Biochemical and Biophysical Research Communications 438 (2013) 346–352

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Pairwise comparisons of ten porcine tissues identify differential transcriptional regulation at the gene, isoform, promoter and transcription start site level

Leila Farajzadeh, Henrik Hornshøj [1], Jamal Momeni, Bo Thomsen, Knud Larsen, Jakob Hedegaard [1], Christian Bendixen, Lone Bruhn Madsen *

Department of Molecular Biology and Genetics, Faculty of Sciences and Technology, Aarhus University, DK-8830 Tjele, Denmark

ARTICLE INFO

Article history:
Received 15 July 2013
Available online 27 July 2013

Keywords:
Gene expression
Pig
Promoter
RNA-seq
Splicing

ABSTRACT

The transcriptome is the absolute set of transcripts in a tissue or cell at the time of sampling. In this study RNA-Seq is employed to enable the differential analysis of the transcriptome profile for ten porcine tissues in order to evaluate differences between the tissues at the gene and isoform expression level, together with an analysis of variation in transcription start sites, promoter usage, and splicing.

Totally, 223 million RNA fragments were sequenced leading to the identification of 59,930 transcribed gene locations and 290,936 transcript variants using Cufflinks with similarity to approximately 13,899 annotated human genes. Pairwise analysis of tissues for differential expression at the gene level showed that the smallest differences were between tissues originating from the porcine brain. Interestingly, the relative level of differential expression at the isoform level did generally not vary between tissue contrasts. Furthermore, analysis of differential promoter usage between tissues, revealed a proportionally higher variation between cerebellum (CBE) versus frontal cortex and cerebellum versus hypothalamus (HYP) than in the remaining comparisons. In addition, the comparison of differential transcription start sites showed that the number of these sites is generally increased in comparisons including hypothalamus in contrast to other pairwise assessments.

A comprehensive analysis of one of the tissue contrasts, i.e. cerebellum versus heart for differential variation at the gene, isoform, and transcription start site (TSS), and promoter level showed that several of the genes differed at all four levels. Interestingly, these genes were mainly annotated to the "electron transport chain" and neuronal differentiation, emphasizing that "tissue important" genes are regulated at several levels. Furthermore, our analysis shows that the "across tissue approach" has a promising potential when screening for possible explanations for variations, such as those observed at the gene expression levels.

1. Introduction

Genomic and transcriptomic data are accumulating at an exceptional rate allowing for the elucidation on how particular functions embodied by specific tissues appear from an universal set of molecular instructions. The understanding of this link is especially challenging in the mammalian brain, where the complexity derives from the abundance and diversity of cell types from which it is comprised [19]. For various tissues including neuronal cell types, deep sequencing of RNAs (RNA-seq) has recently been applied to quantify gene and alternative isoform expression levels [13,27,29]. In RNA-seq experiments all RNAs of a particular sample are randomly fragmented, reverse transcribed, ligated to adapters, and subsequently the fragments are sequenced and levels of gene expression are derived from the number of sequence reads originating from each gene. Furthermore, RNA-seq also allows for the detection of transcription start sites and splice variants, the latter being highly important regarding mammalian brain complexity [15].

Furthermore, it is noteworthy that both genetically engineered pig models and pigs with naturally occurring mutations often more closely recapitulate human disease symptoms than the smaller

* Corresponding author. Address: Department of Molecular Biology and Genetics, Aarhus University, Blichers Allé 20, Postboks 50, 8830 Tjele, Denmark. Fax: +45 87154994.
E-mail address: LoneB.Madsen@agrsci.dk (L.B. Madsen).
[1] Present address: Department of Molecular Medicine (MOMA), Aarhus University Hospital, DK-8200 Aarhus N, Denmark.

0006-291X
http://dx.doi.org/10.1016/j.bbrc.2013.07.074

models [20,23]. Moreover the size of pig genome, gene location and length of coding and non-coding regions resemble that of humans, which for instance has been shown in the characterization of genes important in neurological diseases such as Dystonia [10] and neurodegenerative diseases as Parkinson's and Alzheimer's disease [14,16]. For the past two decades pigs have been used in biomedical research with increasing frequency due to their close resemblance to humans regarding anatomy and physiology [26]. Furthermore, characterization of trinucleotide repeats in genes crucial for the development of neurological diseases associated with expansion of trinucleotide repeat elements, such as Huntington's disease in pigs, shows that these repeats have the potential to expand in pigs, making this species a potential natural model for these diseases [17].

