Predominant contribution of *cis*-regulatory divergence in the evolution of mouse alternative splicing

Dissertation

zur Erlangung des akademischen Grades des Doktors der Naturwissenschaften
(Dr. rer. nat.)

eingereicht im

Fachbereich Biologie, Chemie, Pharmazie
der
Freien Universität Berlin

Vorgelegt von

Qingsong Gao (高青松)

aus Shandong, China

February 2016

1. Gutachter: Prof. Dr. Wei Chen

2. Gutachter: Prof. Dr. Florian Heyd

Datum der Disputation: 2016-05-10

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen verwendet habe. Ich erkläre weiterhin, dass ich die vorliegende Arbeit oder deren Inhalt nicht in einem früheren Promotionsverfahren eingereicht habe.

I hereby declare that this thesis is my own original research work and has not been submitted in any form for another degree of diploma at any university or other institute of education. Contributions from others have been clearly acknowledged in the text and references to literatures are given.

Qingsong Gao

2016-02-01

Preface

All the results presented here are adapted from the following published article:

Gao, Q.*, Sun, W.*, Ballegeer, M., Libert, C., and Chen, W. (2015). Predominant contribution of *cis*-regulatory divergence in the evolution of mouse alternative splicing. Molecular Systems Biology 11, 816.

DOI: 10.15252/msb.20145970

Online link: http://msb.embopress.org/content/11/7/816.long

Author Contributions:

QG, WS and WC conceived and designed the project. WS did the experiments. MB and CL prepared the mice liver tissues. QG analyzed the data. QG, WS and WC wrote the manuscript with input from MB and CL.

Acknowledgement

I would like to open by thanking my wife Xiao, who has accompanied me with her love and unlimited patience since we met and fell in love at college nearly ten years ago. She has been the guiding star of both my life and career. It was her who introduced me into the field of computational biology. I would never be able to accomplish this work without her advice, support and encouragement. I consider myself truly blessed to have married her. I hereby dedicate this thesis to her.

I would like to express my sincere gratitude to my supervisor Prof. Dr. Wei Chen for inviting me to join his lab and leading me into the field of system biology. During all these years, he has been a wonderful mentor for me. His patience and support made it much easier for me to survive in the new field. He also gave me many opportunities to participate in many different projects and attend many scientific conferences. I also appreciate his suggestions on scientific thinking and presentation as well as my career. The knowledge and experience I gained in his lab will be priceless for my life.

I would like to thank Prof. Dr. Florian Heyd for giving me the opportunity to join Free university of Berlin. His generous support has been of great help during my stay, and for all my processes in the university. I also appreciate him for the opportunity to get me involved in a very interesting project in his lab.

I would like to thank Wei (Sunny) Sun for wonderful collaborations in several projects. Moreover, he helped me a lot during my PhD study. Especially at start, he patiently explained my thousands of questions, including basic biological knowledge, novel techniques and ideas of our project.

I would like to thank my collaborators from other labs: Dr. Regina Kanski for penultimate exon project; Prof. Dr. Claude Libert and Marlies Ballegee for hybrid

mouse project. I especially thank Dr. Regina Kanski for the translation of the Zusammenfassung.

I would like to thank my colleagues Claudia Quedenau, Dr. Sebastian Fröhler, Madlen Sohn and Claudia Langnick for great help on many translation-related things including all my contracts in German. They really made my life here much easier.

I would like to thank Dr. Xi Wang and Dr. Xintian You for helpful discussions on RNA decay and mouse exome sequencing projects, respectively. I also thank all the other members of Prof. Dr. Wei Chen's group, including Dr. Yuhui Hu, Dr. Na Li, Dr. Wei Sun, Dr. Yongbo Wang, Dr. Kun Song, Dr. Gangcai Xie, Bin Zhang, Tao Chen, Hang Du, Jingyi Hou and Mirjam Feldkamp for their help, discussions and the wonderful time we had together. Thank Sabrina Deter and Sylvia Sibilak for their great help for all the administrative processes.

I would also like to thank the Chinese community in Campus, especially Dr. Zisong Chang, for both scientific communications and a lot of fun.

I would like to thank the China Scholarship Council (CSC) for providing me a four-year scholarship.

Lastly but very importantly, I would like to express my special thanks to my parents in China for their love and support.

Summary

Divergence of alternative splicing represents one of the major driving forces to shape phenotypic diversity during evolution. However, the extent to which these divergences could be explained by the evolving *cis*-regulatory versus *trans*-acting factors remains unresolved. To globally investigate the relative contributions of the two factors for the first time in mammals, we measured splicing difference between C57BL/6J and SPRET/EiJ mouse strains and allele-specific splicing pattern in their F1 hybrid. Out of 11,818 alternative splicing events expressed in the cultured fibroblast cells, we identified 796 with significant difference between the parental strains. After integrating allele-specific data from F1 hybrid, we demonstrated that these events could be predominately attributed to *cis*-regulatory variants, including those residing at and beyond canonical splicing sites. Contrary to previous observations in *Drosophila*, such predominant contribution was consistently observed across different types of alternative splicing. Further analysis of liver tissues from the same mouse strains and re-analysis of published datasets on other strains showed similar trends, implying in general the predominant contribution of *cis*-regulatory changes in the evolution of mouse alternative splicing.

Zusammenfassung

Differentielles, alternatives Spleißen stellt eine der größten evolutionären Antriebskräfte dar um eine phenotypische Vielfalt zu formen. Es ist jedoch unklar zu welchem Grad diese Unterschiede durch *cis*- oder *trans*-regulatorische Faktoren erklärt werden können. Hier wurde zum ersten Mal in Säugetieren ein relativer Einfluss dieser zwei Faktoren umfassend untersucht. Dazu wurden die Unterschiede im Spleißen zwischen C57BL/6J und SPRET/EiJ Mausstämmen und das allelspezifische Spleißmuster in der ersten hybriden F1-Generation untersucht. Von 11.818 alternativen Spleißvorgängen in kultivierten Fibroblasten wurden 796 identifiziert, die einen signifikanten Unterschied zwischen den Elternstämmen zeigen. Durch die Integration von allelspezifischen Daten der ersten hybriden F1-Generation konnten wir zeigen, dass diese Spleißvorgänge größtenteils durch *cis*-regulatorischen Varianten kontrolliert werden, die konstitutive Spleißstellen und andere Sequenzen betreffen. Dabei waren verschiedene Mechanismen von alternativen Spleißen betroffen, was im Gegensatz zu vorherigen Beobachtungen in Drosophila steht. Des Weiteren zeigten eine Analyse von Lebergewebe aus den gleichen Mausstämmen und eine erneute Analyse von veröffentlichten Daten anderer Mausstämme den gleichen Trend. Diese Daten implizieren, dass in der Maus überwiegend *cis*-regulatorische Veränderungen zur Evolution von alternativem Spleißen beitragen.

Table of contents

1. Introduction

1.1 Alternative splicing

Alternative splicing (AS) is a ubiquitous biological process in eukaryotic organisms by which different combinations of 5' and 3' splice site pairs in precursor mRNA transcripts are selected resulting in the production of diverse mRNA isoforms (Blencowe, 2006; Matlin et al., 2005; Nilsen and Graveley, 2010). Recent studies using high-throughput sequencing indicate that about 25%, 60% and 90% of multi-exon genes in *Caenorhabditis elegans*, *Drosophila melanogaster* and humans, respectively, undergo AS (Gerstein et al., 2010; Graveley et al., 2011; Pan et al., 2008; Ramani et al., 2011; Wang et al., 2008). In addition to enhancing transcriptome plasticity and proteome diversity, AS is also involved in regulation of other gene expression processes including mRNA localization, stability and translation (Moore and Proudfoot, 2009; Proudfoot, 2011; Yap and Makeyev, 2013) thereby playing important roles in cell differentiation, sex differentiation and development (Blekhman et al., 2010; Chen et al., 2012; Hartmann et al., 2011; Kalsotra and Cooper, 2011; Stamm et al., 2005).

1.2 Regulation of alternative splicing

Like other levels of gene control, AS is regulated by complex interactions between *cis*-regulatory sequence elements and *trans*-acting factors (Chen and Manley, 2009; Fu and Ares, 2014; Jangi and Sharp, 2014; Kornblihtt et al., 2013; Lu et al., 2012; Matlin et al., 2005; Wang and Burge, 2008) (Fig 1.1).

The most essential core splicing elements include the 5' splice site and the 3' splice site, which define the exon-intron boundary, as well as the branchpoint, which lies upstream of the 3' splice site. These *cis*-regulatory elements are recognized by the spliceosome, a macromolecular RNA-protein complex that is responsible for introns removal. The sequences of these core splicing signals can module their splicing strength, defined as the probability that they will be recognized by the spliceosome, thereby influencing the frequency with which an exon is selected (Matlin et al., 2005; McManus et al., 2014).

In addition to core splicing signals, other auxiliary *cis*-regulatory elements also participate in the process, either promoting or inhibiting splicing. These elements include exonic splicing enhancers (ESEs), intronic splicing enhancers (ISEs), exonic splicing silencers (ESSs) and intronic splicing silencers (ISSs), depending on their position and function. Such auxiliary elements are recognized by sequence-specific *trans*-acting RNA-binding proteins (RBPs). There regulatory RBPs target components of the spliceosome that associate with the 5' or the 3' splice site to activate or inhibit the use of that site (Fig 1.1). For example, serine/arginine-rich proteins (SR proteins) typically bind to ESEs, where they interact with and recruit various components of the spliceosome to enhance adjacent 5' and 3' splice sites recognition. In contrast, heterogeneous nuclear ribonucleoprotein particles (hnRNPs) usually bind to ESSs or ISSs to inhibit exon selection (Chen and Manley, 2009; Graveley, 2009; Lu et al., 2012). Other auxiliary splicing factors have also been shown to be involved in the regulation of AS, some of which are expressed in a tissue-specific manner (Jelen et al., 2007; Kafasla et al., 2012; Kornblihtt et al., 2013; Lee et al., 2009). In some cases, AS can also be regulated by mechanisms that do not involve auxiliary splicing regulators, suggesting the existence of unconventional mode of splicing regulation (Graveley, 2009; Yu et al., 2008).

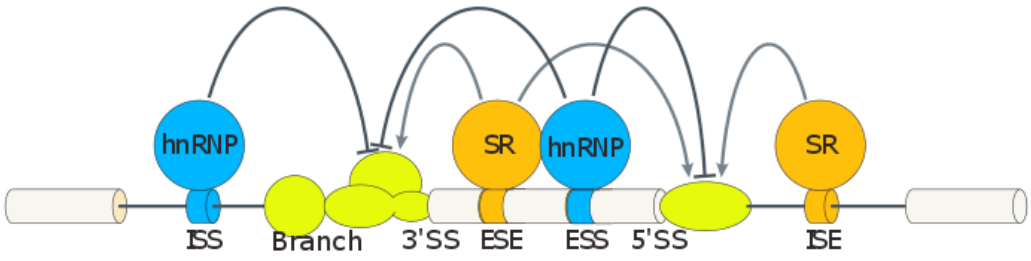


Fig 1.1 Illustration of alternative splicing regulatory network. AS is regulated by the interaction between *trans*-acting factors, such as SR proteins and hnRNPs, and *cis*-regulatory elements within nascent transcripts, including the well defined 5'/3' splice sites and branch sites as well as more diversified exonic/intronic splicing enhancers/silencers (ESE, ESS, ISE and ISS). This figure is adapted from Kornblihtt et al., 2013.

1.3 Types of alternative splicing

There are five major types of AS: exon skipping (or skipped exon, SE), intron retention (or retained intron, RI), alternative 5' splice site (A5SS), alternative 3'splice site (A3SS) and mutually exclusive exon (MXE) (Blencowe, 2006; Cieply and Carstens, 2015; Keren et al., 2010) (Fig 1.2). Intron retention is the most common type in lower metazoans, fungi, protozoa and plants (Keren et al., 2010; Kim et al., 2008). The relative prevalence of exon skipping gradually increases along the eukaryotic tree, which accounts for nearly 40% of AS events in higher eukaryotes (Alekseyenko et al., 2007; Keren et al., 2010). Alternative exons (or cassette exons) can also be spliced or skipped in tandem or spliced in a mutually exclusive manner at much lower frequency. Alternative selections of 5' or 3' splice sites within exon sequences are also frequent, accounting for about 18.4% and 7.9% of all known AS events, respectively (Blencowe, 2006; Keren et al., 2010).

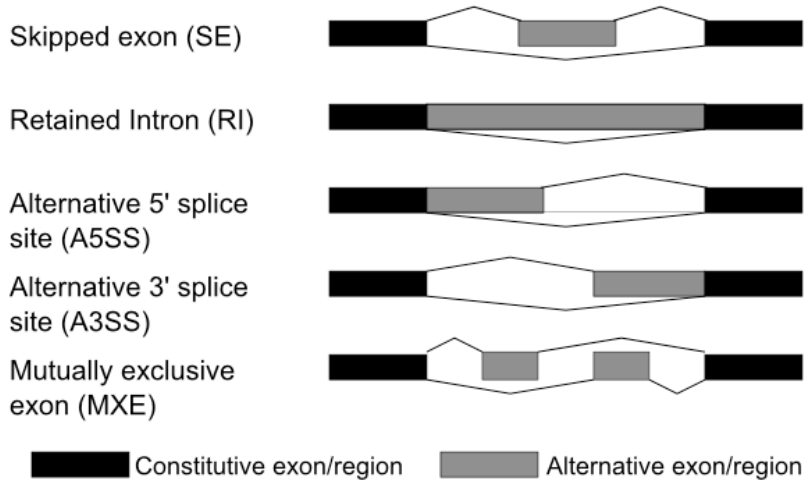


Fig 1.2 Different types of alternative splicing. In all five examples of alternative splicing, constitutive exons/regions are shown in black boxes and alternative exons/introns are shown in grey. Solid lines indicate splicing patterns.

Comparative genomic studies suggest significantly different molecular mechanisms in the regulation and evolution of different splicing types (McManus et al., 2014). In the case of exon skipping, both intron regions that flank the cassette exons are more conserved than the constitutively spliced exons (Ast, 2004; Sorek and Ast, 2003).

Cassette exons also tend to be shorter and have relatively weaker splice sites in comparison with constitutive exons (Clark and Thanaraj, 2002; Koren et al., 2007; McManus et al., 2014). Retained introns have also been found to be associated with weaker splice sites compared to constitutive introns, but they are largely dependent on the combined effects of many different *cis*-regulatory features (or IR code), including GC content, intron length, splice site strength, the location of the intron within gene body as well as characteristics of upstream and downstream exons (Braunschweig et al., 2014; Sakabe and de Souza, 2007). Exons with alternative 5' or 3' splice sites are exhibiting sequence features of an intermediate state between constitutive and alternative cassette exons, which suggest that they are likely to originate via mutations in ancestral constitutive exons, creating new splice sites to compete with the original sites (Sakabe and de Souza, 2007). Mutually exclusive exons are probably generated by exon duplication during evolution, which could be predicted according to exon sequence homology, splice site and reading frame conservation, as well as exon length (Pillmann et al., 2011; Pohl et al., 2013).

