’ (TASS), make up the second-largest AS class in mammals (7–10).

*To whom correspondence should be addressed. Tel: +49 3641 656254; Fax: +49 3641 656254; Email: szafrans@fli-leibniz.de
Present address:
Rileen Sinha, Memorial Sloan Kettering Cancer Center–Computational Biology Center, 1275 York Avenue, New York, NY 10065, USA.

TASS result in subtle transcript variants, either a long or short isoform depending on which splice site is used, and these mostly translate into protein isoforms with subtle differences of just few amino acids (9,11,12). TASS may occur at the 5′SS (splice donor) or the 3′SS (splice acceptor). The vast majority of cases, however, are 3′ TASS with splice sites only 3 nt apart, also called NAGNAG acceptors (7–9,11). The minor extent of isoform differences may explain why TASS has attracted relatively little attention so far. However, about 15–25% of mammalian genes are affected by TASS, and TASS occur ubiquitously throughout the eukaryote kingdom (9,10,13). Evolutionary patterns suggest that purifying selection acts to maintain TASS AS (14,15). Splice site choice in TASS appears to work according to a 'scanning mechanism' (16): Following the first biochemical step of the splicing reaction, the spliceosome is centered at the branchpoint region of the pre-mRNA, upstream of the polypyrimidine tract and the 3′SS. In search for a splice site sequence YAG (Y: = C,U) the ribozyme scans along the polypyrimidine tract. The first such sequence is likely to function as the acceptor in the second step of the splicing reaction but, another splice site sequence located further downstream may compete with the first one. The outcome of this splice site competition is ruled by their distance and quality of splice site motifs. Following this model, the AS propensity of TASS cases can be predicted with quite high reliability, based on the sequence of the splice site region alone (14,17,18). Moreover, splice site choice predictions on TASS are successful across different animal species and may even be transferred to plants (13,18). Importantly, the mechanism of splice site choice in TASS is different from the mechanism of exon inclusion/skipping of CEs. Thus, also TASS splicing regulation likely works via different mechanisms than has been formulated for CEs (19).

In search for functional relevance of TASS, previous studies on TASS have identified examples with differential isoform ratios across various tissues (11,12,20,21). The term 'regulation' is used in this context to name the implicit mechanisms that lead to such quantitative differences of splice site selection, depending on cell type or environmental conditions. Regulation, in this sense, is not directly linked to functional relevance, but a correlate marker for functional regulatory processes. Regarding functional relevance, an earlier study reported that TASS isoforms of the mouse PAX6 transcription factor show differential DNA target affinities (22). Moreover, the TASS isoforms of human DRPLA were reported to show different subcellular localization (12). Meanwhile, another study examined TASS isoform patterns and found almost no tissue-specific differences in 9 genes assayed in 23 tissues, and a differential isoform pattern of one additional TASS case was explained by the interference of sequence polymorphisms (23). Most recently, a genome-wide study of TASS using Illumina RNA-seq data came to the conclusion that at least 25% of human and murine TASS are regulated in tissue-specific patterns and that these regulated TASS cases are under increased purifying selection (21).

We present a systematic study of TASS in comparison to CE and mutually exclusive exon (MXE) splicing in human and mouse tissues as well as cultured cells in order to clarify the extent and patterns of tissue-specific TASS regulation which has been debated. For isoform quantification we relied on capillary electrophoresis (CE-LIF) separation of fluorescence-labeled reverse transcriptase-polymerase chain reaction (RT-PCR) products, a method which has high accuracy and very good scalability on multiple samples (13,18,23,24). Our results indicate that TASS isoform ratios do vary across different tissues and culture conditions, although with relatively small amplitude compared to CEs and MXE. Cultured cells show very ordered regulation in association with cell density and, interestingly, this regulation affects not only most TASS cases but also different AS classes comprising CEs and MXE. These findings indicate a novel regulatory mechanism which is further validated and characterized by genome-wide analyses including RNA-seq.

MATERIALS AND METHODS

Target tandem cases

TASS cases were compiled as previously described (10,20). Isoform-specific counts of expressed sequence tags (ESTs) were used to determine the average isoform ratios. Gene expression was scored substantial if at least 100 ESTs were found in the GenBank database. Homogeneity of expression was scored according to profiles compiled in the SOURCE database and data from the Gene Expression Omnibus (GEO) database at the National Center for Biotechnology Information (NCBI), including Illumina mRNA-seq.

Tissue samples, cell culture and RNA isolation

Human cDNA was purchased as MTC Panels I and II, Fetal and Cell Lines (Clontech). Human leukocytes, the only of these samples that had live cells extensively processed *in vitro*, were excluded from analysis (Supplementary Figure S2). Instead, blood was taken from two female and two male human donors and collected into PAXgene Blood RNA tubes (Qiagen). Male mice, strain C57BL/6J, either 6–8 weeks or 6–11 months old, were killed by cervical translocation, and tissue samples were isolated and kept in RNA Later (Qiagen) at 4°C. All cell lines were obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen (Braunschweig, Germany). Cells were cultured at 37°C under water-saturated 5% carbon dioxide atmosphere. HL-60 and L-5178-Y cells were cultured in 10 ml RPMI-1640 medium (Gibco) supplemented with 10% fetal calf serum (FCS; PAA Laboratories) and penicillin/streptomycin (Sigma) in filter-capped 50-ml polystyrene flasks (Becton Dickinson). HEK-293E cells were cultured on 10-cm polystyrene dishes (Becton Dickinson) with 10 ml Dulbecco's modified Eagle's medium medium (PAA Laboratories) supplemented with 10% FCS and penicillin/streptomycin. Cells were counted in a Neubauer chamber (Marienfeld, Lauda-Königshafen, Germany) in fields of 1 mm^2 size, in four replicates. Dead cells were identified using Trypan blue staining. For harvest, adherent cells were washed with phosphate buffered saline (PBS) and detached mechanically. Cells were then pelleted at 100 $\times g$ for 10 min, washed with PBS and again pelleted. Prior to RNA isolation, mouse tissue was disrupted in the

homogenization buffer of the RNA extraction protocol using a Tissue Lyser instrument (Qiagen). RNA was isolated using RNeasy Mini kit (Qiagen), performing specialized protocols for brain and muscle tissues as recommended by the manufacturer. RNA from whole blood samples was obtained using the PAXgene Blood RNA kit (Qiagen). Synthesis of cDNA was done using AMV Reverse Transcription kit (Clontech), 1 μg total RNA and random primers according to the manufacturer's instructions.