In order to detect abnormal transcription profiles in disease state, it is essential to obtain knowledge of transcription profiles in various tissues including brain regions under normal, healthy conditions. In the current study we have employed RNA-seq on ten different porcine tissues including four brain tissues from two boars to gain an understanding of the variations in transcriptional profiles for these tissues consisting of occipital cortex (OCC), frontal cortex (FCO), hypothalamus and cerebellum along with such diverse tissues as heart, spleen, liver, kidney, lung and *musculus longissimus dorsi*. This has enabled us to perform comparative gene expression analysis of brain regions versus non-brain tissues along with inter-brain tissue comparisons. Hence, we have tested for differentially expressed genes and isoforms, differential splicing, TSS, and differential promoter usage between all ten porcine tissues. Furthermore, our analysis shows that the "across tissue approach" has a promising potential when screening for possible molecular explanations for variations in for instance gene expression levels.

2. Materials and methods

2.1. Animals and sample preparation

Ten tissues from two unrelated one year old Landrace boars were included in the study. Hence, total RNA was extracted from heart, spleen, liver, kidney, lung, *musculus longissimus dorsi*, occipital cortex, hypothalamus, frontal cortex, and cerebellum employing the mirVana™ RNA extraction kit (Ambion) according to manufactures protocol, yielding a total of 20 samples. RNA integrity of the individual RNA samples was assessed on a 2% agarose gel.

2.2. Library preparation and sequencing

Library preparation was performed using the mRNA-seq library prep kit from Illumina according to manufacturer's protocol. Briefly, 10 μg total RNA of each sample was employed in polyA mRNA selection using magnetic beads, followed by thermal fragmentation, resulting in a total of 10 samples for each of the two animals. Subsequently, the fragmented mRNA was reverse transcribed using reverse transcriptase (SuperScript II) and random primers followed by second strand cDNA synthesis using DNA polymerase I. The cDNA was end repaired using T4 DNA polymerase and Klenow DNA polymerase and size selected on a low melting 2% agarose gel and fragments corresponding to sizes of 200 nucleotides were excised from the gel and DNA was recovered applying the QIAquick gel extraction kit (Qiagen). Finally, in order to amplify the libraries, 15 cycles of PCR were employed using primers from the multiplexing sample preparation Oligonucleotide kit (Illumina). The amplified libraries were purified using the QIAquick PCR purification kit (QIAGEN) and the concentration of each library was determined using the Qubit flurometer (Invitrogen). The purified libraries were denatured and diluted to a concentration of 10 nM, and sequenced (50 bp single-end) as multiplex of ten samples per lane in a total of two lanes on a Genome Analyzer (Illumina). Output was transferred to the pipeline computer (Illumina Pipeline version 1.4), quality filtered, and analyzed using the Genome Analyzer Pipeline Software generating the raw fastq files. The fastq files and the processed data have been submitted to ArrayExpress with the accession number E-MTAB-1405.

2.3. Mapping and assembly of fragments

The quality filtered fastq files were mapped to the *Sus scrofa* reference genome build 10.2 [9] employing TopHat version 1.0.12 [24] using default settings, implying that reads were split in segments of 25 bp and mapped independently to the reference genome. Next, Cufflinks version 0.8.0 [25] was applied to assemble the aligned reads into clusters of overlapping reads, then building a graph which represents all possible isoforms, followed by traversing the created graph to assemble isoforms by finding the minimum set of transcripts that explain the intron junctions within the sequenced reads, allowing for the detection of new isoforms and transcripts. Additionally, the normalized RNA-seq fragment counts were used in the Cufflinks program to estimate the relative abundance of each transcript and are reported in fragments per kilobase of exon per million fragments mapped (FPKM). Transcripts were annotated to the human Refseq database downloaded at July 6th 2012.

2.4. Differential expression

Test for differential expression at the gene and isoform level was accomplished using Cuffdiff employing all XLOC IDs established by Cufflinks. Furthermore, test for differential splicing, promoter usage and TSS was also accomplished using the Cuffdiff program. The tests were performed pairwise between all ten tissues yielding a total of 45 tests per condition and all tests were performed using default settings in Cuffdiff. Type I errors were evaluated using Benjamini–Hochberg (1995) correction for multiple testing [3] and a significance level of 0.05. Gene-annotation enrichment analysis and functional annotation clustering were performed using transcripts annotated to the human Refseq database.

2.5. Visualization of spliced alignments

In silico visualization of splice data was accomplished using a local copy of the Integrative Genomics Viewer (IGV) software available at http://www.broadinstitute.org/igv/ [22] and establishment of Venn diagrams was completed in Venny [21].

3. Results and Discussion

It has previously been shown that the pig in terms of genetics, anatomy, and physiology mimics humans extremely well, which is emphasized by the increased use of pigs in biomedical research [26]. In this study massive parallel sequencing was employed to gain a comprehensive understanding of the transcriptome profile including variation in isoforms, TSS, and promoter usage between various porcine tissues including hypothalamus, frontal cortex, occipital cortex, and cerebellum.