1.4 Evolution of alternative splicing

A primary goal of evolutionary biology is to understand changes driving the differences between species (Necsulea and Kaessmann, 2014). The rise of the genomic era provides a solid foundation for a systematic investigation of the molecular basis of phenotypic evolution. Moreover, recent development of large-scale high-throughput sequencing technologies has facilitated genome-wide and less biased comparisons of gene expression and alternative splicing mechanisms between divergent species, which represents a breakthrough in the field of molecular evolution (Necsulea and Kaessmann, 2014; Wang et al., 2009).

Changes in gene expression are thought to underlie many of the phenotypic differences between species. For example, species-specific expression patterns have been linked to mutations on *cis*- and *trans*-regulatory factors, and also to phenotypic divergence (Lynch and Wagner, 2008; Meireles-Filho and Stark, 2009; Wittkopp and Kalay, 2012). However, comprehensive studies of transcriptome profiles across multiple tissues form numerous vertebrate species have shown that the major source of variability of gene expression is the tissue in which they are measured.

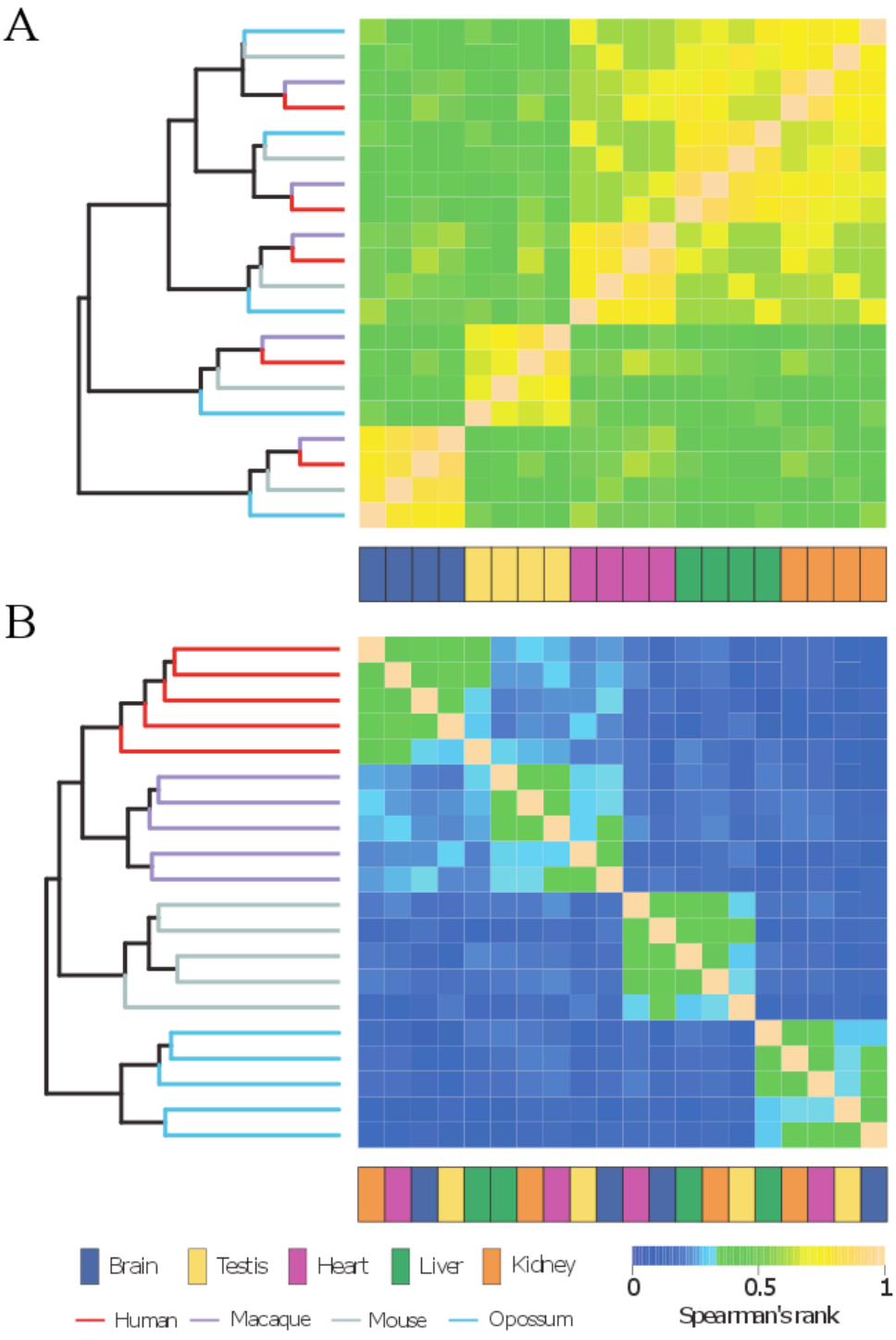


Fig 1.3 Global patterns of evolution for gene expression and alternative splicing. A. Hierarchical clustering of samples based on protein-coding gene expression levels, estimated as the number of reads per kilobase of exon per million mapped reads (RPKM). The heatmap represents Spearman's rank correlation

coefficients between pairs of samples. The sample clustering is represented as a tree with branch colors depicting different species. B. Hierarchical clustering shows that exon skipping frequencies are more similar between different tissues of the same species than between different species for a given tissue. This figure is adapted from Necsulea et al. 2014.

Tissue-dependent expression levels have been largely conserved during vertebrate evolution (Barbosa-Morais et al., 2012; Brawand et al., 2011; Necsulea and Kaessmann, 2014) (Fig 1.3A).

High-throughput sequencing technologies have revealed that the majority of multi-exon genes in mammals undergo AS (Wang et al., 2008). Global analysis in human-mouse and human-chimpanzee comparisons have identified frequent species-specific AS events (Calarco et al., 2007; Pan et al., 2005). Two recent comprehensive cross-mammal and tetrapod studies based on RNA-seq data for multiple organs have confirmed the pervasiveness of species-specific AS (Barbosa-Morais et al., 2012; Merkin et al., 2012). Moreover, these two studies have also shown a species-dominated pattern of clustering: exon skipping frequencies are more similar between different organs of the same species than between different species for a given tissue, which indicates that differences in alternative splicing contribute more significantly to phenotypic variation than gene expression (Necsulea and Kaessmann, 2014) (Fig 1.3B).

1.5 Dissection of *cis*-/*trans*-contributions in alternative splicing

Change in AS, one of the major driving forces to shape phenotypic diversity during evolution, could arise from the divergences in *cis*-regulatory elements and/or *trans*-acting RBPs. The divergences of the two factors with different extent of pleiotropic consequences undergo distinct evolutionary trajectories. Therefore, to better understand evolution in AS, it is important to distinguish the relative contribution of *cis*- and *trans*-effects.

Splicing quantitative trait loci (sQTL)

To globally identify the regulatory variants on AS, splicing quantitative trait loci (sQTL) study is one of the commonly used strategies, similar to expression

quantitative trait loci (eQTL) studies conducted for gene expression variant identification. In a typical sQTL study, the splicing patterns of target exons or the expression levels of different isoforms of the same gene are first estimated across a population, which are then correlated with the genetic variants (e.g. single nucleotide polymorphisms, or SNPs) from the same population. A significant association between SNP with exon splicing/isoform expression suggests that a regulatory mutation on AS is in high linkage disequilibrium with the SNP identified (Goncalves et al., 2012; Ongen and Dermitzakis, 2015; Zhao et al., 2013). Large-scale sQTL studies have found many genetic variants controlling transcriptome diversities or alternative splicing variations in human (Battle et al., 2014; Zhang et al., 2015).

Theoretically, QTL strategy can be used to identify regulatory variants either in *cis*- or in *trans*- effect. However, genome wide eQTL/sQTL studies that test the association between all SNPs against all expression/AS events are statistically underpowered for identifying variants. To overcome this problem, QTL studies typically focus on identifying *cis*-QTL by concentrating only on SNPs which are located close to target genes. This restriction removes the possibility of identifying regulatory variants in distal regions, which are most likely to act in *trans*-effect (Gibson and Weir, 2005; Goncalves et al., 2012). Therefore, the relative *cis*-/*trans*-contributions estimated using QTL methods could be biased toward higher *cis*-effects.

F1 hybrid system

Comparison between two homozygous parents (F0s) and their first generation (F1) hybrids s been shown to be a powerful system for studying regulatory divergence (Wittkopp et al., 2004a). This approach has been successfully used to dissect *cis*- and *trans*- regulatory changes in mRNA abundance (Goncalves et al., 2012; McManus et al., 2010; Tirosh et al., 2009a), RNA decay (Dori-Bachash et al., 2011) , translation (Khan et al., 2012) and recently alternative splicing (McManus et al., 2014). In F1 hybrids, per-mRNA form both parental alleles are subject to the same *trans*-regulatory environments, thus observed differences in allele-specific splicing reflect *cis*-regulatory divergence. If *trans*-regulation diverges between

parental strains, the collection of *trans*-factors in hybrids will be different from that in one or both parental strains, which results in different splicing pattern between F0s and F1 hybrids (Wittkopp et al., 2004a). Therefore, *trans*-acting contributions can then be inferred by comparisons of allele-specific differences in the hybrid to the splicing differences between the parental strains.

1.6 Previous studies on cis-/trans-contributions in alternative splicing

Several studies have tried to distinguish the relative contributions of *cis*- and *trans*-regulatory effects in different species. However, it remains under debate which factor plays more important role in the evolution of AS. Using sQTL strategy, Li *et al.* studied genetic variations of AS in *Caenorhabditis elegans* by comprehensively identifying quantitative trait loci affecting the differential expression of transcript isoforms n a large recombinant inbred population. In total, they found only 22 genes showing evidence for genetic variation of AS, 77% of which were locally regulated, indicating a predominant contribution of *cis*-effects (Li et al., 2010). A more recent study in *Drosophila* used RNA-seq to investigate splicing regulatory evolution among species and showed that whereas RI, A3SS and A5SS were primarily *cis*-directed, *trans*-effect had greater impacts on SE (McManus et al., 2014). In mammals, early work by Lin *et al*., based on the observation of higher sequence divergence flanking divergent SE events, suggested that changes in *cis*-regulatory elements made the major contribution to splicing divergence between human and chimpanzees (Lin et al., 2010). In the study by Barbosa Morais *et al*., the investigation of the splicing pattern of 13 human genes in a mouse strain carrying the majority of human chromosome 21 indicated that *cis*-regulatory changes were sufficient to drive the majority of species-specific pattern of exon inclusion/exclusion between human and mouse (Barbosa-Morais et al., 2012).

1.7 Aim of this study

Although previous mammalian studies implicated a predominant role of *cis*-divergence in evolution of divergent exon skipping events, a direct measurement of global contributions of *cis*- and *trans*-effects towards divergence of AS in

mammals is still lacking. Particularly given the different *cis*-/*trans*- contributions to different types of AS observed in *Drosophila,* it remains unclear whether the same holds true in mammals. In this study, we would like to: 1) globally investigate the relative contribution of *cis*- and *trans*-regulatory changes in mammals using F1 hybrid system; 2) determine the relative frequency of *cis*- and *trans*-regulatory changes for different splicing types; 3) understand the features associated with of *cis*- and *trans*-regulatory divergence.

To do this, we used RNA-seq to study splicing difference between *Mus musculus* C57BL/6J and *Mus spretus* SPRET/EiJ inbred mouse strains, as well as the allele-specific splicing pattern in their F1 hybrid. The two parental strains chosen in this study diverged ~1.5 million years ago, which resulted in about 35.4 million single nucleotide variants (SNVs) and 4.5 million insertion and deletions (indels) between their genome (Dejager et al., 2009; Keane et al., 2011). Such a high sequence divergence allows us to unambiguously determine the allelic origin for a large fraction of short RNA-seq reads, thereby enables accurate quantification of allelic pattern in F1 hybrids.

2. Materials and methods

2.1 Materials and experimental methods

2.1.1 Mouse liver sample collection and fibroblast cell culture

SPRET/EiJ mice were purchased from The Jackson Laboratories (Maine, USA) and C57BL6/J mice were obtained from Janvier (Le Genest-Saint-Isle, France). Both mouse strains were bred further in our animal house (VIB and Ghent University). C57BL6/J females were crossed with SPRET/EiJ males to yield F1 (BxS) hybrid mice. All mice were kept in an air conditioned, temperature controlled conventional animal house and obtained food and water ad libitum. Mice were used at the age of 8 weeks. All animal husbandry and experiments were approved by the local ethical committee (VIB and Ghent University). Mice were killed by acute CO intoxication, and livers were excised under sterile conditions. Livers were snap frozen in liquid nitrogen and kept at -80 °C until further use.

Adult mouse fibroblast cells were isolated and cultured according to the protocol from ENCODE project (https://genome.ucsc.edu/ENCODE/protocols/cell/mouse/Fibroblast_Stam_protocol.pdf) with modification of cell culture medium (RPMI 1640 Medium, GlutaMAX™ Supplement (Gibco, Life Technologies) with 10% FBS and 1% P/S). F1(BxS) mice used for fibroblast cell isolation were obtained as described before (Gao et al., 2013).

2.1.2 RNA sequencing

Total RNAs from cells were extracted using TriZOL reagent (Life Technologies) following manufacturer's protocol. Stranded mRNA sequencing libraries were prepared with 500ng total RNA according to manufacturer's protocol (Illumina). The libraries were sequenced in 2 x 100 +7 manner on HiSeq 2000 platform (Illumina).

2.1.3 RT-PCR and PacBio sequencing

Starting from 5 ug total RNA, polyA RNA was enriched using Dynabeads oligo-dT beads (Life Technologies), and reverse transcription (RT) was performed using

random hexamer and SuperScript II reverse transcriptase. PCR was followed using 1ul of RT product as template in 50ul of GoTaq PCR system (Promega). PCR primers were designed for amplifying the genomic region covering the alternative splicing events (Table E3). PCR program was as following, 4 min at 95 °C, followed by 28 cycles of 30 s at 95 °C, 30 s at 55 °C, and 45 at 72 °C, and a final elongation of 10 min at 72 °C. Different PCR products from the same RT product using different primers were then mixed and purified using Agencourt AMPure XP system (Beckman Coulter) and quantified by Qubit HS dsDNA measurement system (Life Technology). These mixed PCR products were then sequenced on PacBio RS SMRT platform according to the manufacturer's instruction.

Sequence reads from the PacBio RS SMRT chip were processed through PacBio's SMRT-Portal analysis suite to generate circular consensus sequences (CCSs). The CCSs were then mapped to a reference database containing alternative splicing isoforms from both alleles using BLAST with default parameters. The best hit was retained for each aligned sequence read. The reads with multiple best hits were discarded. PSI values were calculated as No. long-isoform-supporting-reads/ (No. long-isoform-supporting-reads + No. short-isoform-supporting-reads).