Quantification of splicing isoforms by CE-LIF

RT-PCR for splice isoform quantification was done using 2 pg poly(A)+ cDNA (human) or 1 μl first-strand cDNA preparation (mouse and cell lines). PCR reactions were set up using Taq BioMix (Bioline) and 10 pmol primers (Metabion; for sequences see Supplementary text 4), one 5'-labeled with 6-carboxyfluorescein (FAM). The cycling conditions were 2 min initial denaturation at 94°C, followed by 40 cycles of 45 s denaturation at 94°C, 50 s annealing at 56°C, 1 min extension at 72°C and a final 30-min extension step at 72°C. The FAM-labeled PCR products were diluted, mixed with formamide (Roth) and GeneScan 500 LIZ (Applied Biosystems), denatured and then separated on an ABI 3730 capillary sequencer (Applied Biosystems) according to the manufacturer's recommendations. The electropherograms were analyzed with the GeneMapper 4.0 software (Applied Biosystems). The fraction of the long isoform (f_{long}) was calculated from the peak areas (a) observed in the electropherogram: $f_{long} = a_{long} / (a_{long} + a_{short})$. From independent triplicate analyses we derived average isoform fractions and standard deviations.

Statistics of differential isoform expression

Tissue-specific isoform ratios were tested by one-way ANOVA using the triplicate measurements from the tissue panels. While the ANOVA P-values describe the signal-to-noise ratio, we describe the cross-tissue variation on an absolute scale using the between-sample standard deviation (SD), corrected by subtracting the standard error of the mean from replicates. Multifactor ANOVA was performed using a standard formula of within/between mean square ratios implemented in in-house software. Tissues that consistently produce extreme isoform ratios were identified using a permutation test. The tissues were ranked according to the isoform ratio, per AS case, and the rank averaged across all cases. Quantiles of the background distribution were obtained from random ranking and subsequent averaging like for the real cases.

Normalization of time series data

Time series data of isoform fractions were normalized for amplitude, resulting in a minimum value of 0.0 and maximum of 1.0. A consensus fit of multiple time series was obtained by computing averages per time point. In addition, each series was allowed to flip the y-axis, mimicking the directional freedom of isoform shift observed in the cell experiments. The best series model, that is the set of y-orientations of the series and its consensus fit, was identified by the least squares method, by summing the deviation squares of all series data against the consensus. A permutation test was used to determine the likelihood of obtaining multiple time series with a given similarity, described by the least squares score of the series consensus model. Random times series were generated by enumeration of random values using a Perl v5.10 script and performing normalization and consensus modeling as done for the real data.

Next-generation sequencing and data analysis

From the GEO database (http://www.ncbi.nlm.nih.gov/gds/) we obtained Illumina RNA-seq data for 12 independent human adult tissue samples and 6 cell lines (1,25). Separately, we constructed a library of potential Δ3 3′ TASS cases using the human genome reference sequence (NCBI Build 36.1) in combination with the RefSeq transcript database as previously described (18). For each of these cases, tiling 32-mer sequence probes, crossing the exon–exon junction with at least 6 nt overlap, were derived. These probes were filtered for perfect matches against the genome sequence and unspecific multigene matches against RefSeq. Perfect matches of the RNA-seq data against the TASS probe library were used to obtain count values for TASS isoforms. For quantitative analysis, we selected TASS cases with substantial isoform count values (≥80 counts and >5% minor isoform, taking all GEO RNA-seq data together), sufficient for reliable estimates for the average isoform ratios. A resulting set of 592 TASS cases was split into 6 bins spanning equally sized ranges of percentage long isoform, 5–20%, 20–35%, 35–50%, 50–65%, 65–80% and 80–95%.

Total RNA from cultured cells, with an RNA integrity number greater or equal to 8.0 (Agilent Technologies), was used for sequencing. Starting with 5 μg RNA, we constructed Illumina TrueSeq RNA-seq libraries according to the manufacturer's protocol. Each of these libraries was sequenced to give 15–24 million single-end readings of 76 nt length. Raw data were deposited in the NCBI Short Read archive under BioProject PRJNA239705. For quantitative analysis, pooled isoform counts were compiled for each bin and cell/tissue sample. In the pooling step, the contribution of single TASS cases to the bin counts, proportional to the gene's expression level, was limited to a maximum of 3%.

RESULTS

Systematic quantification of AS isoforms

In order to obtain a representative picture of TASS versus CE and MXE isoform ratios in native human and murine tissues, we systematically quantified isoforms of 14 AS cases using the CE-LIF method. Replicate assays combined with sample statistics allowed us to infer the measurement accuracy routinely (24). For the experiment we selected 10 human 3′ TASS cases according to following criteria: (i) representation of canonical AG–AG tandems as well as tandems involving non-canonical 3′SS (20), (ii) varying tandem distances, from 3 to 18 nt, (iii) preferably high and ubiquitous mRNA expression and (iv) a minor isoform fraction >10%. The latter two optimality criteria define cases that promise high-resolution quantification results, even for small RNA samples (Supplementary text 1) (24). The chosen AG–AG

tandems occur in the genes *CNOT3* (nt length difference of isoforms: Δ3), *DAXX* (Δ3), *TLE4* (Δ3), *CCNE1* (Δ9) and *RBM39* (Δ18). The chosen non-canonical TG–AG tandems occur in the genes *CNBP* (Δ3), *GNAS* (Δ3), *HNRNPR* (Δ9), *SRRT* (Δ12), *PCBP2* (Δ12) (Supplementary Table S1). Two of these TASS cases have been studied previously for the quantitaties of splicing isoforms, that is *GNAS* (26,27) and *SRRT* (20). For comparison, we included representatives of other AS classes into the study, namely, 5′ TASS in *PUM2* (Δ6), CEs in *GNAS* (Δ45) and *RPS24* (Δ22), as well as MXE in *H2AFY* (Δ9) (Supplementary Table S2) obeying the selection criteria (iii) and (iv). Given the relatively long isoform difference for the CE in *GNAS*, we excluded the possibility that differences in gene expression caused relevant isoform bias in RT-PCR/CE-LIF quantification (Supplementary text 1, Supplementary Figure S1 and Table S3).

Primarily, we studied TASS isoform ratios in 16 human adult and 8 fetal organs/tissues, as well as in 8 human cell lines using CE-LIF (Figure 1, Supplementary Figure S3). The average isoform fraction per TASS case agreed well with values indicated by EST numbers (Fisher's exact test with $\alpha = 0.05$). The TASS isoform data showed a cross-tissue variation larger than the measurement uncertainty (median SD of replicates: 2.2% isoform), statistically significant for 9 of 11 genes (one-way ANOVA with $\alpha = 0.05$). The cross-tissue variation of isoform levels, expressed as tissue-specific SD corrected for measurement uncertainty, was between ±0.3% (*DAXX*) and ±4.3% (*RBM39*) in adult tissues. The most prominent tissue-specific deviation was found for *RBM39* in blood, where the fraction of long isoform was 1.77-fold below the median (21.5% difference). With regard to the variation characteristics, isoform patterns of canonical and non-canonical 3′ TASS cases, as well as the 5′ TASS case were quite similar, and no differences were noted with respect to the dynamic range of isoform ratios. However, overall, the TASS variation was small compared to CE and MXE cases which varied with ±16.0% to ±18.0% isoform fraction (Supplementary Figure S3).