3.1. RNA-Seq data

Totally, 223 million fragments were sequenced, allowing us to cover approximately 80% of the genes expressed. The fragments

were mapped to the high quality *Sus scrofa* reference genome build 10.2 [9] using TopHat enabling the downstream isoform construction [25]. A total of 192 million reads were aligned to the reference genome yielding an overall mapping percentage of 86% with a standard deviation of 9.4%. Following mapping of the RNA-seq reads, transcripts were assembled using Cufflinks, which also reconstructed the various isoforms present in the different porcine tissues. Moreover, the relative abundance of each transcript in fragments per kilobase of exon per kilobase of fragments mapped (FPKM) was estimated by Cufflinks. Furthermore, the transcripts were annotated to the human Refseq database.

The extraction of the number of unique genes (XLOC id's) present in each tissue showed a variation ranging from 38,372 in *musculus longissimus dorsi* to 48,231 in spleen, Table 1, and a total of 59,930 unique genes (XLOC id's) corresponding to approximately 13,899 different annotated genes using homology to the human refseq database. This also means that the majority of genes expressed in a specific tissue are ubiquitously expressed genes. Furthermore, the total numbers of different isoforms have been calculated yielding 290,936 different porcine transcripts (unique TCON id's) ranging from 118,662 in heart to 155,339 in lung. Generally, the data, with a few exceptions show that the relative number of genes and isoforms is highest in the four brain tissues. This is in general agreement with the fact that brain is considered the most complex organ of the body.

From the 59,930 unique XLOC IDs, 5817 ID's equaling 9.7%, were found to be new. Moreover, a substantial part of the new transcripts, are expressed in all tissues, emphasizing the potential endogenous importance of these transcripts. Interestingly, 3.2% of the newly identified transcripts are specifically expressed in only one of the tissues, and, hence, further analysis is needed to clarify if these transcripts have a function which could be of tissue specific importance.

3.2. Differential analysis

In this study the FPKM values generated by Cufflinks were applied in Cuffdiff to test for differential gene expression and were performed pairwise between all tissues. Furthermore, in addition to simple changes in gene expression Cuffdiff also reports differential transcript expression, analysis of differential splicing, differential regulation via promoter switching, and differential changes ins TSS, the latter being a central point in the regulated control of gene transcription when RNA polymerase II enzyme, along with its cofactors forms the pre-initiation complex and begins transcription of the downstream gene [12].

The number of differentially expressed genes (XLOC id's) between the various tissues ranges from 286 to 7064, the lowest numbers belonging to tissues originating from the brain, i.e. hypothalamus versus frontal cortex, and the highest numbers of differential expressed genes originates from comparisons between tissues consisting of cells which are functionally very different, i.e. occipital cortex versus *musculus longissimus dorsi*, Fig. 1A.

This tendency also holds true when calculating the relative number of expressed genes in relation to the total number of differential events. The proportion is illustrated for all differentially calculated comparisons and shown in Fig. 1B together with the absolute numbers of differentially expressed events for each of the 45 comparisons, Fig. 1A.

Overall, the proportion of genes differentially expressed for each of the comparisons is lower than the proportion of differentially expressed isoforms. This is also in agreement with the observation that isoform expression by a gene does not pursue a minimalistic expression approach meaning that there is a tendency for genes to express several isoforms simultaneously [4]. Moreover the major isoform of genes expressed from a specific cell or cell line has been shown to vary with cell type and even point in time [4], making it important to stress that the transcriptome landscape illustrated in our study is a snapshot of a variety of cell types. However, alternative isoform expression is a very important aspect of the total transcriptome picture, since it has been shown to be crucial for translation and very important during development [18].

The comparison of differential TSS show that the number of these sites are generally proportionally higher when comparing hypothalamus to the remaining tissues than for the other TSS tissue comparisons including the remaining brain versus non brain tissues. This means that a higher proportion of transcription start site regulated genes exist between hypothalamus and spleen, lung, liver, kidney, and occipital cortex than for the other contrasts. Interestingly, the comparison for differential TSS usage between hypothalamus and frontal cortex shows the lowest proportional percentage, meaning that the transcriptomic differences between these two tissues are relatively low at the TSS level, whereas higher proportional differences exist at the isoform level, Fig. 1B.