2.1.4 Minigene plasmids construction and in vitro minigene splicing reporter assay

Two C57BL/6J homologue genomic regions from Trim26 gene were amplified from 100 ng of C57BL/6J genomic DNA using 50 ul of Phusion PCR system (Thermo Scientific), respectively, with PCR program of 3 min at 98 °C, followed by 40 cycles of 30 s at 98 °C, 30 s at 57 °C, and 1 min at 72 °C, and a final elongation of 10 min at 72 °C. For the PCR of the first C57BL/6J homologue genomic region, the PCR primers were designed as following: one targeting on exon 1 (MG1-1-F: AAGCTGGCTAGCGTTTAAACTTAAGCTTGCTTGCTCAGGACCTACCCCG CGG); the other targeting on the region from the exon 2 to the adjacent region in intron 2 with four versions containing different combinations of SPRET/EiJ variants, respectively (MG1-1-no_variant-R: TAAACAGATACATAAATATAAGACCTGCTTCTGGTCATGCAGGGCTCCA AGCCACCAGGTGGAACGTCATCCGGGTC; MG1-1-insert-R: TAAACAGATACATAAATATAAGACCTGCTTCTGGTCATGCAGGGCTCCA

AGCCCAAGCTCCAACCAGGTGGAACGTCATCCGGGTC; MG1-1-SNV-R: TAAACAGATACATAAATATAAGACCTGCTTCTGGTCATGCAGGGCTCCA AGCCAGCAGGTGGAACGTCATCCGGGTC; MG1-1-SNV_insert-R: TAAACAGATACATAAATATAAGACCTGCTTCTGGTCATGCAGGGCTCCA AGCCCAAGCTCCAAGCAGGTGGAACGTCATCCGGGTC). For the PCR of the second C57BL/6J homologue genomic region, the PCR primers were designed as following: one targeting on intron 2 region adjacent to exon 3 with 5' overhang sequence overlapping with intron 2 part of the first PCR product (MG1-2-F: GCAGGTCTTATATTTATGTATCTGTTTATTTTTTTTTATTTATTTATCCTC AGAGTCATAGCCCGGGACAGCCACAGAGGA); the other targeting on exon 3 (MG1-2-R: TCTAGACTCGAGCGCGGATCCATATGGGGCGGATATCACTTGTGCAG).
The PCR products from above were purified using Agencourt AMPure XP system (Beckman Coulter). Then the overlapping PCR were performed between 15 ng of PCR products from the first and second Trim26 genomic regions using 50 ul of Phusion PCR system (Thermo Scientific) with PCR program of 3 min at 98 °C, followed by 8 cycles of 30 s at 98 °C, 30 s at 55 °C, and 1 min at 72 °C, then adding 10 nmol of MG1-1-F and MG1-2-R primers, followed by 27 cycles of 30 s at 98 °C, 30 s at 55 °C, and 1 min at 72 °C, and a final elongation of 10 min at 72 °C. Overlapping PCR products were purified using Agencourt AMPure XP system (Beckman Coulter), cut by NheI and XhoI restrict enzymes (NEB) and subcloned into pcDNA3.1/Hygro(+) vector (Invitrogen). Final minigene constructs were sequenced to verify the sequences and variants.

HEK293T and NIH3T3 cell lines (ATCC) were grown in DMEM (Invitrogen) with 10%FBS (Invitrogen). Cells were plated in 6-well plates and transfected using Lipofectamine 2000 (Invitrogen) according to manufacturer's protocol. TotalRNA were purified 48hr after transfection using TriZOL reagent (Invitrogen) and reverse-transcribed into ss-cDNA using oligo dT primer with superscript II reverse transcription system (Invitrogen). PCR were then performed using 50 ul of GoTaq PCR system with 1 ul of cDNA, 10 nmol of PCR primers T7-Promoter (TAATACGACTCACTATAGGG) and BGH-reverse (TAGAAGGCACAGTCGAGG), and PCR program of 2 min at 95 °C, followed by either 25 cycles (HEK293T) or 40 cycles (NIH3T3) of 30 s at 95 °C, 30 s at 54 °C,

and 1 min at 72 °C, and a final elongation of 10 min at 72 °C. PSI values of RT-PCR products were measured by Bioanalyser DNA 1000 chip (Agilent).

2.2 Computational methods

2.2.1 Reference sequences and gene annotation

The reference sequences and the Ensembl gene annotation of the C57BL/6J genome (mm10) were downloaded from the Ensembl FTP server (ftp://ftp.ensembl.org, version GRCm38, release 74). The SNVs and indels between C57BL/6J and SPRET/EiJ were downloaded from Mouse Genome Project website (http://www.sanger.ac.uk/). The vcf2diploid tool (version 0.2.6) in the AlleleSeq pipeline was used to construct the SPRET/EiJ genome by incorporating the SNVs and indels into the C57BL/6J genome (Rozowsky et al., 2011). The chain file between the two genomes was also reported as an output, which was further used with the UCSC liftOver tool.

2.2.2 RNA-seq read preprocessing and alignment

Flexbar was first used to trim the RNA-seq reads that pass the Illumina filter to remove library adapter sequences with parameters -f i1.8 -x 6 -u 0 -m 90 -k 90 -ae RIGHT (Dodt et al., 2012). Here in addition to the adapter sequences, we trimmed the first 6 bases on the 5' end to remove the sequence artifact due to the use of random hexamer as RT primers (-x 6). We retained only the read pairs with both reads of length >= 90 nucleotides after adapter removal (-m 90) and trimmed all of them from 3'end to the same length of 90 nucleotides (-k 90).

The remaining RNA-seq reads were aligned to the mouse genomes reference sequences (see above) using TopHat with default mapping parameter and Ensembl gene annotation (version 2.0.8) (Trapnell et al., 2009). For RNA-seq samples from parental strains, reads were aligned to the corresponding genome. For mixed (mock F1 hybrid) and F1 hybrid samples, reads were first aligned to both genomes, and then assigned to the parental allele with less mapping edit distance. The reads with equal mapping distance to both genomes were discarded and only the allele-specific reads were retained for further analysis. Genomic alignment coordinates for SPRET/EiJ were then converted to the corresponding locations in the C57BL/6J

reference genome using the UCSC liftOver tool and their chain files.

2.2.3 Alternative splicing analysis

Mixture of Isoforms (MISO) Bayesian Inference model (version 0.4.9) was used for quantification and comparison of alternative splicing events (Katz et al., 2010). The MISO algorithm counts the numbers of reads that are common to both isoforms and the reads that are exclusive to one isoform or the other, in order to estimate the Percent-Spliced-In (PSI) values in a given sample. The MISO events database (mm10) was downloaded from the MISO website (http://genes.mit.edu/burgelab/miso). Only the events from autosome were considered in this study. Splicing analysis was performed for the events supported with at least 20 RNA-seq reads (spliced-in + spliced-out) in all the replicate samples.

The Bayesian Factor (BF) was used as a measure of statistical significance for PSI difference. Based on prior work, BF > 5 in all the replicates and average $|\Delta PSI| > 0.1$ was used as the threshold for determining significant splicing difference between two parental strains or two alleles. To check whether our conclusion was sensitive to different thresholds, we also tried different cutoffs of $|\Delta PSI|$ values ($|\Delta PSI| > 0.0$, 0.05 and 0.15, respectively) corresponding to different FDRs.

Trans-regulatory divergence in alternative splicing was estimated using the method of Altman and Bland (Altman and Bland, 2003; McManus et al., 2014). In brief, the ratio of PSI values between strains were compared to allele-specific PSI ratios from F1 hybrid. The standard error of the difference in parental and allelic PSI ratios was calculated and used to derive Z-scores and p-values. Q-values were further calculated using the 'qvalue' module in R, and a same FDR cutoff as for *cis*-regulatory divergence was applied to determine *trans*-regulatory splicing divergence (Storey and Tibshirani, 2003).

2.2.4 False discovery rate estimation

To estimate the FDR, we used a method based on bootstrapped label permutation, as described before (Sterne-Weiler et al., 2013). In brief, for each value of x from 0.01 to 0.20 increasing by 0.01, we performed independent 100 bootstrapped label permutations of other replicates. For each of the 100 shuffled sets, we calculated the

number of events passing the threshold (false positives), *i.e.* BF > 5 in all the replicates and average |ΔPSI| >*x*. Then for each of the 100 permutations of each value *x*, the FDR was estimated as false positives divided by the number of real events passing the threshold, including both false positives and true positives.

2.2.5 Filter with mock F1 hybrid

In F1 hybrid, only the reads that could be unambiguously assigned to either genome were retained for estimation of alternative splicing. Therefore, the events with low variation density could have low coverage in F1 hybrid sample, or inconsistent PSI values between the parental strains and their F1 hybrid. To avoid potential errors, we mixed C57BL/6J reads and SPRET/EiJ reads to create mock F1 hybrid samples, which were then processed in the same way as the real F1 hybrid samples (*i.e.* mapping to both genomes and assignment to the parental alleles for identification of allele-specific reads according to edit distance). To evaluate the variations of PSI values for the events without assignment bias, we also down-sampled the C57BL/6J reads to the same coverage as the C57BL/6J allele in mock F1 hybrid, and then mapped these reads to C57BL/6j genome, and likewise for SPRET/EiJ reads.

To detect the events with inconsistent PSI values between the parental strains and the mock F1 hybrid, we applied a Z-value transformation, *i.e.* ΔPSI (the difference between the PSI values and the mock F1 hybrid PSI values) by a local standard deviation which we computed using a sliding window approach as following. In the down-sampled data, after sorting the events according to the total number of spliced-in and spliced-out reads used for computing the PSI values, we calculated for each data point the standard deviation of the respective values inside a window consisting 1% events. The local standard deviations were then smoothed using loess regression before we used them for calculating Z values and P-values in mock F1 hybrid sample. P-values were then adjusted using Benjamini-Hochberg method and a false discovery rate of 0.05 was applied to filter out the events with inconsistent PSI values after assignment.

2.2.6 C57BL/6J, CAST/EiJ and their F1 hybrid liver data analysis

The C57BL/6J, CAST/EiJ and their F1 hybrid liver data were downloaded from previous study and processed in the same way as our data (Goncalves et al., 2012).

Due to lower sequencing depth and lower density of sequence variants between these two strains, we pooled their dataset into two replicates for C57BL/6J, CAST/EiJ and their F1 hybrid, respectively. Specifically, ERR185942, ERR185943 and ERR120684 were pooled into C57BL/6J replicate 1; ERR120686, ERR120702 and ERR120704 were pooled into C57BL/6J replicate 2; ERR120692, ERR120694 and ERR120698 were pooled into CAST/EiJ replicate 1; ERR185946, ERR185947 and ERR185948 were pooled into CAST/EiJ replicate 2; ERR120672, ERR185940, ERR185941, ERR120678, ERR185945 and ERR120700 were pooled into F1 hybrid replicate 1; ERR185944, ERR120696, ERR185949, ERR185950, ERR185951 and ERR185952 were pooled into F1 hybrid replicate 2.

2.2.7 Control events without cis-regulatory divergence

To compare with the events with *cis*-regulatory divergence, we selected a separate group of AS events that passed the minimum threshold of 20 supporting reads but did not show splicing divergence between the two strains (BF < 1 and 0.05 < PSI < 0.95 in all three replicates as well as average $|\Delta PSI| < 0.05$).

2.2.8 Splicing site strength score analysis

For each splicing event, the nucleotide sequences of 5' and 3' splice sites were first extracted from the C57BL/6J and SPRET/EiJ genomes according to their locations (in .fasta format). These sequences were then uploaded to the "Analyzer Splice Tool" server (http://ibis.tau.ac.il/ssat/SpliceSiteFrame.htm) to calculate the splicing site strength score. For SE, RI and MXE, the strength scores of 5' and 3' splice site were combined.

2.2.9 Five mouse strains brain data analysis

The C57BL/6NJ, PWK/PhJ, WSB/EiJ, CAST/EiJ and SPRET/EiJ brain data were downloaded from previous study (accession number: ERP000614) (Danecek et al., 2012) and then MISO (version 0.4.9) was used for quantification of alternative splicing events in each data set.

3. Results

3.1 Divergence in alternative splicing between C57BL/6J and SPRET/EiJ

3.1.1 Quantification and comparison of alternative splicing pattern

To characterize the divergence of alternative splicing between C57BL/6J and SPRET/EiJ, we derived fibroblast cell lines from the two mouse strains and sequenced three biological replicates of polyA RNAs isolated from them on an Illumina HiSeq 2000/2500 platform (Fig 3.1). Paired-end sequencing resulted in an average of 169.4 million read pairs from each parental sample (Table 3.1). These reads were then mapped to the corresponding genome using splicing-aware alignment tools TopHat (Trapnell et al., 2009).

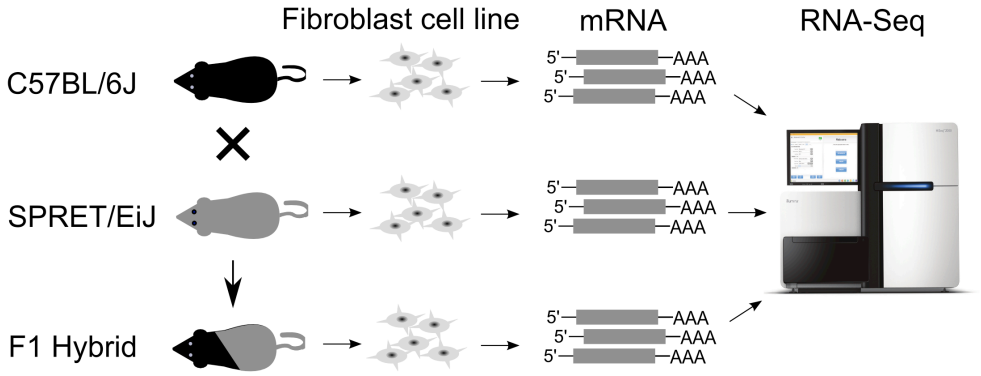


Figure 3.1. Study design. Fibroblast cells were isolated from adult C57BL/6J, SPRET/EiJ and the F1 hybrid mice and cultured. PolyA RNAs prepared from each cell line were sequenced on an Illumina HiSeq 2000/2500 platform.