To assess TASS isoform patterns also across species, we additionally analyzed mouse adult tissues. This left seven informative cases for analysis (Supplementary Figure S4). Ten male mice, split into two age groups (either 6–8 weeks or 6–11 months), were analyzed separately and showed high congruence in their tissue-specific TASS pattern (Figure 1B and C, Supplementary Figure S4). In the orthologous mouse genes, 12 of 14 studied AS events were conserved (all except 3′ TASS cases *Ccne1* and *Daxx*). Among these, TASS isoform ratios varied with low percentage across tissues (corrected SD ±0.8% to ±4.6%, median ±2.4%), though, tissue-specific differences were highly significant in all nine cases (one-way ANOVA, with *P*-values 4E-12 or smaller). Isoform ratios of CE and MXE cases were tissue-specific with high significance and, like in human, the cross-tissue isoform variation was much higher than for TASS cases, ±20.1% to ±26.6%. Factor analysis substantiated that the outstanding major determinant of varying TASS isoform ratios was tissue origin in all cases (median R^2 of 0.587; multifactor ANOVA) in contrast to age (median R^2 of 0.013) or replicate number.

Overall, isoform ratios were found quite stable across the examined human samples and TASS cases. While this finding is close to that of other systematic studies (21,23), it is incompatible with reports on a number of other TASS cases to show remarkable tissue-specific patterns of their isoforms (11,12,20). To investigate this discrepancy more directly, we have re-evaluated six of these cases (human *ITGAM*, *BTNL2*, *SMARCA4*, *BRUNOL4*, *NOXO1* and murine *Ccl20*) using the state-of-the-art CE-LIF method. The results indicate tissue-specific differences in only one case (*BRUNOL4*; $P = 0.26$) but do not support the existence of strong tissue-specific differences for any of these TASS isoforms (Supplementary text 2 and Supplementary Figure S5). In all six cases, we observed high measurement variance, at least for some tissue samples. qRT-PCR demonstrated that this variance was due to very low gene expression, which was critical at ≤200 molecules per isoform assay (Supplementary text 2 and Supplementary Figure S5) (24). Likely, previous studies using singlet isoform measurements were mislead by high scatter from very low expressed mRNAs.

AS isoforms are co-regulated in association with cell density

In addition to tissues, two human and one mouse cell lines were tested for splicing isoform ratios. Fist, human leukemia HL-60 cells were incubated in a constant volume of standard media over several days until a plateau of cell density reached (Supplementary Figure S2A). Of these cultures, cell batches were harvested every 24 h for monitoring of the isoform fractions using the CE-LIF method. In order to avoid unphysiological conditions over long cultivation periods, the culture medium was refreshed every 24 h. As cells grew to high density, the TASS isoform ratios shifted between day 2 and day 5 to reach a new plateau (Figure 2B and Supplementary Figure S6). In 9 of the 11 cases, the fraction of the long isoforms decreased, whereas in two cases, *CCNE1* and *RBM39*, it increased (Figure 2B). As seen in tissues, different TASS cases showed different amplitudes of isoform changes, with the cross-sample SDs ranging from ±0.2% (*TLE4*) to ±6.4% (*SRRT*). However, these amplitudes showed weak gene-by-gene correlation in comparison to tissue data or results from other cell experiments (Supplementary Tables S5 and S6). Surprisingly, also the isoform patterns of CE and MXE cases showed clear time-dependent changes, with shift amplitudes of ±2.5% to ±5.5% which where within the range of the TASS cases (*t*-test, $P = 0.19$). One of these showed a decrease of the long isoform while the others showed a decrease. After accounting for case-specific direction and normalizing the amplitude, time series of isoform shifts show a remarkable similarity between all AS cases, far from random (Figure 2C; $P < 1E\text{-}10$; permutation test). In 10 of the 14 cases, the time series show better fit to the sigmoid consensus profile than to an alternative model of individual linear functions (based on least squares). We also checked whether significant changes in mRNA expression levels occurred which might have caused a technical bias in the measured isoform quantities. Expression levels changed up to 4.5-fold between between low and high cell density (*SRRT*; Supplementary