Analyzing the data in Cuffdiff using default settings only reveals few (<5) differentially spliced genes between the various tissues. However, it is likely that the result shown in Fig. 1 for the differentially spliced genes is an underestimation of the true number of differentially spliced genes since the criteria in Cuffdiff for reporting a differentially spliced gene not only requires the test result to be below the level of significance (0.05), but also, that the number of sequences for a particular variant to be tested are at least ten. Furthermore, differential splicing events of genes with low or intermediate expression levels can be missed by RNA-seq [29], requiring a high sequencing depth to avoid a detection bias against highly expressed genes. However, interestingly, the highest number of relative differentially spliced genes is seen between comparisons where brain tissues are included, i.e. cerebellum versus

Table 1
Number of cDNA fragments sequenced and mapped to the *Sus scrofa* genome build 10.2 using TopHat. Number of genes and isoforms (FPKM > 1) represented in each tissue is presented in the last two columns.

Tissue	Total no of reads	Reads mapped	Mapped reads [%]	No of genes (XLOC id's)	No of isoforms (TCONS)
Heart	20,228,265	15,983,983	79%	40,860	118,662
Spleen	27,088,408	23,008,315	85%	45,533	144,722
Lung	25,073,509	24,633,093	98%	47,870	155,339
M. l. dorsi	32,528,975	27,521,194	85%	42,660	133,404
Liver	30,343,507	27,870,121	92%	38,372	110,493
Kidney	18,874,691	18,177,590	96%	43,230	130,286
Frontal cortex	9,008,663	6,365,159	71%	46,256	132,892
Occipital cortex	38,190,450	3,2591,392	85%	47,410	153,616
Cerebellum	15,997,390	11,557,438	72%	48,231	149,059
Hypothalamus	5,877,539	4,532,964	77%	45,545	126,471
Experiment wise	223,211,397	192,241,249	84%	59,930	290,936