After mapping, a previously developed Bayesian inference methodology - Mixture of Isoforms (MISO) - was applied for quantification (measured by Percent Spliced In, PSI) and comparison (ΔPSI) of alternative splicing events between the two parental strains C57BL/6J and SPRET/EiJ (Katz et al., 2010). Five major types of alternative splicing events were considered: SE, RI, MXE, A5SS and A3SS. A total

Table 3.1 - Summary of sequencing results

	Samples		Total read pairs [million]	Read pairs after trimming [million]	Concordantly mapped read pairs [million (%)][1)]	Assigned to allelic origin [million (%)][2)]
Fibroblast cell line	C57BL/6J	Replicate 1	168.6	150.5	129.4 (86.0%)	
		Replicate 2	156.0	137.7	120.2 (87.3%)	
		Replicate 3	175.0	156.1	133.3 (85.4%)	
	SPRET/EiJ	Replicate 1	176.4	157.1	125.2 (79.7%)	
		Replicate 2	183.1	164.2	129.7 (79.0%)	
		Replicate 3	157.7	140.4	110.7 (78.8%)	
	F1 hybrid	Replicate 1	374.9	338.6	287.9 (85.0%)	176.2 (61.2%)
		Replicate 2	405.6	366.1	309.8 (84.6%)	189.4 (61.1%)
		Replicate 3	383.6	346.1	288.5 (83.4%)	176.3 (61.1%)
Liver tissue	C57BL/6J	Replicate 1	156.2	141.7	116.7 (82.4%)	
		Replicate 2	157.0	143.0	115.1 (80.4%)	
	SPRET/EiJ	Replicate 1	164.2	149.6	113.8 (76.1%)	
		Replicate 2	175.4	159.0	122.1 (76.8%)	
	F1 hybrid	Replicate 1	268.1	242.6	204.0 (84.1%)	125.0 (61.3%)
		Replicate 2	301.2	273.9	218.6 (79.8%)	132.8 (60.8%)

1) The alignment rate was calculated as the number of concordantly mapped read pairs divided by the number of read pairs after trimming.

2) The percentage of F1 hybrid reads which could be unambiguously assigned to allelic origin was calculated as the number of read pairs assigned to allelic origin divided by the concordantly mapped read pairs.

of 30,199 annotated splicing events in mouse genome downloaded from MISO webpage (http://genes.mit.edu/burgelab/miso) were considered in this study (Table 3.2). To ensure higher accuracy, we required the quantification of a splicing event to be supported with at least 20 sequencing reads in all samples. In total, 11,818 events were retained for further analysis, including 5,615 SE, 1,768 RI, 696 MXE, 2,236 A3SS and 1,503 A5SS (Table 3.2).

We utilized the Bayesian factor (BF) as a measure of statistical significance for splicing difference (ΔPSI). After applying a threshold of BF > 5 in all the three replicates and average |ΔPSI| > 0.1, a criterion previously shown to maximize the number of significant events and minimize the false discovery rate (Sterne-Weiler et al., 2013), we identified in total 796 events showed significant splicing divergence between the two parental strains (Table 3.3 and Fig 3.2, false discovery rate (FDR) =2.5%). These divergent events covered all the five AS types (Table 3.3).

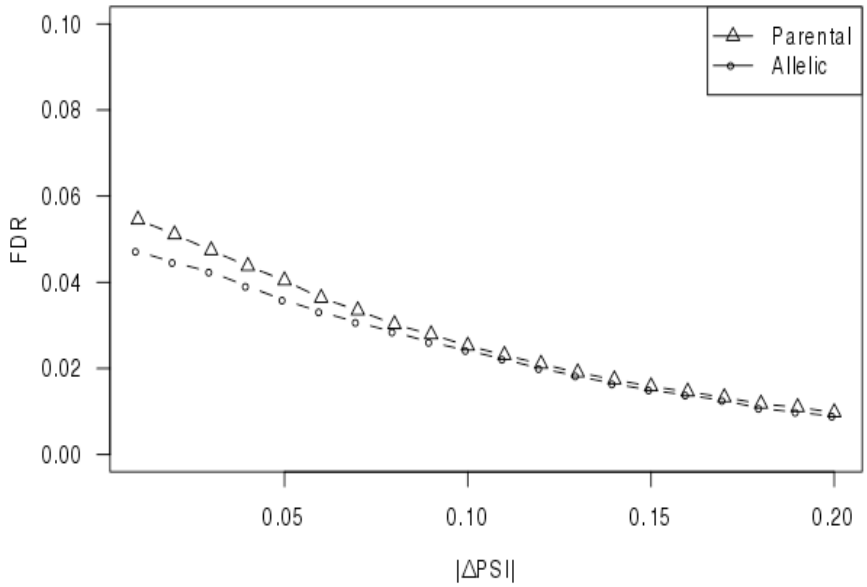


Fig 3.2 FDR estimation for each |ΔPSI| cutoff. FDR for parental (triangle) and allelic (circle) splicing comparison (y axis) was plotted against different |ΔPSI| cutoffs (x axis). For each value of *x* from 0.01 to 0.20 increasing by 0.01, we performed independent 100 bootstrapped label permutations of replicate 2 and replicate 3, respectively. For each of the 100 shuffled sets, we calculated the number of events passing the threshold (false positives), *i.e.* BF > 5 in all the replicates and average |ΔPSI| >*x*. Then for each of the 100 permutations of each value *x*, the FDR

was estimated as false positives divided by the number of real events passing the threshold, including both false positives and true positives.

Table 3.2. Summary of AS events in this study

Samples	Events	SE	RI	MXE	A3SS	A5SS	Total
	Total number	14,959	3,260	1,666	6,474	3,840	30,199
Fibroblast cell line	Expressed in parental strains[1)]	5,615	1,768	696	2,236	1,503	11,818
	Divergent between parental strains[2)]	418	124	54	101	99	796
	Retained for allelic comparison[3)]	2,667	953	245	1,158	779	5,802
	Retained divergent events between parental strains[4)]	203	69	21	63	61	417
	Divergent events in F1 hybrid[5)]	156	77	26	58	64	381
Liver tissue	Expressed in parental strains[1)]	4,088	1,590	245	1,650	1,186	8,759
	Divergent between parental strains[2)]	286	143	18	84	76	607
	Retained for allelic comparison[3)]	1,872	788	89	794	581	4,124
	Retained divergent events between parental strains[4)]	147	80	9	51	49	336
	Divergent events in F1 hybrid[5)]	121	58	6	38	47	270

[1)] Number of expressed events in parental strains with at least 20 spliced-in + spliced-out supporting reads in all replicates

[2)] Number of divergent event between parental strains at threshold BF>5 in all replicates and average |ΔPSI| >0.1.

[3)] After filtering using mock F1 hybrid, number of expressed events in F1 hybrid with at least 20 spliced-in + spliced-out supporting reads in all replicates

[4)] After filtering using mock F1 hybrid, number of retained divergent events between parental strains.

[5)] Number of divergent event between the two alleles in F1 hybrid at threshold BF>5 in all replicates and average |ΔPSI| >0.1

Table 3.3. Comparison of alternative splicing between C57BL/6J and SPRET/EiJ

		Total expressed events	Differential events (%)	P-value (Fisher's exact test)
Total number		11,818	796 (6.7%)	
Event type	SE	5,615	418 (7.4%)	
	RI	1,768	124 (7.0%)	
	A3SS	2,236	101 (4.5%)	
	A5SS	1,503	99 (6.6%)	
	MXE	696	54 (7.8%)	
Event effect	Non-coding regions*	3,400	317 (9.3%)	1.1e-10
	Coding regions	8,418	479 (5.7%)	
	-Frame-neutral event	4,235	273 (6.4%)	4.8e-3
	-Frame-shifting events	4,183	206 (4.9%)	

* Non-coding regions include non-coding genes and untranslated regions (UTRs) of coding genes.

3.1.2 Alternative splicing effects

Alternative splicing can affect either protein-coding sequences or non-coding ones (including non-coding genes and untranslated regions of coding genes). The former might be subject to stronger selection during evolution. Consistent with this, among the divergent AS events, the frequency of divergent splicing in non-coding regions was significantly higher than that in coding region (Table 3.3). Furthermore, within the set of divergent event in protein-coding regions, frame-preserving events were more likely to be divergent compared to frame shifting events (Table 3.3). These results demonstrated that in general AS with functional relevance was under stronger negative selection.

3.2 Predominant contribution of *cis*-regulatory variants underlying divergent AS between C57BL/6J and SPRET/EiJ

Alternative splicing divergence between species can arise from *cis*- and/or *trans*-regulatory differences. After identifying alternative splicing differences between the two parental strains, we next addressed the relative contributions of *cis*-regulatory differences in AS divergence using their F1 hybrids. *Trans*-acting contributions can then be inferred by comparing allele-specific differences in the hybrid to the splicing differences between the parental strains.

3.2.1 Quantification and comparison of alternative splicing pattern in F1 hybrid

Paired-end sequencing of polyA RNAs isolated from F1 fibroblast cell line resulted in on average 388.0 million read pairs for each of the three replicates (Table 3.1). The high density of sequence variants between the genomes of C57BL/6J and SPRET/EiJ allowed the unambiguous assignment of allelic origin for an average of 180.6 million read pairs in each replicate, which were used for further quantification of allelic alternative splicing (Table 3.1).

3.2.2 Filter with mock F1 hybrid

To avoid bias due to the potential misalignment of reads to the wrong allele, we first created a mock F1 hybrid RNA-seq dataset by mixing equal amounts of RNA-seq reads derived from the two parental strains. We then compared the PSI values of 11,818 expressed splicing events for both strains estimated based on the separate RNA-seq data from the parental strains to the allelic PSI values calculated using only those reads in the mock F1 dataset that could be unambiguously assigned to either allele. 2,595 events supported with <20 allelic reads in the mock dataset and 2,689 events with significant difference between the two PSI values for either allele were filtered out (Fig 3.3). Fig 3.4 showed that for the remaining 6,534 “well-behaved” events, both the PSI and ΔPSI values in the parental strains correlated well with the allele-specific values in mock F1 hybrid

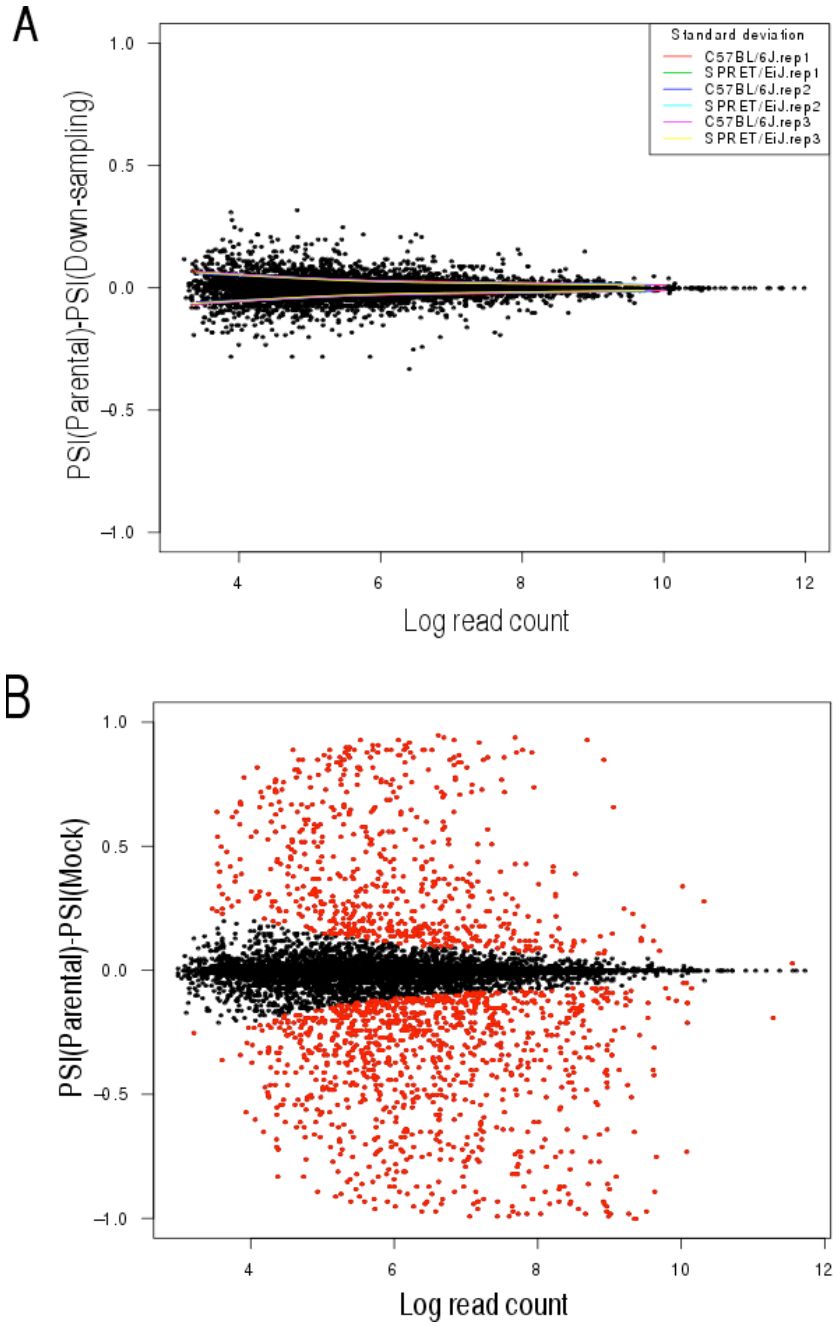
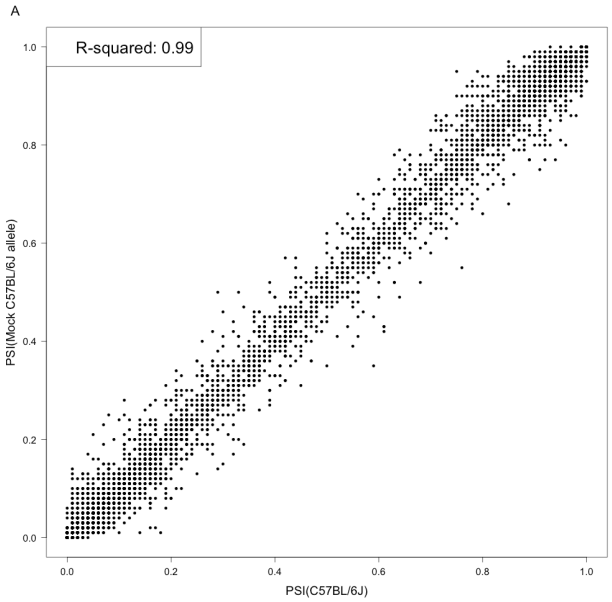
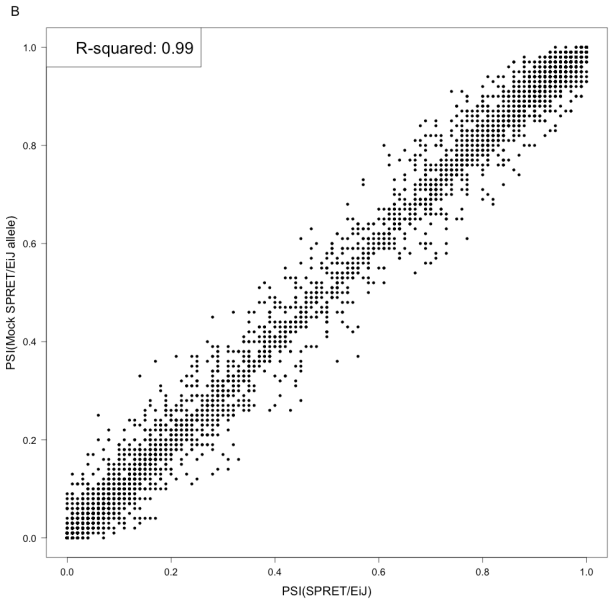


Fig 3.3 Illustration of data filtering based on mock F1 hybrid. A. MA plot comparing the PSI values in parental strains and their down-sampling datasets. The local standard deviation for each comparison was also indicated. B. MA plot comparing the PSI values in parental strains and those estimated based on mock F1 dataset. The red dots represented the outliers with inconsistent PSI values between parental strain and mock F1 dataset.