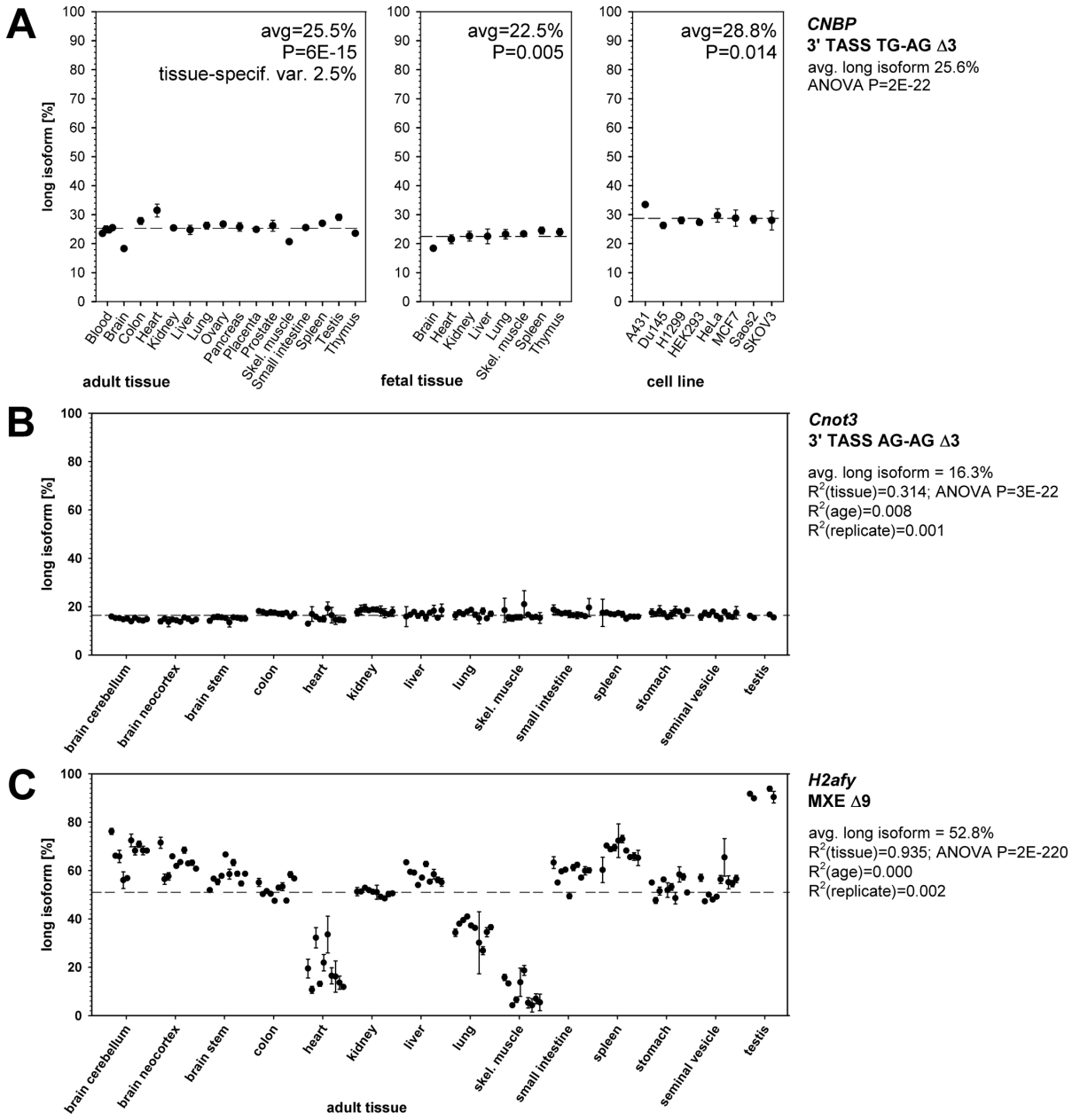


Figure 1. Isoform fractions in different tissues and cell lines. Shown are **(A)** human *CNBP*, **(B)** mouse *Cnot3* and **(C)***H2afy*. Gene name, AS characteristics and statistical measures are given right next to each plot. All other cases are given as Supplementary material (Supplementary Figure S3). Each mouse tissue was analyzed independently in 5 young (left half data block) and 5 old mice (right half). Standard deviations of triplicate measurements are indicated by variance bars. *P* denotes the significance level of overall sample differences (one-way ANOVA).

Table S7), but the direction of expression changes could not consistently explain the shift of isoform ratios.

Since the cell type may have influenced the regulatory patterns observed in leukemia HL-60 cells, we studied human embryonic kidney (HEK-293) cells cultivated under equivalent conditions. These cells showed the same effect of shifted isoform ratios when reaching about 100% confluence (Supplementary Figure S7). Between the two cell lines, the direction of the isoform shifts were identical in 13 of the 14 cases, except *CCNE1* which was ambiguous, and these shift directions were stable case characteristics in all replicate experiments (data not shown). Another experiment was performed with cultures of the mouse lymphoma cell line L-5178-Y. Here, following pre-culture of cells at low density (between 0.25 and 1.0 $*10^6$ cells/ml) for 48 h, cells were reseeded to high density (from 1.0 to 3.0 $*10^6$ cells/ml; Figure 2D). The isoform quantities shifted within 16 h (8 out of the 10 AS cases that could be analyzed), with the direction corresponding to the human orthologous isoforms. Another case shifted within 32 h (*Rps24*), one showed an ambiguous to adverse shift (*Tle4*; Supplementary Figure S9). These results demonstrate that an immediate transition to high cell density provokes isoform shift with faster kinetics than observed in the human cells which underwent

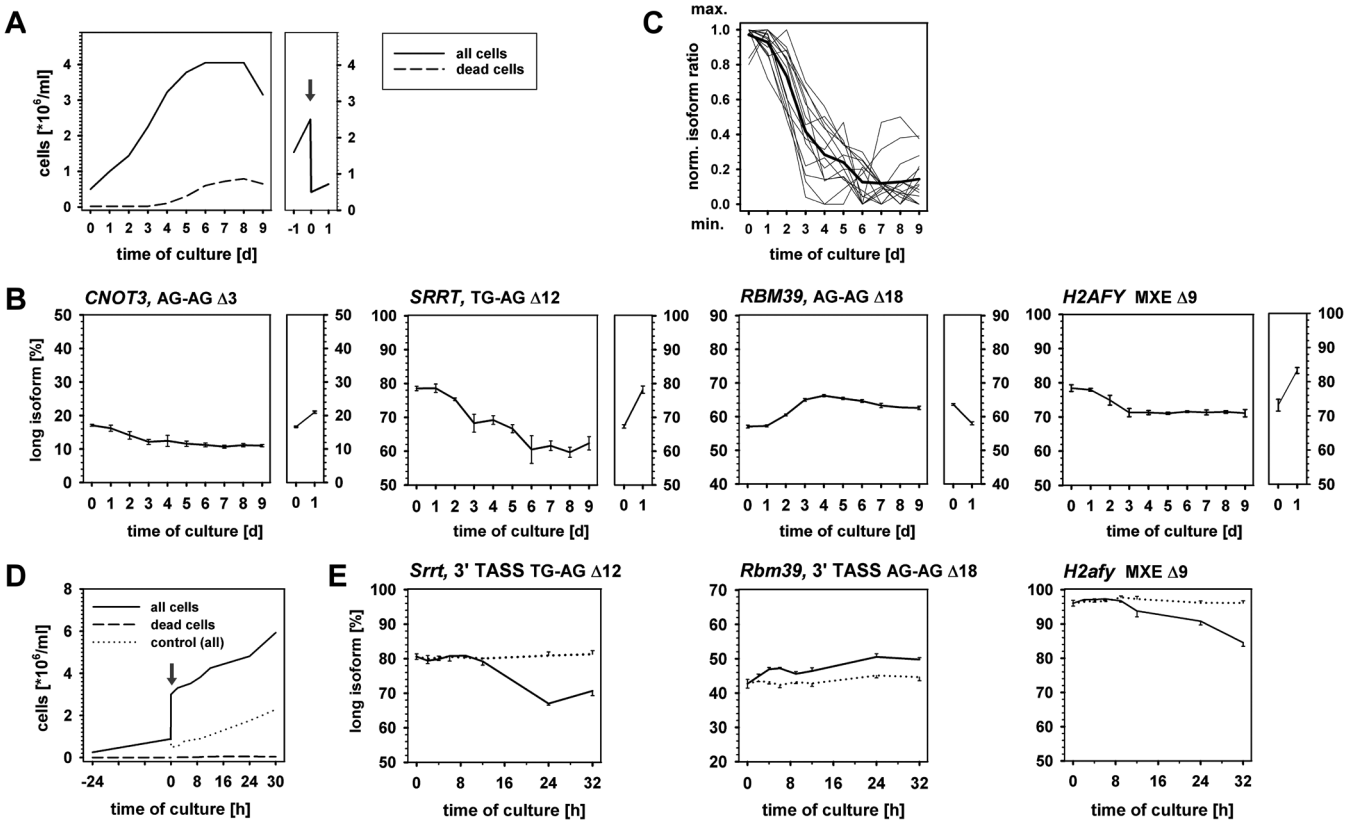


Figure 2. Splicing isoform fractions change in cultured human HL-60 cells and mouse lymphoma L-5178-Y cells depending on cell density. **(A)** Cell density in two HL-60 culture experiments, either cells seeded at low density (left diagram; seeding at day 0), or re-seeded from higher to lower cell density (right diagram; arrow indicates re-seeding at day 0). **(B)** Isoform fractions in the cell culture experiments represented in A, after harvest at indicated time points, per day (d); representative results from three 3′ TASS cases and one MXE (additional cases are represented in Supplementary Figure S6). **(C)** Time series profiles obtained via normalization of the isoform data taken from (B) (normal lines; all TASS cases), pooled to derive a consensus time series (bold line), as described in Materials and Methods. **(D)** Density monitoring of L-5178-Y cells, pre-cultured at low density (24 h indicated), then re-seeded at t_0 (indicated by arrow) to either, high density (plotted with solid line) or low density as a control (dotted line). **(E)** Isoform fractions of 3′ TASS and MXE cases (additional cases represented in Supplementary Figure S9) in the L-5178-Y experiments illustrated in panel D, at harvest 0, 2, 4, 6, 9, 12, 24 and 32 h after re-seeding; dense cells represented by solid line, dilute cells by dotted line. In all panels, SDs from triplicate assays are indicated by variance bars.

a gradual change of cell density over days. Notably, under these kinetics conditions, some cases showed isoform responses with an oscillatory pattern, most prominent in *Rbm39* (Figure 2E), but also in *Cnot3*, *Cnbp* and *Srrt* (Supplementary Figure S9).