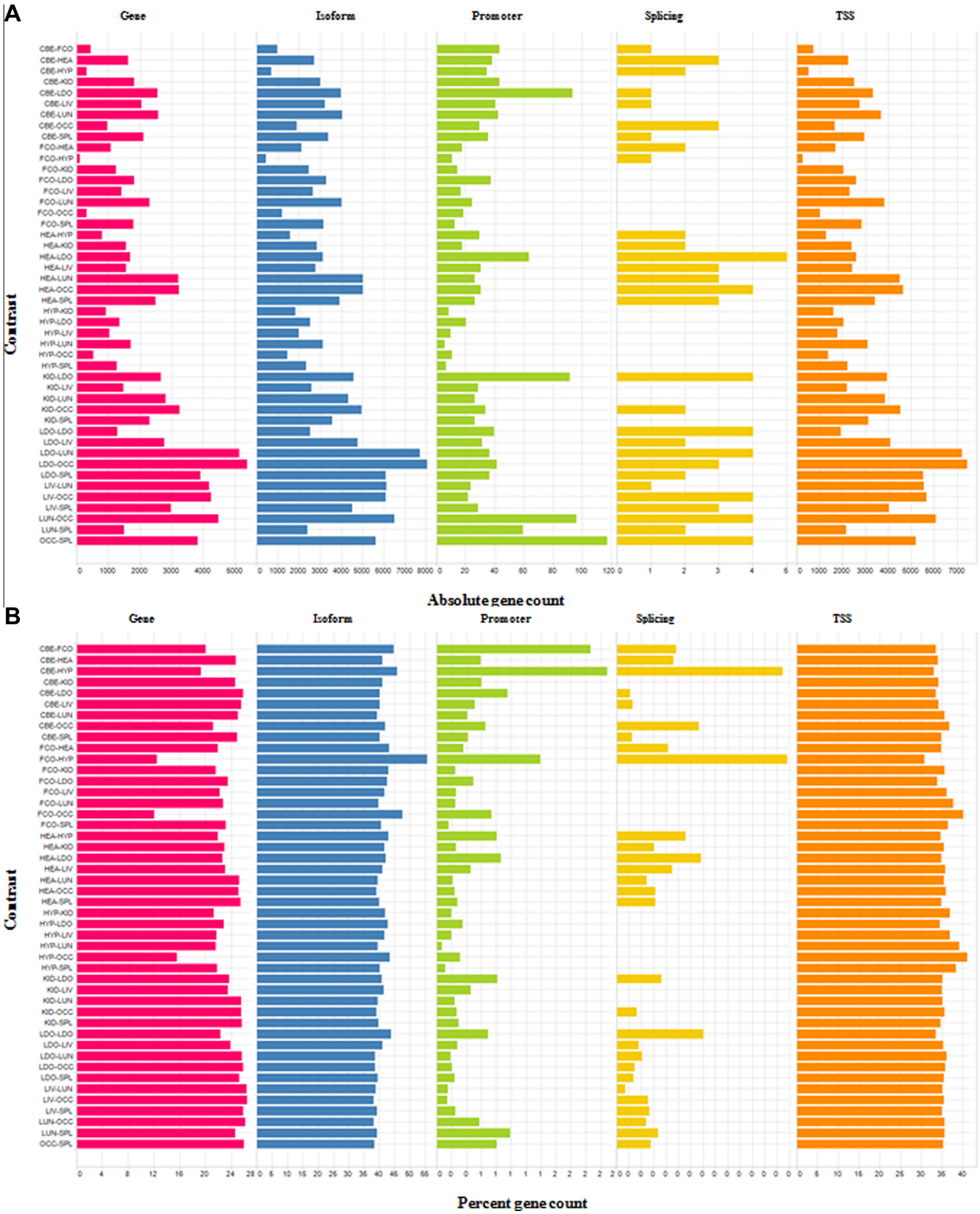

Fig. 1. Bar chart showing the differential events at the gene, isoform, promoter usage, splicing, and TSS level for all 45 pairwise tissue comparisons. (A) The absolute gene count defined as Cufflinks id's, at each level for all contrasts. (B) The gene count in percent within each contrast across the five levels for the differential events.

hypothalamus and frontal cortex versus hypothalamus, indicating that differential splicing is an important gene regulating mechanism in the porcine brain. This is in good agreement with studies of alternative splicing in primate brain transcriptomes, showing that even between various primate species there are considerable splice differences in the brain transcriptome which can possibly

be linked to evolution of neuronal gene regulation and function [15].

One of the genes which in this study is identified as differentially spliced both between brain tissues, i.e. CBE versus OCC, and between brain and non-brain tissues, i.e. CBE versus HEA, CBE versus liver, and OCC versus liver, is PJA1 encoding Praja Ring Finger Protein 1, where seven different isoforms are detected, Fig. 2A.

Previous studies of PJA1 has shown that it is ubiquitously expressed in the brain including the cerebellum and occipital lobe and that it is likely to participate in ubiquitination in the brain and in humans several multiple splice forms of the gene have been identified [31]. Furthermore, PJA1 has been linked to X-linked mental retardation, (MRX) and craniofrontonasal syndrome (CFNS) [30], making the splicing pattern for this particular gene interesting due to the X-linked nature of these disorders. Furthermore, it is important to stress that even alternative splicing events that generally are present in a variety of tissues of ubiquitously expressed genes, could have major consequences in the nervous system. An example is the disease related to abnormal alternative splicing of MAPT in patients with frontotemporal dementia and parkinsonism linked to chromosome 17 [11], where the MAPT gene is ubiquitously expressed, and the causal mutation causes splicing defects in various tissues, although the disease phenotype of the abnormal splicing is apparently limited to neuronal cells [6]. Therefore, general knowledge on differential splicing between tissues provides valuable information regarding interpretation of information in relation to alternative splicing and disease [28].

Pairwise testing between the various tissues using Cuffdiff also enabled the identification of alternative promoter usage between the various tissues, and totally 1,068 genes (XLOC id's) were statistically significant in relation to differential promoter usage. Both in absolute frequencies and relative frequencies differential promoter usage is most common in brain tissues and in particularly in cerebellum versus frontal cortex and cerebellum versus hypothalamus, the proportion of differential promoter usage is increased in comparison to the remaining contrasts. One of the genes which in the contrast between cerebellum and frontal cortex is subjected to differential promoter usage is Zinc finger protein of the cerebellum 1 (ZIC1), a zinc-finger transcription factor, which during gestational development mainly is expressed in the dorsal neural tube and plays an important role in maturation of the cerebellum [7]. Furthermore, ZIC1 is involved in the most common malformation of