A
R-squared: 0.99
PSI(Mock C57BL/6J allele)
PSI(C57BL/6J)
0.0
0.2
0.4
0.6
0.8
1.0


B
R-squared: 0.99
PSI(Mock SPRET/EiJ allele)
PSI(SPRET/EiJ)
0.0
0.2
0.4
0.6
0.8
1.0

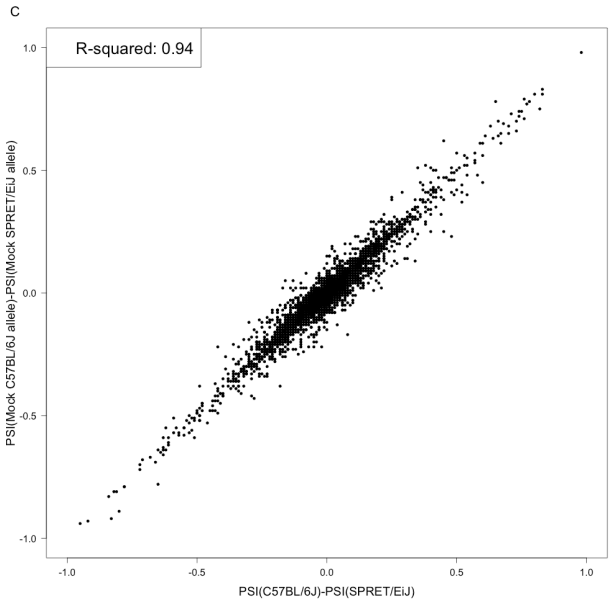

C
R-squared: 0.94
PSI(Mock C57BL/6J allele)-PSI(Mock SPRET/EiJ allele)
PSI(C57BL/6J)-PSI(SPRET/EiJ)
-1.0
-0.5
0.0
0.5
1.0

Fig 3.4 Correlation between parental strains and mock F1 hybrid after filtering. The PSI values for C57BL/6J (A), SPRET/EiJ (B) and their difference (C) correlated well between parental strains and mock F1 hybrid (R^2=0.99, 0.99 and 0.94, respectively) after filtering.

Out of 6,534 AS events, 5,802 supported with at least 20 sequencing reads in all three F1 hybrid sequencing replicates were retained for further analysis (Table 3.2). After applying the same threshold as that for parental strain, *i.e.* BF > 5 in all the three replicates and average |ΔPSI| > 0.1, we could detect a total of 381 divergent events between the two alleles in F1 hybrid (Fig 3.1, FDR=2.4%).

3.2.3 Independent validation of splicing difference using PacBio RS system

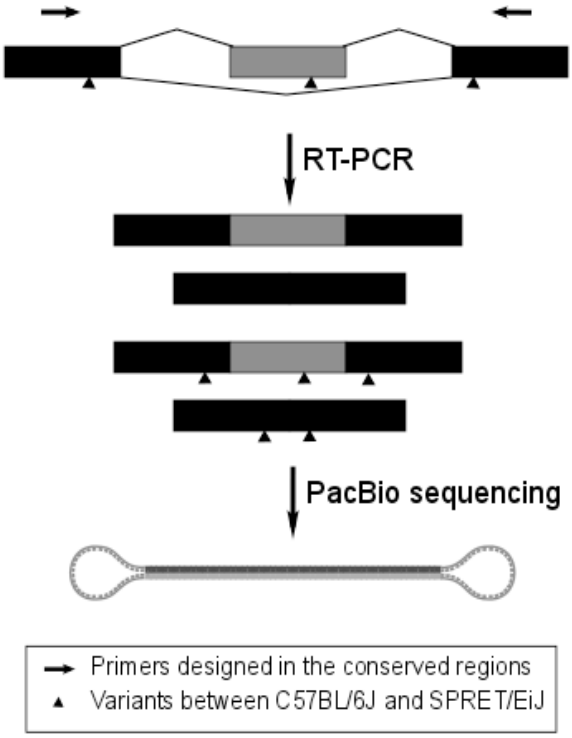


Fig 3.5 Illustration of PacBio sequencing of splicing event spanning cDNA PCR products. For each candidate event, RT-PCR primers were designed in the conserved regions of the constitutive exons to amplify both isoforms from the two alleles/strains. The PCR products were then sequenced at full length using PacBio RS system.

To assess the accuracy of our allele specific splicing analysis, we selected 20 candidate events consisting of all five different AS types (8 SE, 3 RI, 3 MXE, 2 A3SS, and 4 A5SS) for validation. Using long sequencing PacBio RS system, we deep sequenced the AS-spanning RT-PCR products amplified from either parental strains or F1 hybrid at full length using primers targeted at flanking constitutive regions with no sequence variant between the two strains (Fig 3.5) (Eid et al., 2009; Sun et al., 2013). Compared to Illumina sequencing, the longer read length facilitated the assignment of the PacBio reads to the parental alleles without any ambiguity in F1 hybrid. With the number of reads for each isoform of each strain/allele, the strain/allele-specific PSI could be calculated (Appendix Table 1). As shown in Fig 3.6, the splicing changes estimated in this way were significantly correlated with those determined by RNA-seq (R^2 =0.92).

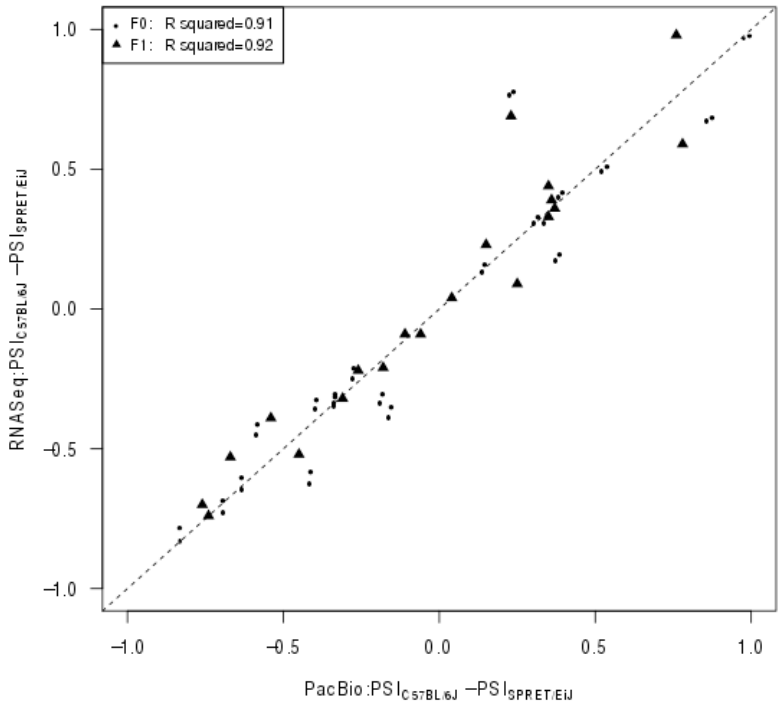


Fig 3.6 Scatterplot comparing parental splicing differences (dots, denoted as F0) or allelic splicing differences (triangle) estimated based on Illumina RNA-seq results (y axis) to those based on PacBio sequencing of splicing event spanning cDNA products (x axis) (R^2=0.91 and 0.92 for comparison of parental and allelic difference, respectively).

3.2.4 Dissection of *cis*- and *trans*-regulatory contributions in alternative splicing

We then compared the allelic divergent AS to the divergent AS between the parental strains. Out of 5,802 retained events, 417 had divergent regulation between parental strains, of which 255 and 62 exhibited *cis*- and *trans*- divergence, respectively (Fig 3.7). Fig 3.8 showed two representative examples for the divergent splicing events with predominant *cis*- and *trans*- contribution respectively. Such predominant *cis*-contributions were evident for all the five different types of AS (Fig 3.9).

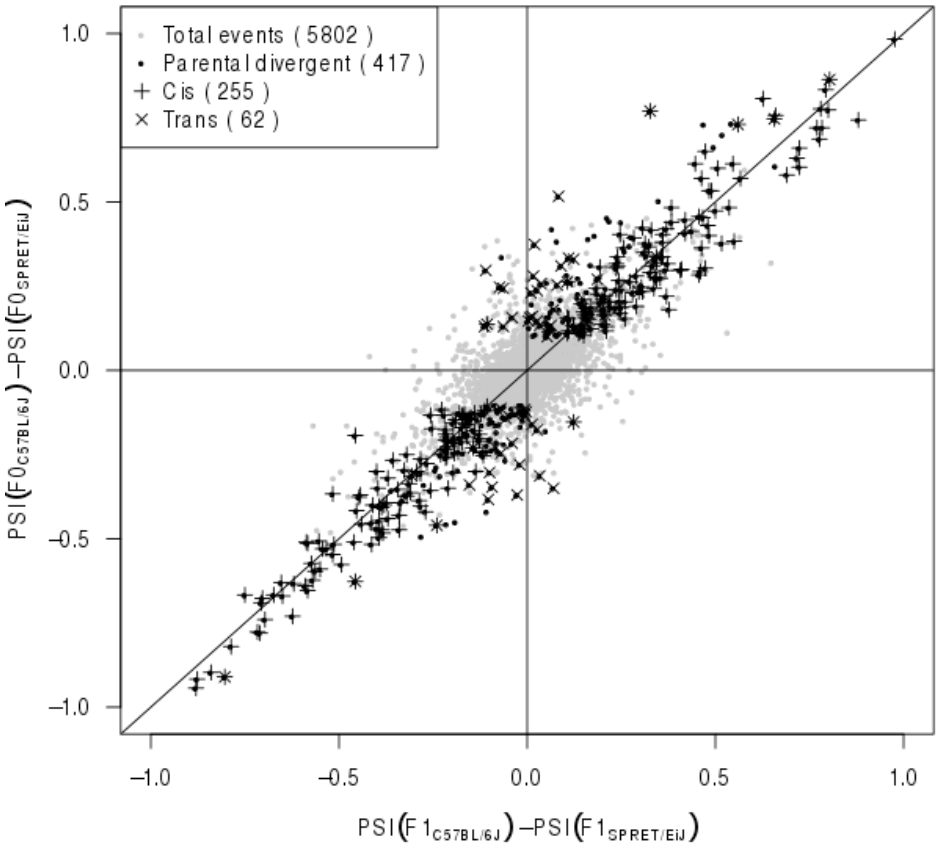


Fig 3.7 Scatterplot comparing splicing difference in parental strains (y axis) versus the allelic difference in F1 hybrid (x axis). After filtering using mock F1 hybrid, 5,802 AS events were expressed in F1 hybrid (grey dots). Among these, 417 AS events were divergent between parental strains (black dots), of which 255 (indicated as "+") and 62 (indicated as "x") exhibited significant *cis*- and *trans*-regulatory divergence, respectively.

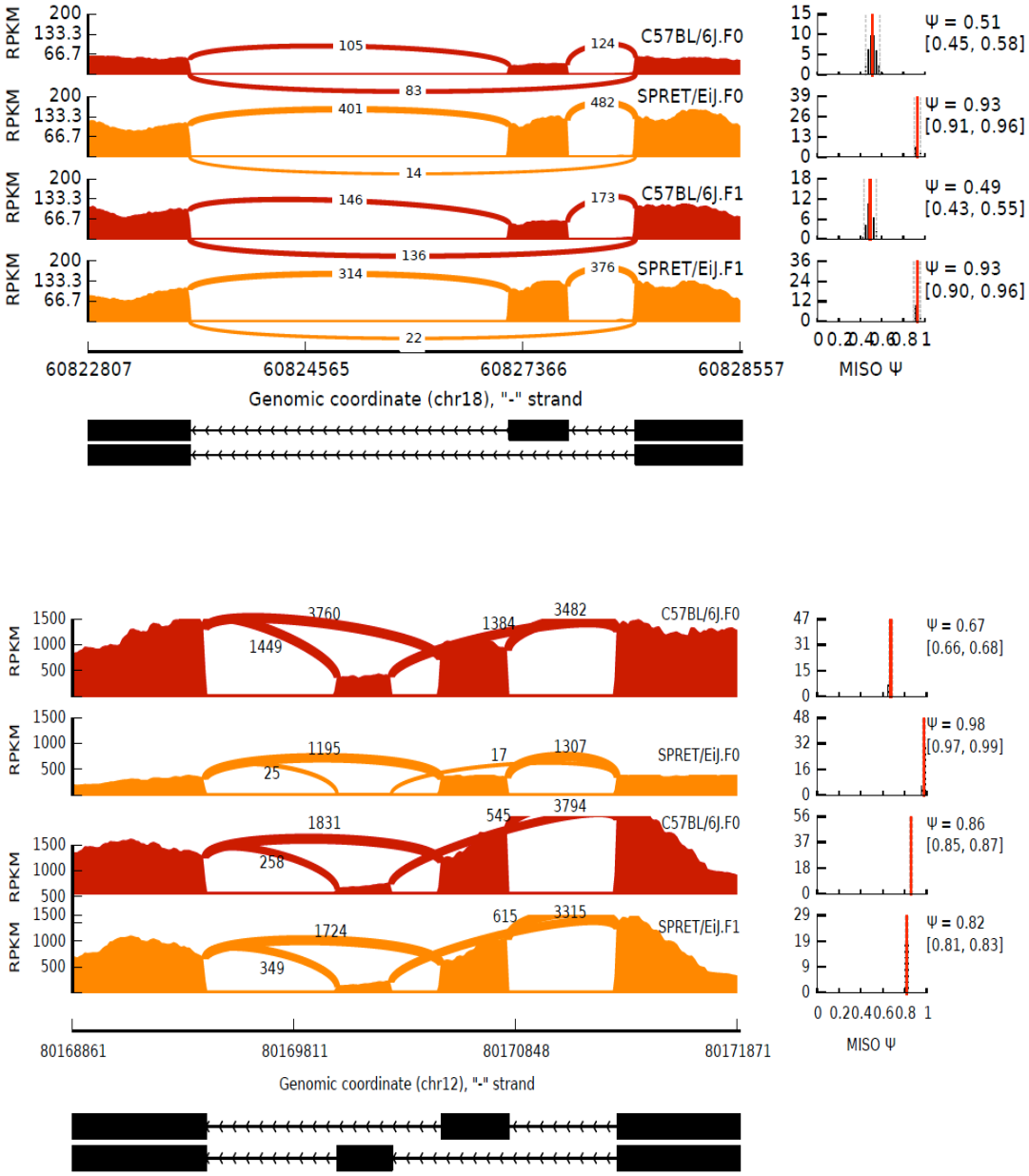


Fig 3.8 Examples of *cis*- (upper panel) and *trans*- (lower panel) regulatory divergence in alternative splicing. The RNA-seq read densities supporting the inclusion and exclusion of exons were shown in the left plot. The estimated PSI values and 95% confidence intervals were shown in the right plot.