Next, we asked if the effects were reversible. Upon dilute reseeding of high-density cultured HL-60 cells, the isoforms shifted back to initial levels within 24 h in all 13 analyzed cases (Figure 2B and Supplementary Figure S6, right diagrams; $P = 0.001$, one-sided binomial test), as well as 6 of 7 tested cases in similarly treated HEK-293 cells (Supplementary Figure S7; $P = 0.07$). This reversal of the isoform ratio was about 80% complete within 24 h, which indicates that the regulatory response is a result of a steady interaction with the culture environment and that cells remained viable in responding to the environmental conditions throughout the experiment. The overall viability of the harvested cells was also demonstrated by trypan blue exclusion (Figure 2A), indicating that the physiological conditions of the cell culture were well tolerated by the cells. In an attempt to characterize the physiological conditions that lead to the TASS isoform shift we performed an experiment in which cell culture medium was not replaced by fresh medium but medium that was conditioned by high-density pre-culture. The data revealed that the TASS splicing shift was accelerated (Supplementary Figure S8), further supporting the idea that isoform ratios change due to a feedback with the environment.

RNA-seq confirms the co-regulatory pattern genome-wide

We wondered if the results from our gene-directed approach could be generalized, because bias might have been introduced by case selection or the chosen quantification method. Next-generation transcriptome sequencing (RNA-seq) allowed us to sample a large number of Δ3 3′ TASS cases, and made a technically independent validation of the CE-LIF quantification possible. Basically, the occurrences of RNA-seq reads can be taken as a quantitative mirror of the mRNA molecules, including different mRNA isoforms of the same gene (21). We performed RNA-seq for culture samples of HL-60 and HEK-293 cells, each at low- and a high-density time points, which were previously analyzed by CE-LIF. By focussing on Δ3 3′ TASS, we chose the largest TASS subgroup with high structural homogeneity, minimizing the contribution of overlaid regulatory ef-

fects including the unexplained direction flipping of isoform shifts. Thus, we screened the data for sequence fragments that are specific for the different Δ3 3′ TASS isoforms expressed in the cell lines (504 and 532 cases, respectively) and obtained measures of their isoform ratios. Although RNA-seq yielded 15–24 million reads per sample, isoform counts greater or equal to 100 were obtained for only few individual TASS cases (7 and 13 cases passing in both samples from the two cell lines, respectively). Only one of these indicated a significant shift of the isoform ratios (*SRSF11*; decrease of long isoform in both cell lines), the others were inconclusive (Supplementary text 3).

To increase the statistical power, we combined counts by pooling the TASS cases into six evenly sized bins of average isoform percentage ranges (total range 5–95%, bin size 15%) and analyzed these bins for changes of isoform ratios when comparing the different cell culture states (Figure 3). This approach gave 832–6416 isoform counts per TASS bin and cell sample, which allowed determination of the isoform ratios with a resolution of 0.43–2.6% SD. For HL-60 cells, all six bins showed decreased long isoforms at high cell density (Figure 3; $P < 0.05$ for four bins, Fisher's exact test), highly consistent with the CE-LIF measurements. Similarly, HEK-293 cells showed decreased long isoforms in all six TASS bins at high cell density (Supplementary Figure S11). Interestingly, in both experiments, the bins with predominantly short or long isoforms (bins 1 and 6, respectively) showed a comparably small shift of isoform quantities. This may indicate that the isoform ratio is a result of fold-changed reaction rates of splice site choice leading to either, the long or the short isoform. Under this kinetic model, the change of isoform ratios (long over short) is 0.79 ± 0.09-fold in HL-60 cells and 0.82 ± 0.07-fold in HEK-293 cells, fairly homogeneous across all bins. However, bin 6 consistently remained an outlier in showing weaker changes than did other bins (0.89-fold in HL-60, 0.95-fold in HEK-293).

3′ TASS isoforms show a co-regulatory pattern *in vivo*

Having shown the simultaneous shift of AS isoform ratios in cultured cells, we sought for evidences of such a co-regulation *in vivo*. CE and MXE clearly differed from TASS in their differentiated, strong tissue-specific isoform patterns. We therefore left these aside and further analyzed the TASS isoform data from the tissue panels and asked if certain tissues consistently produced less or excess long isoforms. In the cell culture experiments, two TASS cases (*CCNE1*, *RMB39*) changed their isoform ratio in a direction opposite to that of the other genes. Taking this case-specific signature into account, we scored tissues for being more alike low- or high-density cultured cells.

A comparison of human adult and fetal tissues revealed that TASS patterns of the latter were more similar to high-density cultured cells, consistent for 9 of the 10 analyzed TASS cases ($P = 0.011$, two-tailed binomial test). Among the 16 adult human tissues, brain or skeletal muscle had isoform ratios most similar to high-density cultured cells, indicated by 7 of the 10 TASS cases (Supplementary Table S4; $P = 0.02$, two-tailed permutation test). When we pooled the data for the different TASS cases, a complete rank list of the adult tissues was obtained (Figure 4A, left). Consistently, brain and skeletal muscle significantly scored toward high-density cultured cells ($\alpha = 0.05$, two-tailed permutation test, $N = 10\,000$). On the other hand, human kidney showed a TASS pattern with significant similarity to low-density cultured cells. Next, when we analyzed tissue rank lists obtained from mouse data (Figure 4A, right), mouse adult brain showed high consistency to human. Three different brain regions significantly scored towards high-density cultured cells (two-tailed permutation test, $N = 10\,000$, $\alpha = 0.05$). Moreover, mouse liver scored significantly toward low-density cultured cells, as did the human liver sample, though with borderline significance. Mouse skeletal muscle, however, unlike the human counterpart, showed no significant trend, and mouse colon tissue scored significantly toward low-density cultured cells, while the corresponding human tissue showed only slight correspondence.