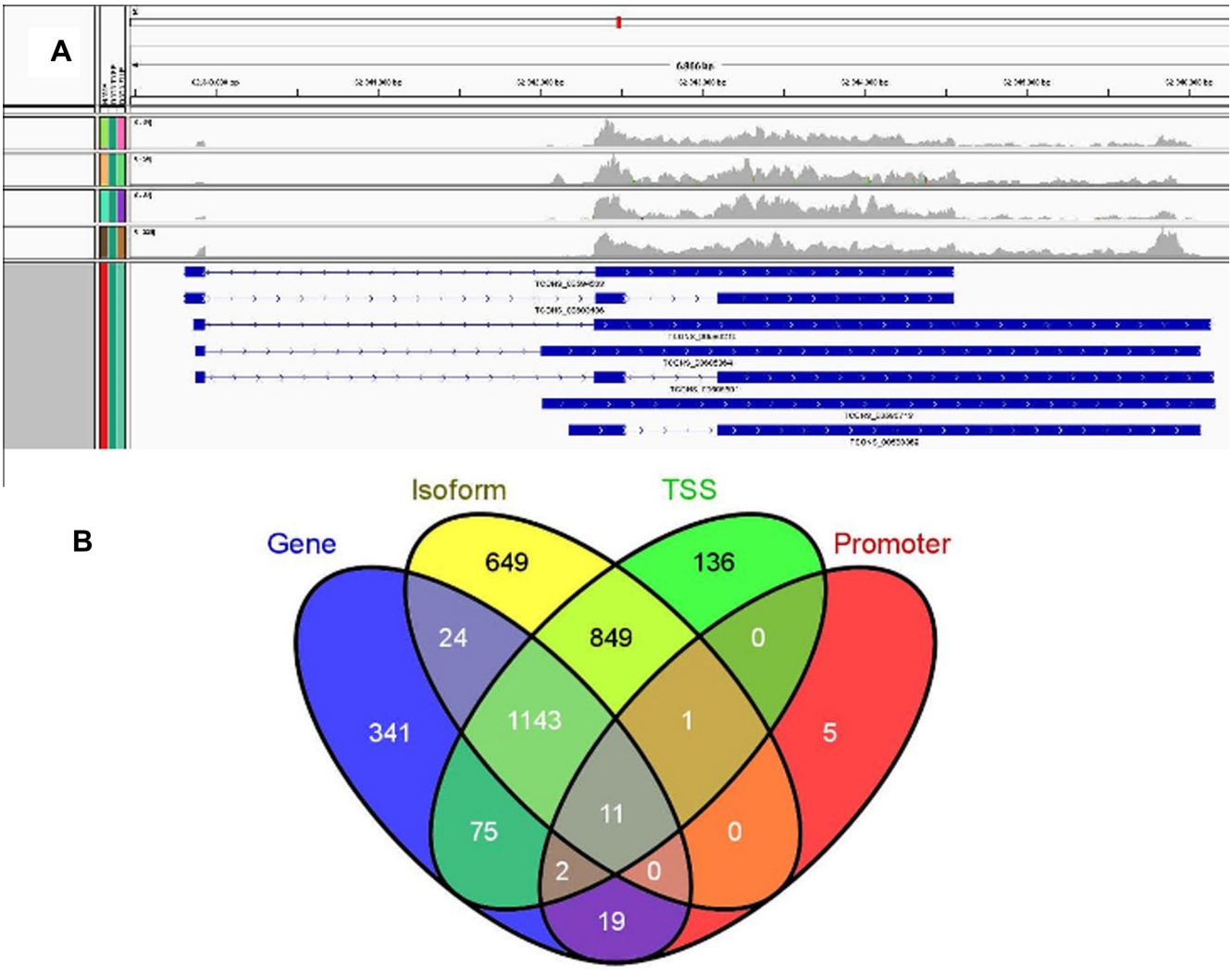

Fig. 2. (A) In silico visualization in IGV of seven splice variants of the porcine PJA1 gene. Differential splicing is significant between CBE versus OCC, CBE versus HEA, CBE versus liver and OCC versus liver. The difference in splice forms between the various tissues is not seen as a presence or not presence of a particular splice variant but is seen as a difference in level of expression of the individual splice forms. (B) Venn diagram showing number of differential XLOC ID's within the gene expression, isoform, TSS and promoter level between cerebellum and heart. The diagram illustrates number of genes (XLOC id' being unique to each of the four levels and number of genes (XLOC IDs) which, according to Cuffdiff, are significantly differentially "expressed" at several levels. Hence, the grey area depicts 11 genes (XLOC id's), which are differentially expressed at both gene and isoform level, and also are significant regarding promoter usage and transcription start sites.

Table 2
Genes which are represented as significantly differentially regulated at the gene, isoform, promoter, and TSS level between cerebellum and heart.

Gene symbol	Gene description	Gene function	Functional annotation, DAVID
FRRS1L	Ferric-chelate reductase 1-like protein	Reduce Fe(3+) to Fe(2+) before its transport from the endosome to the cytoplasm (by similarity to FRRS1)	Electron transport chain
ZNF385C	Zing finger protein 385C	Transcription factor	–
HLF	Hepatic leukemia factor	Bzip transcription factor	–
RNASEK	Ribonuclease, RNase K	Endonuclease function	–
B3GNT1	UDP-GlcNAc:betaGal β-1,3-N-acetylglucosaminyltransferase1	Synthesis of aminoglycans	Neuron differentiation
UQCRB	Ubiquinol-cytochrome c reductase binding protein	Redox linked proton pumbing	Electron transport chain
BHLHE22	Basic helix loop helix family member 22	Inhibits DNA binding	Neuron differentiation
CA8	Carbonic anhydrase VIII	Unknown, but does not have carbonic anhydrase activity	–
MRF	Myelin gene regulation factor	Transcription regulator, important for myelination	–
RELN	Reelin	Serine protease, role in layering of neurons in cerebellum and cerebral cortex	Neuron differentiation
NDUFB11	NADH dehydrogenase 1 β subcomplex 11	Accessory subunit of the mitochondrial membrane respiratory chain NADH dehydrogenase	Electron transport chain