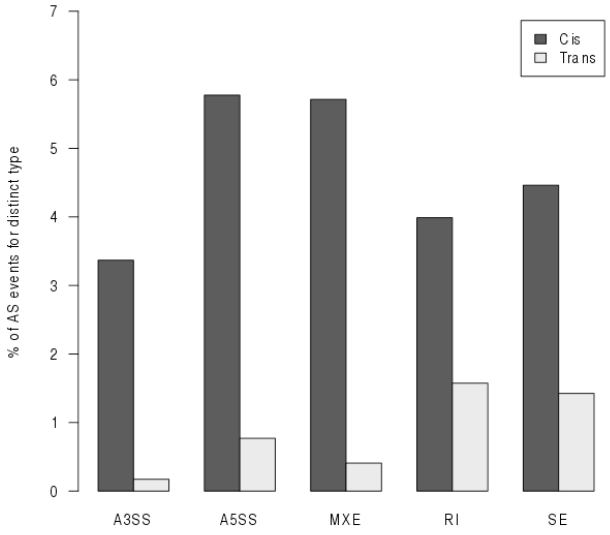
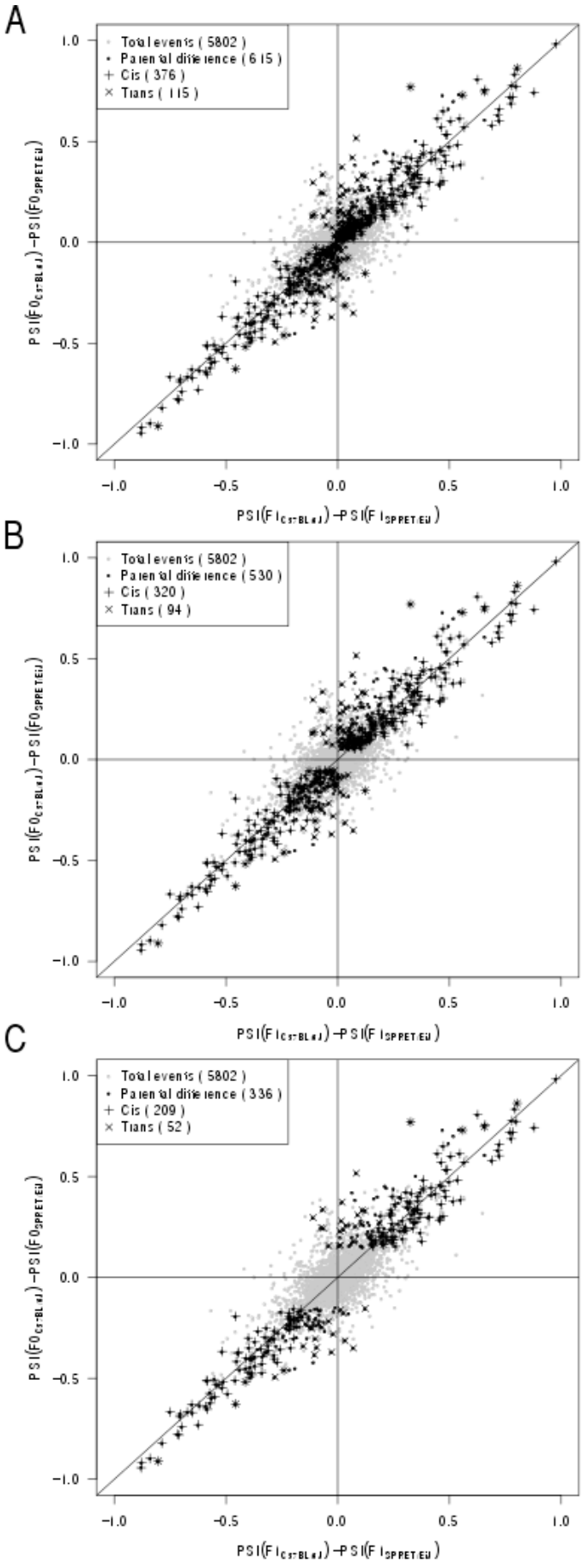


Fig 3.9 Percentage of *cis*- and *trans*-divergent events for the five AS types separately. *Cis*-regulatory contributions were predominant for all the five different types of AS.

3.2.5 Dissection using different thresholds

To check whether our conclusion was sensitive to different thresholds, we tried different cutoffs of |ΔPSI| values to determine the divergent AS events (Figure 3.1). As shown in Figure 3.10 A-C, *cis*-regulatory divergence always showed predominant contribution at different thresholds (|ΔPSI|>0.0, 0.05, and 0.15, respectively) and this trend also held true for all the five AS types (Fig 3.10 D-F). Furthermore, we also checked whether the contributions of *cis*-/*trans*- regulatory divergence were different for parental divergent events with different effect sizes (*i.e.* |ΔPSI|). For this, we grouped the 417 divergent events between the parental strains into 7 categories according to the |ΔPSI| values: (0.1, 0.2], (0.2, 0.3], (0.3, 0.4], (0.4, 0.5], (0.5, 0.6], (0.6, 0.7] and (0.7, 1.0]. As shown in Figure 3.11, while *cis*-regulatory divergence always played the predominant role in determining parental AS divergence with different effect sizes, its relative contribution slightly decreased with the decreasing effect size.


A
Total events (5802)
Parental difference (615)
Cis (376)
Trans (115)
B
Total events (5802)
Parental difference (530)
Cis (320)
Trans (94)
C
Total events (5802)
Parental difference (336)
Cis (209)
Trans (52)

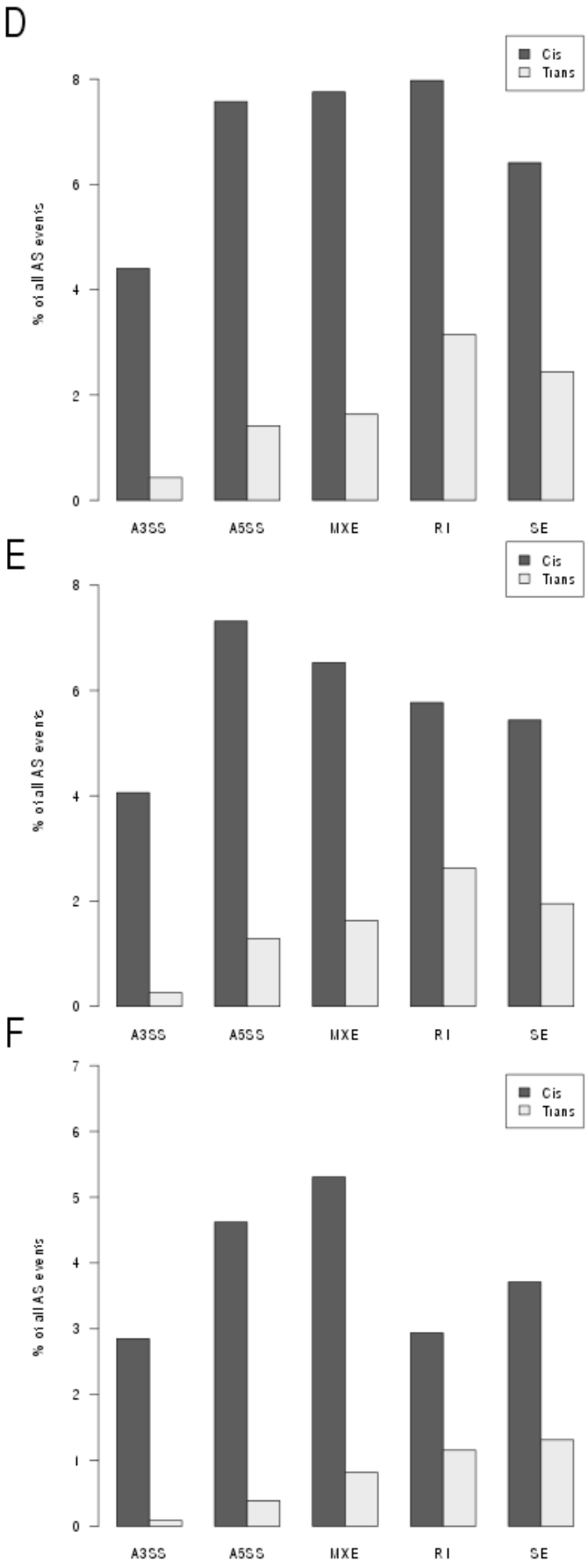

D
Cis
Trans
% of all AS events
0
2
4
6
8
A3SS
A5SS
MXE
RI
SE
E
Cis
Trans
% of all AS events
0
2
4
6
8
A3SS
A5SS
MXE
RI
SE
F
Cis
Trans
% of all AS events
0
1
2
3
4
5
6
7
A3SS
A5SS
MXE
RI
SE

Figure 3.10 Dissection of *cis*- and *trans*-regulatory contributions in alternative splicing at different |ΔPSI| cutoffs. A-C. Scatterplot comparing splicing differences in parental strains (y axis) versus the allelic differences in F1 hybrid (x axis) at different |ΔPSI| cutoffs (|ΔPSI|>0 (A), 0.05 (B) and 0.15 (C)). After filtering using mock F1 hybrid, 5,802 AS events were expressed in F1 hybrid (grey dots). Among these, 615 (A)/530 (B)/336 (C) AS events were divergent between parental strains (black dots), of which 376 (A)/320 (B)/209 (C) (indicated as "+") and 115 (A)/86 (B)/43 (C) (indicated as "x") exhibited significant *cis*- and *trans*-regulatory divergence, respectively. D-F. Percentage of *cis*- and *trans*-divergent events for the five AS types separately at different |ΔPSI| cutoffs (|ΔPSI|>0 (D), 0.05 (E) and 0.15 (F)).

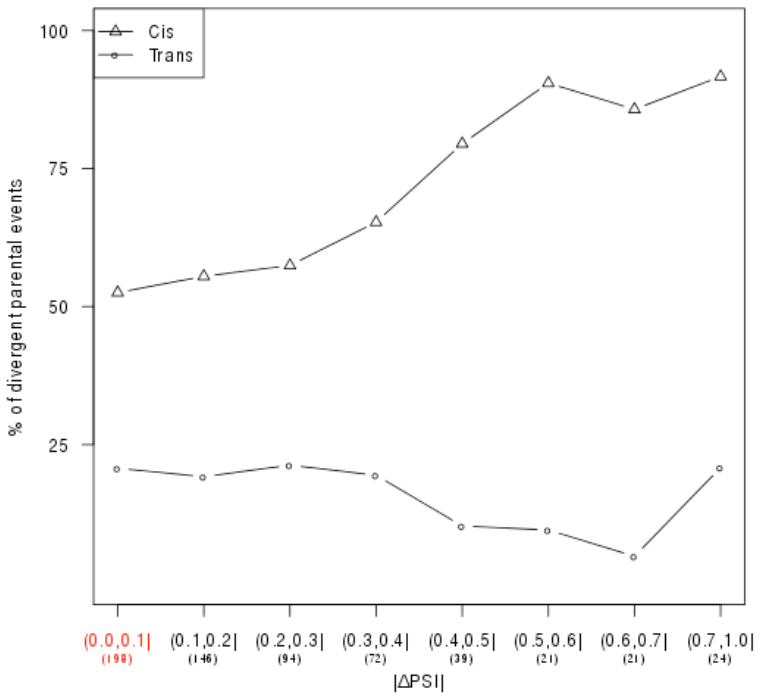


Fig 3.11 Contributions of *cis* (indicated as triangle)-/*trans* (indicated as circle)-regulatory divergence (y axis) to parental divergent AS events with different effect sizes (*i.e.* |ΔPSI|, x axis). 417 divergent events between parental strains (see Fig 2B) were grouped into 7 categories according to the |ΔPSI| values: (0.1, 0.2], (0.2, 0.3], (0.3, 0.4], (0.4, 0.5], (0.5, 0.6], (0.6, 0.7] and (0.7, 1.0]. The number of events in each category was marked.

3.2.6 Dissection using different statistical methods

To check whether our conclusion could be affected by the specific statistical methods applied in this study, we tried a different statistical test - Fisher's exact test - to determine the statistical significance in calculating splicing divergence. As shown in Fig 3.12, more divergent events in both parental and allelic comparisons could be identified using Fisher's exact test, and indeed nearly all the significantly divergent events found by MISO could also be detected using Fisher's exact test. We then compared the divergent AS identified by Fisher's exact test in parental strains to those in F1 hybrid. As shown in Fig 3.13, *cis*-regulation showed again predominant contributions for all the five AS types, demonstrating that our conclusion on predominant *cis*-contribution in splicing divergence was not test-dependent.

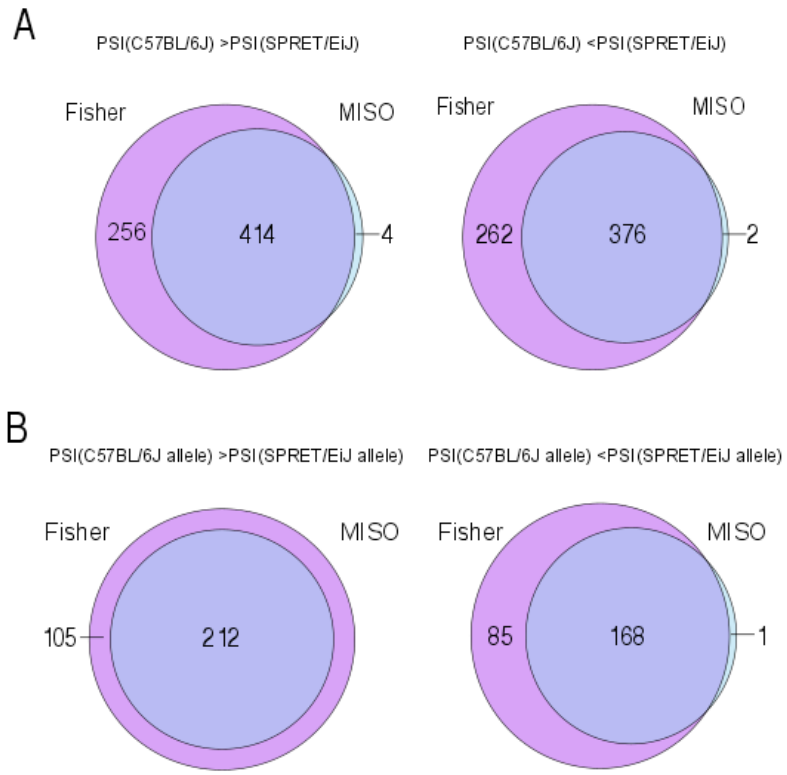


Fig 3.12 Venn diagram showing the overlap of the divergent events identified by Fisher's exact test and MISO in parental strains (A) and in F1 hybrid (B).