Since the human and mouse tissues showed only partial congruence in their TASS splicing patterns based on CE-LIF, we sought for independent data. Again, RNA-seq data served here as a methodologically independent data set, having the additional advantage to display a large spectrum of TASS cases. A total of 592 Δ3 3′ TASS cases were used to rank the tissues for preferably producing long or short TASS isoforms (Figure 4B). Again, brain tissue (two samples) showed significantly less long TASS isoforms (two-tailed permutation test, $N = 10\,000$, $\alpha = 0.05$). Liver showed significantly increased long TASS isoforms, as was found in human and murine adult tissues using CE-LIF. The RNA-seq data gave no consistent support for an extreme status of skeletal muscle (low long isoforms in human and murine tissue panel), heart muscle (low long isoforms in murine tissue panel) or kidney (high long isoforms in human and murine tissue panels). However, across all experimental TASS data, brain very consistently showed low fractions of long isoforms, like densely cultured cells, and liver consistently showed high fractions of long isoforms, like dilute cell cultures.

DISCUSSION

The aim of this work was to investigate quantitative patterns of TASS isoforms. First, we systematically evaluated 11 TASS cases in 46 different mammal tissues and cell lines, thereby producing the largest-available high-resolution quantitative data set for TASS isoforms. The obtained data resolve disagreements between previous studies: TASS isoforms do neither show constant ratios across tissues (23), nor extreme variations of the ratios (11,12,20). The observed patterns are intermediate: (i) the isoform ratios are significantly non-constant for almost all analyzed TASS cases, but (ii) the dynamic range of varying TASS isoform ratios is relatively small (cross-tissue SD ±4.6%) and (iii) outstanding tissue-specific deviations, more than 1.5-fold from the median isoform ratio, occur for only one of 11 TASS cases in a single out of 24 tissues (1.77-fold in *RBM39* from adult human blood). Furthermore, a previously reported strong tissue-specificity of TASS isoforms from five genes was not confirmed by a thorough re-analysis. For these, our results suggest that previous studies were probably mislead by scatter from low mRNA concentrations and lacking replicates that might have controlled for measure-

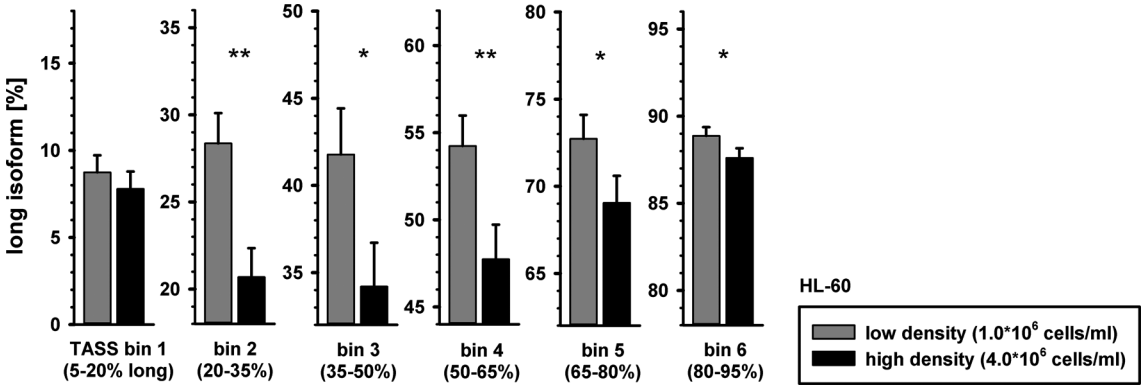
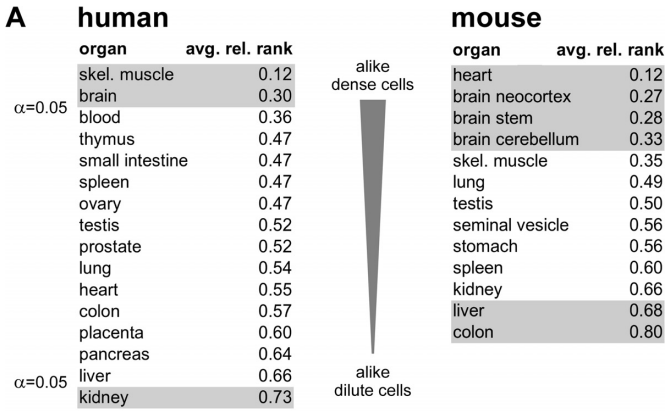


Figure 3. Genome-wide comparison of the isoform ratios of 504 Δ3 TASS cases in low- versus high-density cultured HL-60 cells based on RNA-seq. TASS cases were grouped into six bins according to their average percentage long isoform in human samples (note the different scaling of the subgraph *y*-axes). RNA was isolated from low-density (gray bars; day 1 culture, same experiment as shown in Figure 2) and high-density cultured HL-60 cells (black bars; day 7). Variance bars indicate the amount of variation expected from the Bernoulli distribution, based on the sample size of isoform counts. Asterisks indicate a significant decrease of long isoforms in the densely cultured cells (Fisher's exact test, *significant for $\alpha = 0.05$; **for $\alpha = 0.01$).



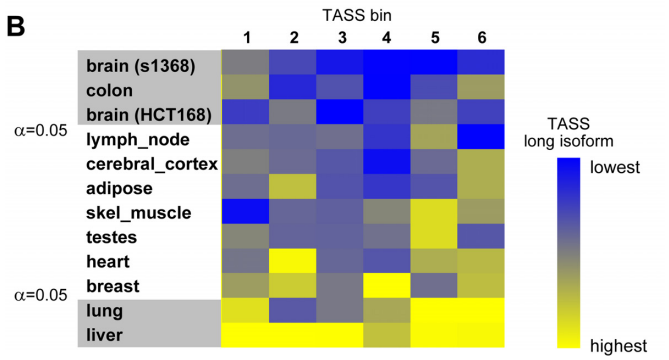


Figure 4. Brain tissue consistently produces TASS isoform ratios much alike high-density cultured cells. **(A)** Isoform ratios of TASS cases were measured in human (left; 10 cases) or mouse adult tissues (right; 7 orthologous cases) using CE-LIF as described. For every TASS case, the tissues were ordered according to their splicing ratio being more alike that of high-density cultured cells or dilute cells. Superimposition of the lists and averaging of the ranking positions (given as 'avg. rel. pos.') yields the lists as shown. The significance threshold α is derived from a permutation test as described in Materials and Methods. **(B)** Isoform ratios of 592 TASS cases, binned as described in Figure 3, were determined from GEO RNA-seq data for 12 human adult tissues. The heat map indicates the relative ranks of long isoform abundance (color legend on the right). Abundance of long TASS isoforms serve as a proxy for the splicing pattern of low-density cultured cells. Superimposition of the tissue lists and averaging yields the list as shown on the left. The significance threshold α is determined as in (A).

ment uncertainty (Supplementary text 2 and Supplementary Figure S5). Thus, our results on TASS isoform expression profiles are representative for these and likely many other cases. The vast majority of tissue-specific isoform patterns may be appropriately described as 'subtle variation', by reusing the term 'subtle AS' that has been proposed to describe the small mRNA isoform differences introduced by NAGNAG acceptor motifs (12). The notion of subtle differences is mostly consistent with the data from a recent genome-wide *in silico* analysis of TASS splicing (21), and it is consistent with the previous notion that TASS splicing is highly predictable using computational approaches, even when based on EST data from very heterogeneous tissue sources (17,18). The current finding of a small dynamic range of TASS isoform ratios is important because, so far, tissue specificity has been explored as a proxy for gene-specific regulatory functions of TASS. In this respect, we conclude that the narrow tissue-specific TASS isoform variation is unlikely to imply a gene-specific regulatory role in most of the cases.