the cerebellum, named Dandy–Walker malformation [8] and high levels of ZIC1 expression have been found in medulloblastoma, a childhood brain tumor [1]. The differential promoter usage elucidated in this study could very well result in the differential expression which is present between porcine cerebellum and the remaining porcine tissues, where the FPKM values are very low (<10). Since the expression pattern shown in this study for porcine ZIC1 is similar to the expression pattern for human ZIC1, it is likely that differential promoter usage could also be the reason for the expression pattern of ZIC1 in humans. However, experimental testing needs to be performed in order to draw a final conclusion on this.

3.3. Gene regulation within specific tissue contrasts

To link the differential results from the Cuffdiff tests on the gene, isoform, TSS and promoter level, we compared the lists of genes (XLOC id's) within the cerebellum versus heart contrast. The result of the comparison is shown in Fig. 2B.

In order to evaluate if the genes present in the intersection between the gene, isoform, TSS and promoter level could be associated to any common functions, the 11 genes represented at all four levels were investigated further. The result is shown in Table 2.

A functional annotation of the 11 genes presented in Table 2 using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) showed that six of the genes could be ascribed to either neuron differentiation (*P*-value 0.026) or the electron transport chain (*P*-value 0.0019). Since the electron transport chain is a "housekeeping" function common in nearly all cells, it is of interest that these data show a difference in regulation of genes between cerebellum and heart which can be ascribed to this particular term. Although the data have not been functionally validated, the presence of these three genes at the gene, isoform, TSS, and promoter level makes it probable that a difference exist at the gene regulatory level between these two tissues regarding the electron transport chain.

Moreover, also genes which can be ascribed to neuronal differentiation are significantly regulated at all four levels included in the comparison. Furthermore, several of the remaining genes including CA8 and MRF can through literature study also be associated to important functions in the CNS. Even though the exact role of CA8 is still not known, it has been investigated that a mutation in the gene can be associated with ataxia and a mild retardation in humans, leading to the assumption that the gene has an important function in the brain [2]. The MRF gene encodes a transcriptional regulator important for myelination of oligodendrocytes and mice lacking the gene fail to myelinate and die because of seizures during the third week after birth [5]. Our data from pigs further reveals that CA8 and MRF at least in porcine cerebellum is regulated at several levels, which further support the importance of the genes in the brain.

Conclusively, by comparing very diverse tissues we might obtain regulatory information which otherwise would be unavailable, when comparing related tissues, such as comparison between tissues within the brain. Therefore, the approach applied in this study is a valuable tool when screening for mechanisms associated to gene regulation and, hence, cell and tissue function.

Acknowledgments

The authors wish to thank Mette Jeppesen and Mahesha Perera for technical assistance with sequence library preparation and Illumina sequencing.

References

[1] J. Aruga, Y. Nozaki, M. Hatayama, Y.S. Odaka, N. Yokota, Expression of ZIC family genes in meningiomas and other brain tumors, BMC Cancer 10 (2010) 79.
[2] A. Aspatwar, M.E. Tolvanen, C. Ortutay, S. Parkkila, Carbonic anhydrase related protein VIII and its role in neurodegeneration and cancer, Curr. Pharm. Des. 16 (2010) 3264–3276.
[3] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. R. Stat. Soc. 57 (1995) 289–300.
[4] S. Djebali, C.A. Davis, A. Merkel, et al., Landscape of transcription in human cells, Nature 489 (2012) 101–108.
[5] B. Emery, D. Agalliu, J.D. Cahoy, et al., Myelin gene regulatory factor is a critical transcriptional regulator required for CNS myelination, Cell 138 (2009) 172–185.
[6] M.A. Garcia-Blanco, A.P. Baraniak, E.L. Lasda, Alternative splicing in disease and therapy, Nat. Biotechnol. 22 (2004) 535–546.
[7] I. Grinberg, K.J. Millen, The ZIC gene family in development and disease, Clin. Genet. 67 (2005) 290–296.
[8] I. Grinberg, H. Northrup, H. Ardinger, C. Prasad, W.B. Dobyns, K.J. Millen, Heterozygous deletion of the linked genes ZIC1 and ZIC4 is involved in Dandy–Walker malformation, Nat. Genet. 36 (2004) 1053–1055.
[9] M.A.M. Groenen, A.L. Archibald, H. Uenishi, et al., Pig genomes provide insight into porcine demography and evolution, Nature 491 (2012) 393–398.
[10] C. Henriksen, L.B. Madsen, C. Bendixen, K. Larsen, Characterization of the porcine TOR1A gene: the first step towards generation of a pig model for dystonia, Gene 430 (2009) 105–115.
[11] M. Hutton, C.L. Lendon, P. Rizzu, et al., Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17, Nature 393 (1998) 702–705.
[12] J.T. Kadonaga, Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors, Cell 116 (2004) 247–257.