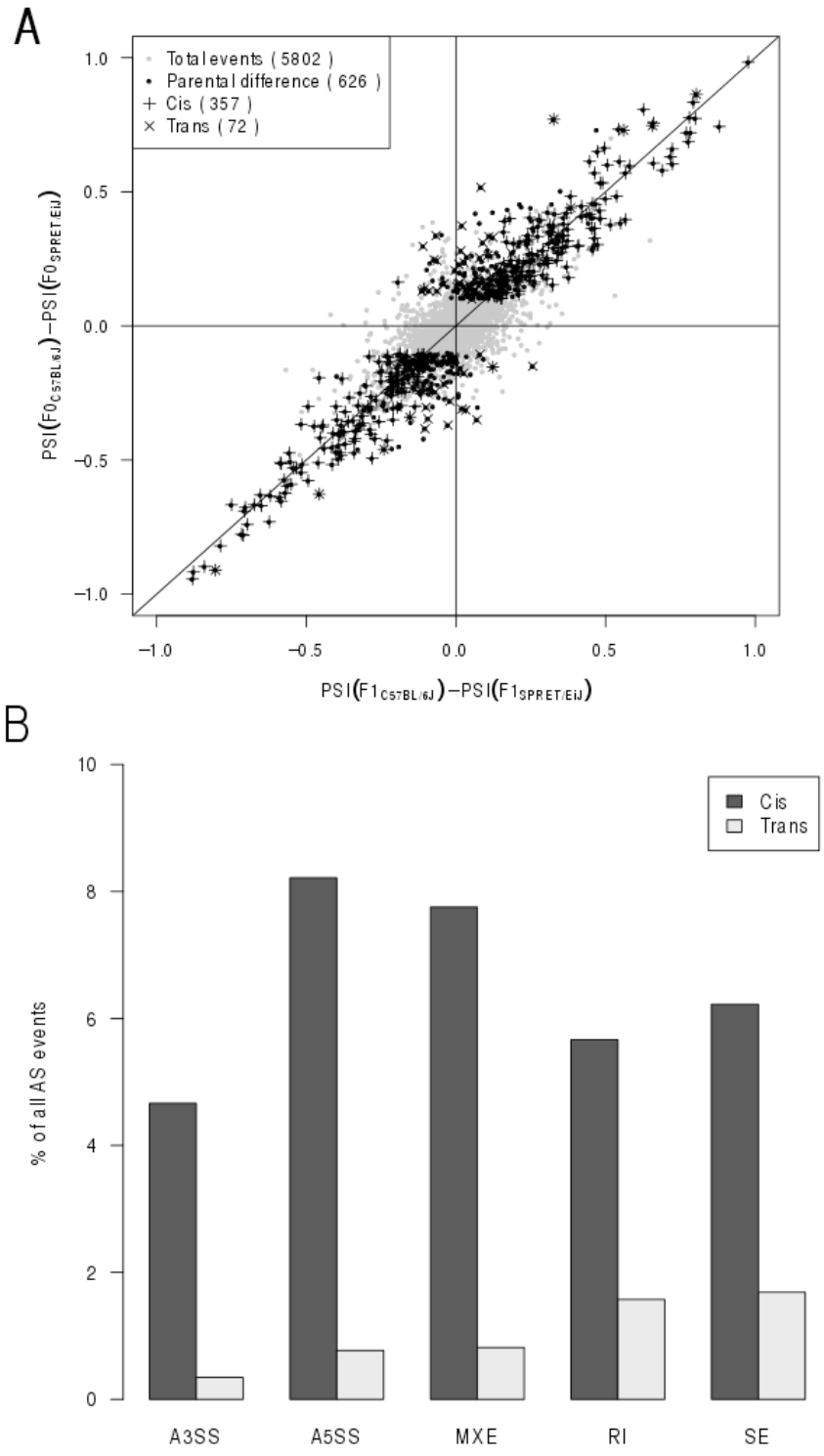


Figure 3.13 Divergent AS events identified using Fisher's exact test. A. Scatterplot comparing splicing difference in parental strains versus the allelic difference in F1 hybrid identified by Fisher's exact test. After filtering using mock F1 hybrid, 5,802 AS events were expressed in F1 hybrid (grey dots). Among these, 626 AS events were divergent between parental strains (black dots), of which 357 (indicated as "+") and

72 (indicated as "x") exhibited significant *cis*- and *trans*-regulatory divergence, respectively. B. Percentage of *cis*- and *trans*-divergent events for the five AS types separately using Fisher's exact test.

3.2.7 Dissection using tissue samples

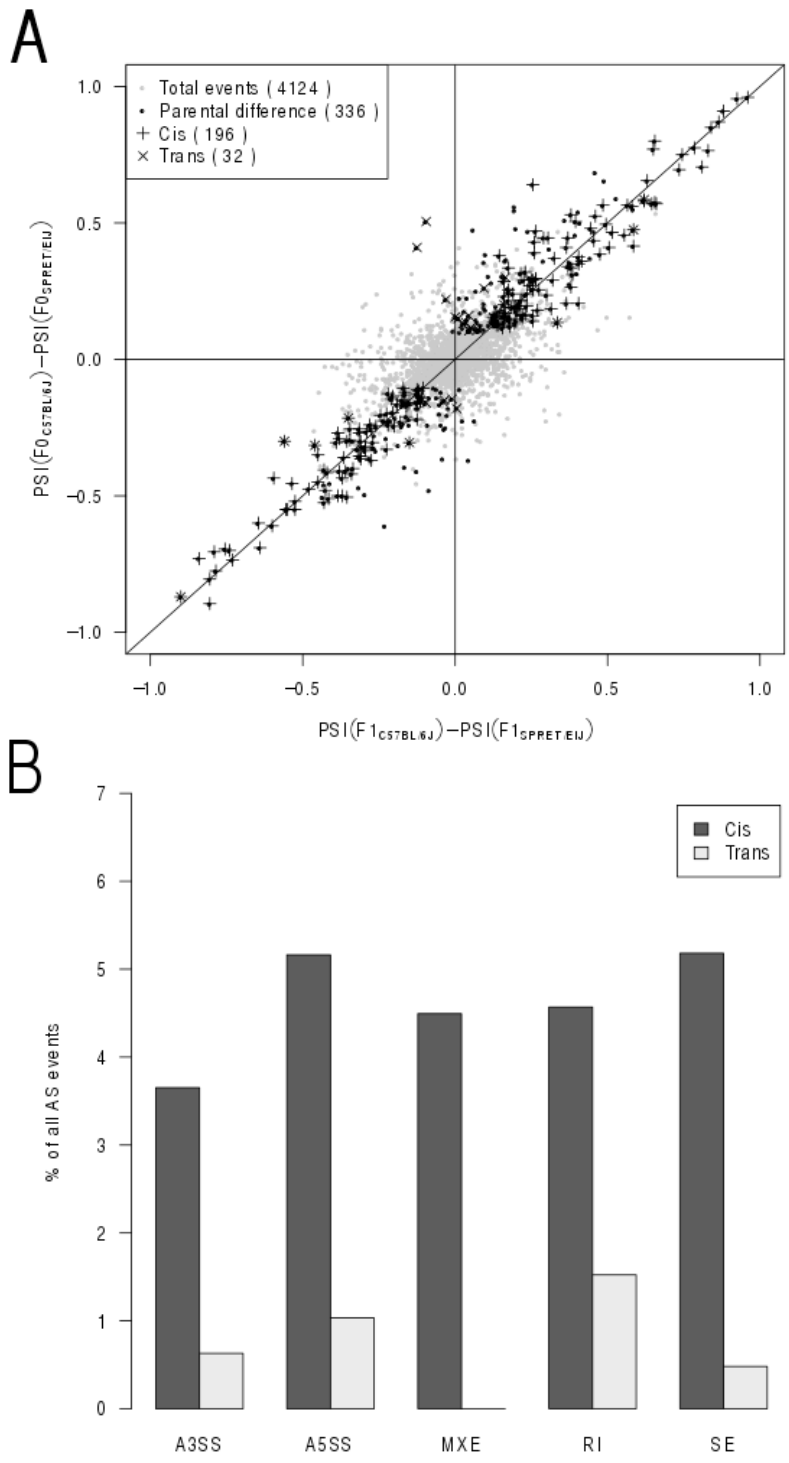


Figure 3.14 Dissection of *cis*- and *trans*-regulation in alternative splicing between C57BL/6J and SPRET/EiJ liver samples. A. Scatterplot comparing splicing difference between C57BL/6J and SPRET/EiJ liver samples versus their allelic difference in F1 hybrid liver sample. After filtering using mock F1 hybrid, 4,124 AS events were expressed in F1 hybrid (grey dots). Among these, 336 AS events were divergent between parental strains (black dots), of which 196 (indicated as "+") and 38 (indicated as "x") exhibited significant *cis*- and *trans*-regulatory divergence, respectively. B. Percentage of *cis*- and *trans*-divergent events for the five AS types separately.

To check whether our conclusion from cultured cells could be extended to mouse tissues, we performed RNA-seq on two replicates of the liver samples from C57BL/6J, SPRET and their F1 hybrid, respectively (Table 3.1). Out of 8,759 AS events expressed in the parental samples, 607 were identified as significantly divergent between the parental strains (BF>5 in both replicates and average $|\Delta PSI|>0.1$). After the similar filtering based on mock F1 dataset, 4,124 and 336 total expressed and divergent events retained, respectively (Table 3.2). Then by applying the same threshold as that for parental strains, we detected 270 divergent events between the two alleles in F1 hybrid (Table 3.2). Finally we compared the allelic divergent to the parental divergent AS. Out of 336 parental divergent events retained after filtering, 196 and 38 exhibited significant *cis*- and *trans*-regulatory divergence, respectively (Fig 3.14A). Such predominant contributions of *cis*-regulatory divergence were also evident for all the five splicing types (Fig 3.14B).

3.2.8 Dissection using different mouse strains

To check whether our conclusion could be generalized to other mouse strains, we compared the AS patterns between C57BL/6J and CAST/EiJ using previously published dataset (Goncalves et al., 2012). These two strains diverged about 1 million years ago, resulting in 17.7 million SNVs and 2.7 million indels between their genome sequences (Keane et al., 2011). The lower density of sequence variants, together with shorter sequencing reads (2x72), allowed in their F1 hybrid RNA-seq data only about 30.2% of the mappable reads to be unambiguously assigned to their

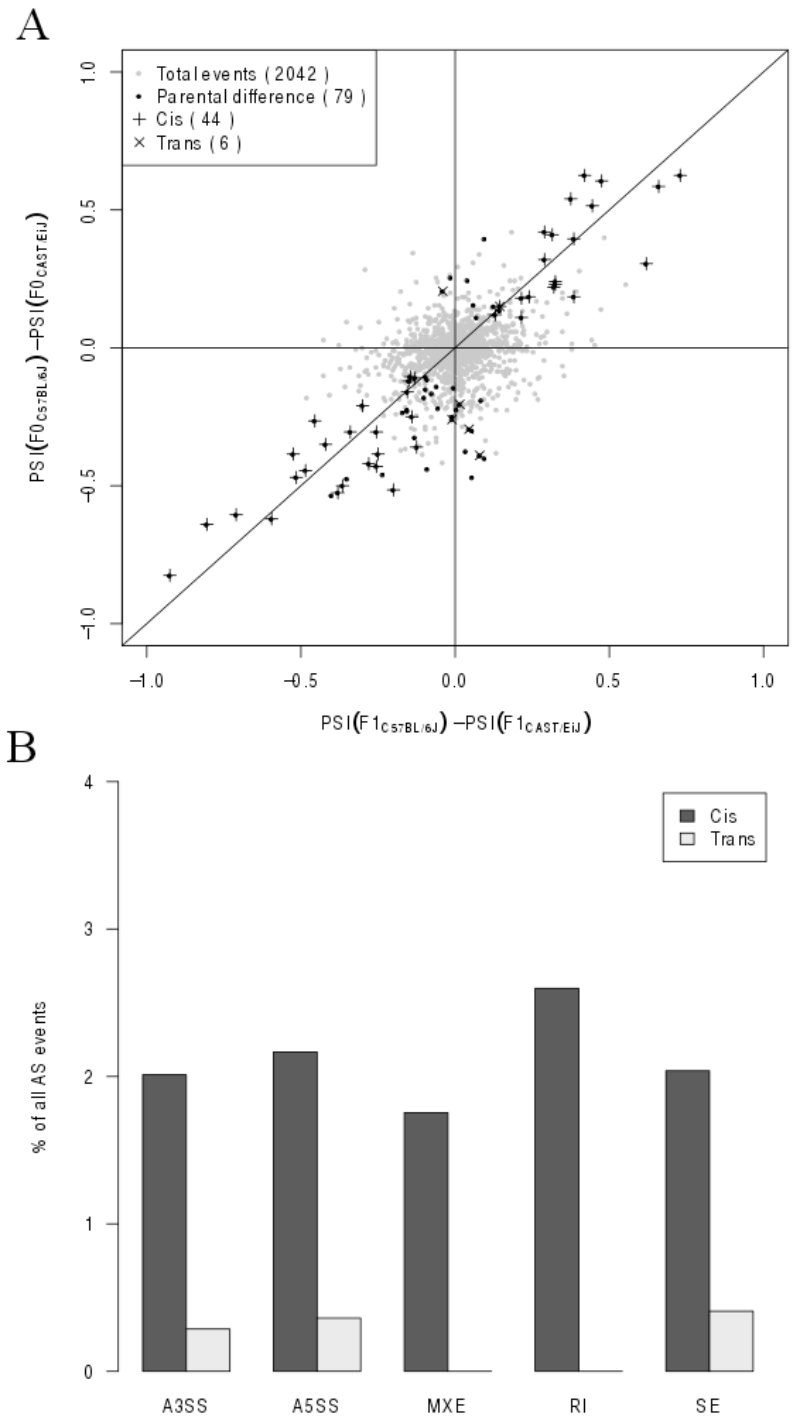


Figure 3.15 Dissection of *cis*- and *trans*-regulation in alternative splicing between C57BL/6J and CAST/EiJ. A. Scatterplot comparing splicing difference between C57BL/6J and CAST/EiJ versus their allelic difference in F1 hybrid. After filtering using mock F1 hybrid, 2,042 AS events were expressed in F1 hybrid (grey dots). Among these, 79 AS events were divergent between parental strains (black dots), of which 44 (indicated as "+") and 6 (indicated as "x") exhibited significant *cis*- and

trans-regulatory divergence, respectively. B. Percentage of *cis*- and *trans*-divergent events for the five AS types separately.

parental alleles (compared to about 61.1% in our F1 hybrid of C57BL/6J and SPRET/EiJ, Table 3.1).

Therefore, to obtain a sufficient number of reads for accurate PSI quantification, we pooled the data from three individuals together and generated two replicate datasets for C57BL/6J, CAST/EiJ and their F1 hybrid, respectively. We then performed the same analysis as described before. Although the absolute numbers of divergent events identified both between parental strains and between alleles in F1 hybrid were understandably lower, the predominant contribution of *cis*-regulatory divergence (44 *cis* versus 6 *trans*) was still evident (Fig 3.15A), and this trend held true for all the five splicing types (Fig 3.15B). This implied that, in general, predominant *cis*-contribution in the evolution of mouse alternative splicing.