We observed that almost all AS isoform ratios changed in human cell culture with increasing cell density (Figure 2) and did so with similar time series profiles (Figures 2C and 3). The data from human and mouse adult tissues corroborate the view that the *in vivo* cellular state (including cell type) influences the isoform ratio of most analyzed 3′ TASS cases (Figure 4A and B). In particular, different brain tissues consistently produce lower amounts of long isoforms than do other tissues. Liver cells, on the other hand, consistently produce higher amounts of long isoforms. In addition, RNA-seq data confirm these tissue-specific splicing patterns and allow us to generalize this finding to the genome-wide repertoire of Δ3 TASS (Figures 3 and 4C). Most remarkably, representatives of structurally different AS classes, 5′ TASS as well as the unrelated group of CEs, showed isoform responses in the cell system very similar to 3′ TASS with respect to time course and amplitude of isoform change. This extensive co-regulation, despite the structural differences between and within the AS classes, suggests a novel common regulatory mechanism which functions independently of the structural properties.

It appears that TASS splicing displays this mechanism in a pure form, while other AS classes, especially CEs, are additionally regulated by the action of accessory splicing factors that bind the pre-mRNA sequence-specifically and very often show prominent tissue-specific isoform patterns (2,3). Since the novel regulatory mechanism is ubiquitous and not sequence-specific, it is likely mediated by the ubiquitous core of the spliceosome, comprising five snRNPs and a number of ubiquitous co-factors. This hypothesis is consistent with the widespread occurrence of TASS across the eukaryote kingdom and a high predictability of TASS isoform patterns across distantly related species (13,18). Consistently, in a previous study on Arabidopsis 3′ TASS it was noted that isoform patterns are regulated in a condition-specific manner rather than tissue-specifically, and regulatory patterns were remarkably similar across genes (13). Previous studies have suggested the constitutive splicing factors SLU7 and SPF45 (coded by *RBM17*) as mediators of 3′ TASS regulation (16,28). However, it remains to be tested if these factors may play a role in a generalized regulatory model that comprises CEs.

Splicing isoform changes in the cell culture experiments were continuous in time and non-stochastic. Moreover, the isoform changes reversed upon reversal of the cell culture conditions, that is, upon dilution of the cells using fresh medium. These results strongly indicate an underlying regulatory mechanism which is associated with cell environment changes during cultivation. We are just at the beginning to characterize the physiological trigger of this splicing regulation. However, some preliminary conclusions can be drawn from the experiments. Cell–cell contacts probably do not play a role in the regulatory mechanism, since regulation was observed in suspension cultures of HL-60 cells just like adherently growing HEK-293 cells. Moreover, the splicing effects are reproduced in HepG2 cells (results not shown), which are defective in contact inhibition. A conditioning experiment revealed that the splicing effect can be accelerated by pre-incubated culture medium transferred to a low-density culture (Supplementary Figure S8). This indicates that the regulatory effect is mediated through the cell culture medium, suggesting that the trigger either is the shortage or abundance of soluble factor(s). Beside nutrients like glucose or amino acids, the limited availability of ions may cause starvation effects in the cells. For example, zinc is not explicitly included in culture medium formulas, and some second-step splicing factors (e.g. SLU7) are zinc finger proteins. In fact, a zinc chelator has been reported to influence the product formation of *in vitro* splicing reactions (29). Finally, it should be noted that high cell density in culture experiments has been described to activate hypoxia-inducible factor, even at subcritical oxygen concentrations, and to promote transcription of target genes (30). Hypoxia-associated pathways may explain the extreme AS patterns found in brain and skeletal muscle (Figure 4). Brain has a chronically weak oxygen supply due to the blood–brain barrier, and skeletal muscle may experience significant hypoxia upon activity. Liver, on the other hand, has optimal arterial blood supply, in accord with its opposing AS pattern. However, the actual physiological trigger, as well as the signaling pathway that couples it to AS regulation, need to be identified in future studies. The cell culture experiment described here seems a promising system to approach this question.

According to the mechanistic definition of regulation, the TASS regulatory patterns described here do not necessarily imply functional relevance. However, the results provide new conceptual implications for the functional characterization of TASS splicing. TASS splicing regulation affects a very large number of mammalian genes simultaneously, about 2000 if we apply conservative thresholds of specificity and isoform abundance (9,10). This means that the functional potential of this regulatory mechanism is substantial, even if the amplitudes of individual isoform shifts are relatively small. Moreover, a putative relevance of TASS on the phenotypic level is likely rarely dependent on individual genes. Rather, TASS may be lost and gained with high frequency because the selective pressure on individual genes is relatively low. This could explain why the evolutionary conservation of TASS was demonstrated only with the use of differentiated statistical methods (15). We note that this change of perspective suggests shifting the research focus from individual TASS-containing genes toward the global functional signatures of TASS-affected genes. Earlier work has noted that TASS-containing genes preferentially code for splicing-regulatory proteins containing the RNA recognition motif (Pfam PF00076) (11). And 4 of the 10 genes selected for this study happen to code for RNA-binding proteins, as opposed to only 4.7% of all characterized proteins (http:// amigo.geneontology.org). This hints at a feedback regulation of the splicing machinery, as was suggested earlier (11). The current finding that TASS isoform ratios appear rather stable across tissues and cell lines, but respond to a physiological trigger from the environment, further suggests that the ratios are the steady-state result of homeostasis. In addition, oscillatory response patterns observed upon acute changes of the cellular environment do specifically support a model of negative feedback regulation (Figure 2E, Supplementary Figure S9). Thus, TASS may act as the mRNA relay station in a complex homeostasis pathway involving as yet unidentified physiological factor(s).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dagmar Kruspe, Nadine Zeise and especially Ivonne Görlich for expert technical assistance. We also thank Steffen Priebe and Sebastian Müller (Hans Knöll Institute, Jena, Germany) for helpful discussions and two anonymous reviewers for their help to improve the manuscript.

FUNDING

Bundesmisterium für Bildung und Forschung (BMBF) [01GS08182 to M.P. and J.H.]. Source of open access funding: Leibniz Institute for Age Research - Fritz Lipmann Institute, basic funding.
Conflict of interest statement. None declared.