[13] H.J. Kang, Y.I. Kawasawa, F. Cheng, et al., Spatio-temporal transcriptome of the human brain, Nature 478 (2011) 483–489.
[14] K. Larsen, L.B. Madsen, Sequence conservation between porcine and human LRRK2, Mol. Biol. Rep. 36 (2009) 237–243.
[15] L. Lin, S. Shen, P. Jiang, S. Sato, B.L. Davidson, Y. Xing, Evolution of alternative splicing in primate brain transcriptomes, Hum. Mol. Genet. 19 (2010) 2958–2973.
[16] L.B. Madsen, B. Thomsen, K. Larsen, et al., Molecular characterization and temporal expression profiling of presenilins in the developing porcine brain, BMC Neurosci. 8 (2007) 72.
[17] L.B. Madsen, B. Thomsen, C.A. Solvsten, C. Bendixen, M. Fredholm, A.L. Jorgensen, A.L. Nielsen, Identification of the porcine homologous of human disease causing trinucleotide repeat sequences, Neurogenetics 8 (2007) 207–218.
[18] C.S. Maxwell, I. Antoshechkin, N. Kurhanewicz, J.A. Belsky, L.R. Baugh, Nutritional control of mRNA isoform expression during developmental arrest and recovery in *C. elegans*, Genome Res. 22 (2012) 1920–1929.
[19] S.B. Nelson, C. Hempel, K. Sugino, Probing the transcriptome of neuronal cell types, Curr. Opin. Neurobiol. 16 (2006) 571–576.
[20] V.H. Nielsen, C. Bendixen, J. Arnbjerg, C.M. Sorensen, H.E. Jensen, N.M. Shukri, B. Thomsen, Abnormal growth plate function in pigs carrying a dominant mutation in type X collagen, Mamm. Genome 11 (2000) 1087–1092.
[21] J.C. Oliveros. Venny: an interactive tool for comparing lists with Venn diagrams. (2007). (http://bioinfogp.cnb.csic.es/tools/venny/index.html).
[22] J.T. Robinson, H. Thorvaldsdottir, W. Winckler, et al., Integrative genomics viewer, Nat. Biotechnol. 29 (2011) 24–26.
[23] A. Sironen, B. Thomsen, M. Andersson, V. Ahola, J. Vilkki, An intronic insertion in KPL2 results in aberrant splicing and causes the immotile short-tail sperm defect in the pig, Proc. Natl. Acad. Sci. USA 103 (2006) 5006–5011.
[24] C. Trapnell, L. Pachter, S.L. Salzberg, TopHat: discovering splice junctions with RNA-Seq, Bioinformatics 25 (2009) 1105–1111.
[25] C. Trapnell, B.A. Williams, G. Pertea, et al., Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation, Nat. Biotechnol. 28 (2010) 511–515.
[26] E.M. Walters, E. Wolf, J.J. Whyte, et al., Completion of the swine genome will simplify the production of swine as a large animal biomedical model, BMC Med. Genomics 5 (2012).
[27] E.T. Wang, R. Sandberg, S. Luo, et al., Alternative isoform regulation in human tissue transcriptomes, Nature 456 (2008) 470–476.
[28] G.S. Wang, T.A. Cooper, Splicing in disease: disruption of the splicing code and the decoding machinery, Nat. Rev. Genet. 8 (2007) 749–761.
[29] Z. Wang, M. Gerstein, M. Snyder, RNA-Seq: a revolutionary tool for transcriptomics, Nat. Rev. Genet. 10 (2009) 57–63.
[30] I. Wieland, C. Weidner, R. Ciccone, et al., Contiguous gene deletions involving EFNB1, OPHN1, PJA1 and EDA in patients with craniofrontonasal syndrome, Clin. Genet. 72 (2007) 506–516.
[31] P. Yu, Y. Chen, D.A. Tagle, T. Cai, PJA1, encoding a RING-H2 finger ubiquitin ligase, is a novel human X chromosome gene abundantly expressed in brain, Genomics 79 (2002) 869–874.