REFERENCES

1. Wang,E.T., Sandberg,R., Luo,S., Khrebtukova,I., Zhang,L., Mayr,C., Kingsmore,S.F., Schroth,G.P. and Burge,C.B. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature*, **456**, 470–476.
2. Black,D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.*, **72**, 291–336.
3. Stamm,S., Ben-Ari,S., Rafalska,I., Tang,Y., Zhang,Z., Toiber,D., Thanaraj,T.A. and Soreq,H. (2005) Function of alternative splicing. *Gene*, **344**, 1–20.
4. Makeyev,E.V., Zhang,J., Carrasco,M.A. and Maniatis,T. (2007) The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. *Mol. Cell*, **27**, 435–448.
5. Singh,R.K. and Cooper,T.A. (2012) Pre-mRNA splicing in disease and therapeutics. *Trends Mol. Med.*, **18**, 472–482.
6. Lareau,L.F., Inada,M., Green,R.E., Wengrod,J.C. and Brenner,S.E. (2007) Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature*, **446**, 926–929.
7. Dou,Y., Fox-Walsh,K.L., Baldi,P.F. and Hertel,K.J. (2006) Genomic splice-site analysis reveals frequent alternative splicing close to the dominant splice site. *RNA*, **12**, 2047–2056.
8. Akerman,M. and Mandel-Gutfreund,Y. (2006) Alternative splicing regulation at tandem 3' splice sites. *Nucleic Acids Res.*, **34**, 23–31.
9. Hiller,M. and Platzer,M. (2008) Widespread and subtle: alternative splicing at short-distance tandem sites. *Trends Genet.*, **24**, 246–255.
10. Sinha,R., Lenser,T., Jahn,N., Gausmann,U., Friedel,S., Szafranski,K., Huse,K., Rosenstiel,P., Hampe,J., Schuster,S. *et al.* (2010) TassDB2—a comprehensive database of subtle alternative splicing events. *BMC Bioinformat.*, **11**, 216.
11. Hiller,M., Huse,K., Szafranski,K., Jahn,N., Hampe,J., Schreiber,S., Backofen,R. and Platzer,M. (2004) Widespread occurrence of alternative splicing at NAGNAG acceptors contributes to proteome plasticity. *Nat. Genet.*, **36**, 1255–1257.
12. Tadokoro,K., Yamazaki-Inoue,M., Tachibana,M., Fujishiro,M., Nagao,K., Toyoda,M., Ozaki,M., Ono,M., Miki,N., Miyashita,T. *et al.* (2005) Frequent occurrence of protein isoforms with or without a single amino acid residue by subtle alternative splicing: the case of Gln in DRPLA affects subcellular localization of the products. *J. Hum. Genet.*, **50**, 382–394.
13. Schindler,S., Szafranski,K., Hiller,M., Ali,G.S., Palusa,S.G., Backofen,R., Platzer,M. and Reddy,A.S. (2008) Alternative splicing at NAGNAG acceptors in Arabidopsis thaliana SR and SR-related protein-coding genes. *BMC Genom.*, **9**, 159.
14. Akerman,M. and Mandel-Gutfreund,Y. (2007) Does distance matter? Variations in alternative 3' splicing regulation. *Nucleic Acids Res.*, **35**, 5487–5498.
15. Hiller,M., Szafranski,K., Sinha,R., Huse,K., Nikolajewa,S., Rosenstiel,P., Schreiber,S., Backofen,R. and Platzer,M. (2008) Assessing the fraction of short-distance tandem splice sites under purifying selection. *RNA*, **14**, 616–629.
16. Chua,K. and Reed,R. (1999) The RNA splicing factor hSlu7 is required for correct 3' splice-site choice. *Nature*, **402**, 207–210.
17. Chern,T.M., van Nimwegen,E., Kai,C., Kawai,J., Carninci,P., Hayashizaki,Y. and Zavolan,M. (2006) A simple physical model predicts small exon length variations. *PLoS Genet.*, **2**, e45.
18. Sinha,R., Nikolajewa,S., Szafranski,K., Hiller,M., Jahn,N., Huse,K., Platzer,M. and Backofen,R. (2009) Accurate prediction of NAGNAG alternative splicing. *Nucleic Acids Res.*, **37**, 3569–3579.
19. Busch,A. and Hertel,K.J. (2012) Extensive regulation of NAGNAG alternative splicing: new tricks for the spliceosome? *Genome Biol.*, **13**, 143.
20. Szafranski,K., Schindler,S., Taudien,S., Hiller,M., Huse,K., Jahn,N., Schreiber,S., Backofen,R. and Platzer,M. (2007) Violating the splicing rules: TG dinucleotides function as alternative 3' splice sites in U2-dependent introns. *Genome Biol.*, **8**, R154.
21. Bradley,R.K., Merkin,J., Lambert,N.J. and Burge,C.B. (2012) Alternative splicing of RNA triplets is often regulated and accelerates proteome evolution. *PLoS Biol.*, **10**, e1001229.
22. Kozmik,Z., Czerny,T. and Busslinger,M. (1997) Alternatively spliced insertions in the paired domain restrict the DNA sequence specificity of Pax6 and Pax8. *EMBO J.*, **16**, 6793–6803.
23. Tsai,K.W. and Lin,W.C. (2006) Quantitative analysis of wobble splicing indicates that it is not tissue specific. *Genomics*, **88**, 855–864.
24. Schindler,S., Heiner,M., Platzer,M. and Szafranski,K. (2009) Comparison of methods for quantification of subtle splice variants. *Electrophoresis*, **30**, 3674–3681.
25. Pan,Q., Shai,O., Lee,L.J., Frey,B.J. and Blencowe,B.J. (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.*, **40**, 1413–1415.
26. Pollard,A.J., Krainer,A.R., Robson,S.C. and Europe-Finner,G.N. (2002) Alternative splicing of the adenylyl cyclase stimulatory G-protein G alpha(s) is regulated by SF2/ASF and heterogeneous nuclear ribonucleoprotein A1 (hnRNPA1) and involves the use of an unusual TG 3'-splice Site. *J. Biol. Chem.*, **277**, 15241–15251.
27. Kramer,M., Huse,K., Menzel,U., Backhaus,O., Rosenstiel,P., Schreiber,S., Hampe,J. and Platzer,M. (2011) Constant splice-isoform ratios in human lymphoblastoid cells support the concept of a splico-stat. *Genetics*, **187**, 761–770.
28. Lallena,M.J., Chalmers,K.J., Llamazares,S., Lamond,A.I. and Valcarcel,J. (2002) Splicing regulation at the second catalytic step by Sex-lethal involves 3' splice site recognition by SPF45. *Cell*, **109**, 285–296.
29. Shomron,N., Malca,H., Vig,I. and Ast,G. (2002) Reversible inhibition of the second step of splicing suggests a possible role of zinc in the second step of splicing. *Nucleic Acids Res.*, **30**, 4127–4137.
30. Kaluz,S., Kaluzova,M., Chrastina,A., Olive,P.L., Pastorekova,S., Pastorek,J., Lerman,M.I. and Stanbridge,E.J. (2002) Lowered oxygen tension induces expression of the hypoxia marker MN/carbonic anhydrase IX in the absence of hypoxia-inducible factor 1 alpha stabilization: a role for phosphatidylinositol 3'-kinase. *Cancer Res.*, **62**, 4469–4477.