3.3 Genomic features that correlate with *cis*-regulatory AS divergence

3.3.1 Sequencing variants density

Cis-regulatory divergence should result solely from sequence variants in pre-mRNA sequences, particularly those residing close to the affected splicing events. To investigate this, we calculated the frequencies of SNVs and indels in the regions flanking the AS events with or without *cis*-regulatory divergence (Fig 3.16). As shown in Fig 3.17, compared with those without *cis*-divergence (control events, see Materials and Methods), the regions flanking AS events with *cis*-divergence contained significantly higher density of sequence variants between the two strains (see also Appendix Fig 1 for the comparison of different AS types separately).

3.3.2 Sequencing variants at splice sites

We then checked how sequence variants at the exact splicing sites could contribute to the events with *cis*-regulatory divergence. As shown in Fig 3.18, 36.2% of these events with *cis*-regulatory divergence had at least one sequence variants at the

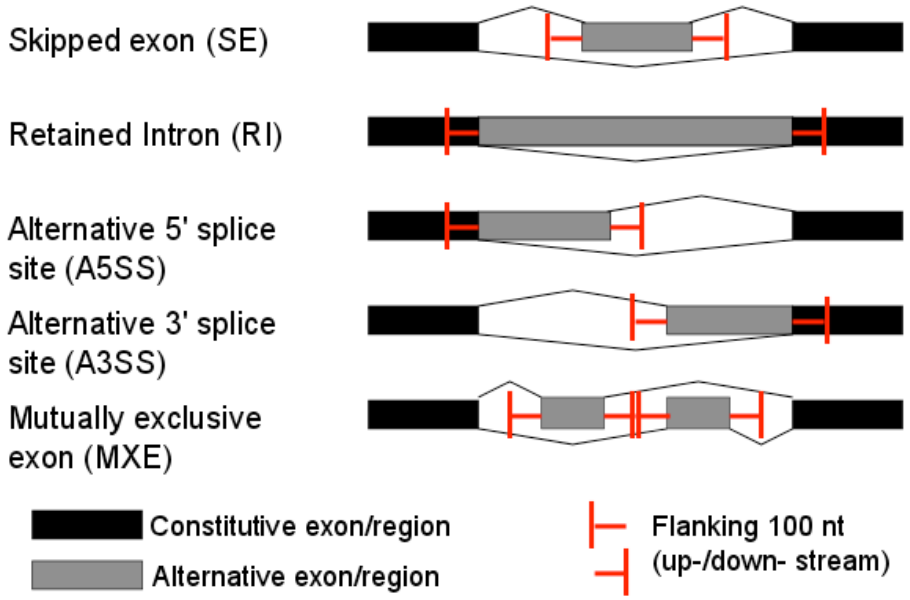


Figure 3.16 Illustration of the regions flanking the AS events. For SE, the alternative exons and their flanking 100nt intron sequences were considered; For RI, the retained introns and their flanking 100nt exon sequences were considered. For A3SS or A5SS, the alternative exon regions and their flanking 100nt exon/intron sequences were considered. For MXE, both alternative exons and their flanking 100nt intron sequences were considered.

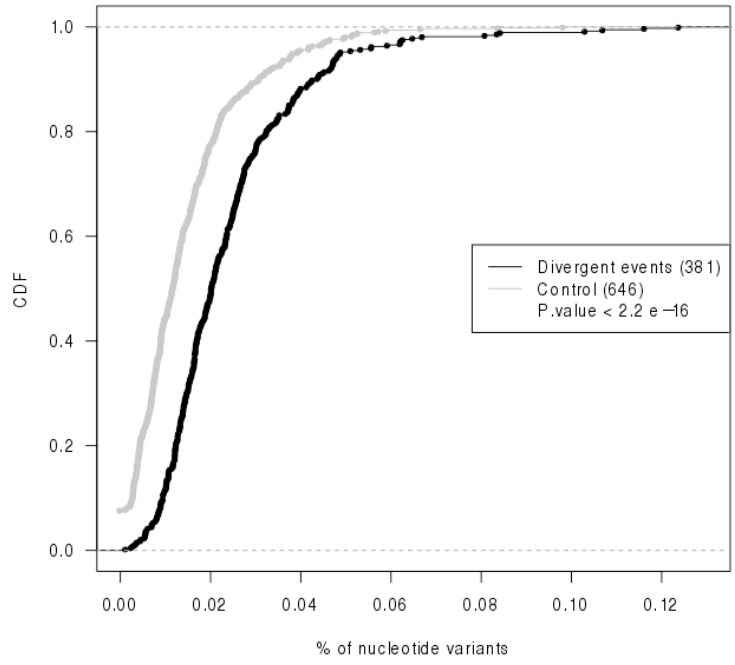


Fig 3.17 The cumulative distribution function (CDF) of frequencies of nucleotide variants in the AS flanking regions for the events with *cis*-regulatory divergence (black) and controls (gray). Compared with controls, the events with significant *cis*-regulatory impact had higher sequence divergence in the flanking regions. The p-values were calculated by the Mann-Whitney U test.

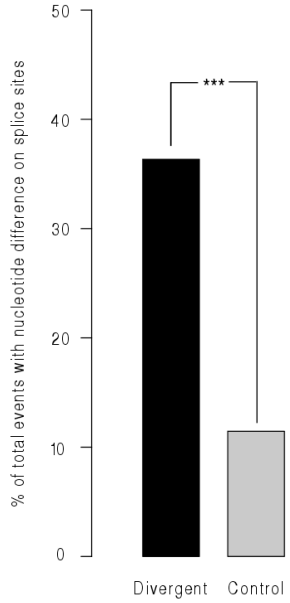


Fig 3.18 36.2% and 11.5% of the events with significant *cis*-regulatory divergence (black) and control events (grey) had sequence divergence at their exact splice sites, respectively.

respective splicing sites, compared to 11.5% of control events (P.value=9.2e-14, Fisher's exact test, see also Appendix Fig 1 for the comparison of different AS types separately). Sequence variants at splice sites could regulate alternative splicing by affecting splice site strength – the probability that the splice sites could be recognized by the spliceosome (McManus et al., 2014). To investigate how sequence variants at the splicing sites could affect splicing site strength, we calculated the splicing site strength score for the two alleles containing variants at the exact splice sites (Material and Methods) and compared the allelic difference of such score between the events with *cis*-regulatory divergence and those without. As shown in Fig 3.19, the sequence variants at the splicing sites of *cis*-divergent events affected the splicing site strength

more than those at splicing sites of control events. As expected, variants changing the canonical GU/AG splicing donor/acceptor sites severely affected the splicing site strength, which resulted in complete functional abortion of the corresponding splicing site, as exemplified in Fig 3.20. Importantly, the same analysis of the liver data showed a similar correlation of all these genomic features (Fig 3.21). Taken together, sequence variants at the canonical splicing sites could affect splicing site strength and thereby lead to divergent AS.

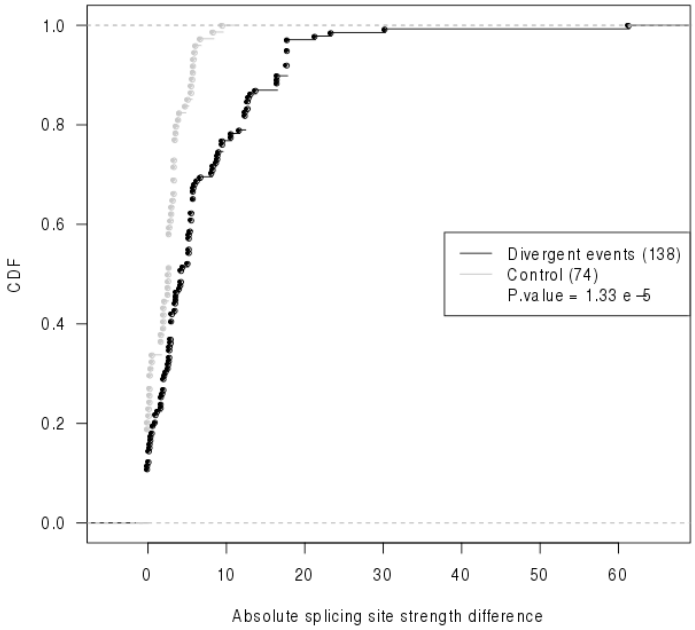


Fig 3.19 CDF of allelic differences in splicing site strengths due to sequence variants at the exact splicing sites plotted for *cis*-regulatory divergent events (black) and control events (gray), separately. The splicing site strengths changed more in the events with *cis*-regulatory events than in those without. The p-values were calculated by the Mann-Whitney U-test.

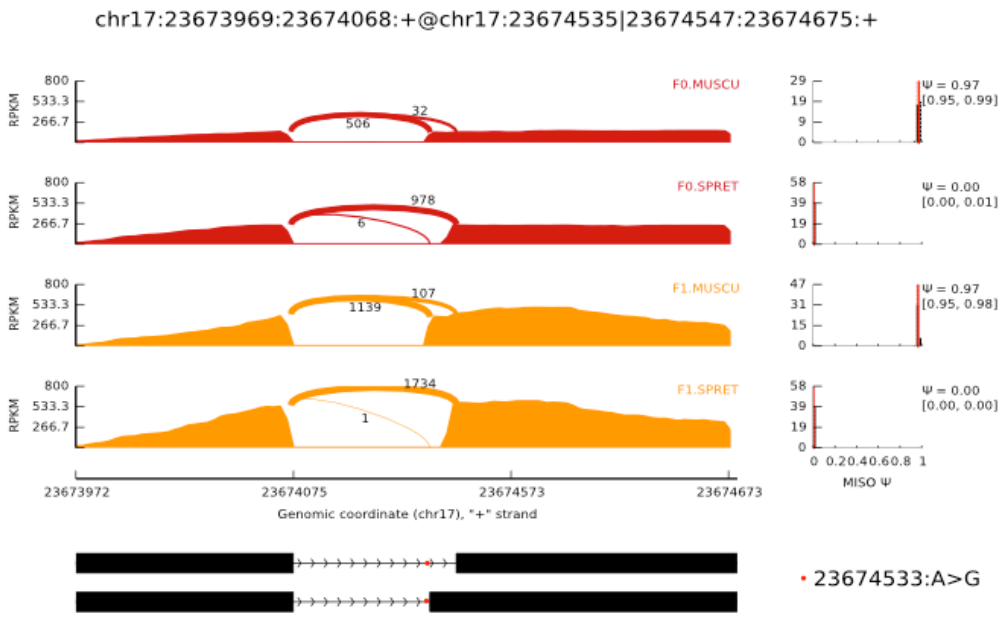
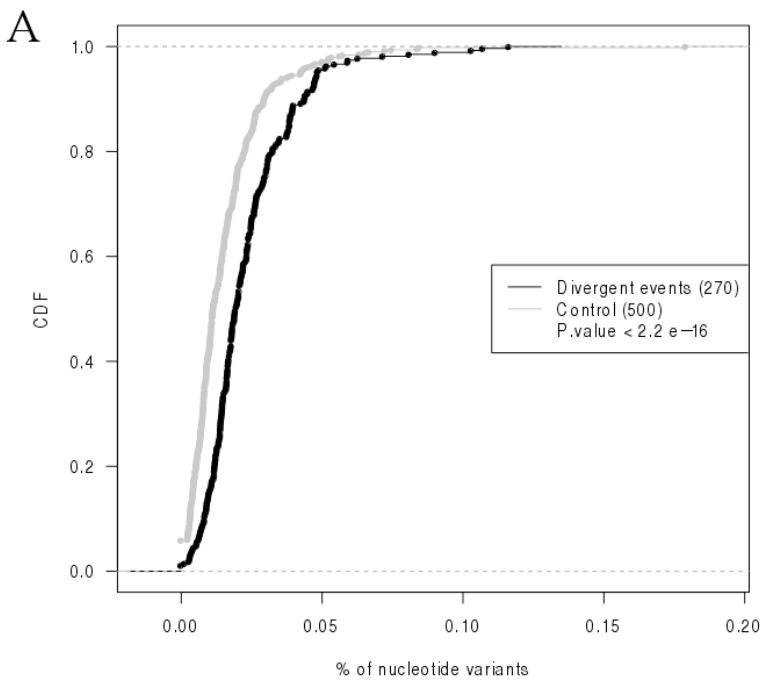


Fig 3.20 An example showing that a SNV at the canonical GU/AG sites (indicated as an arrow) resulted in complete functional abortion of the corresponding splice sites. The substitution of the AG to GG in SPRET/EiJ disrupted the splicing site and thereby facilitated the use of a downstream splicing acceptor.



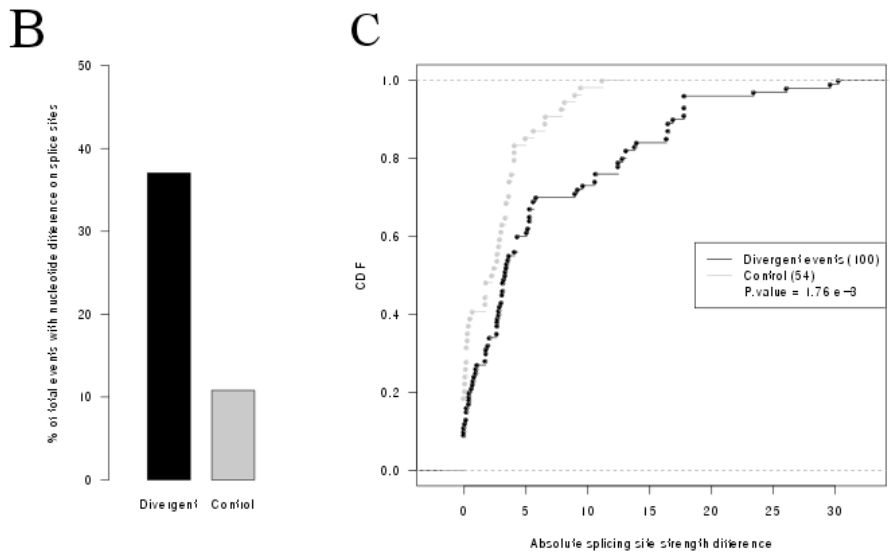


Fig 3.21 Genomic features that correlate with *cis*-regulatory alternative splicing divergence identified in the liver sample. A. CDF of frequencies of nucleotide variants in the AS flanking regions for the events with *cis*-regulatory divergence (black) and controls (gray) identified in liver sample. Compared with controls, the events with significant *cis*-regulatory impact also had higher sequence divergence in the flanking regions. B. In liver sample, 37.0% and 10.8% of the events with significant *cis*-regulatory divergence (black) and controls (grey) had sequence divergence at the exact splice sites, respectively. C. CDF of allelic differences in splicing site strengths due to sequence variants at the exact splicing sites plotted for *cis*-regulatory divergent events (black) and controls (gray) identified in liver sample. The splicing site strengths changed more in the events with *cis*-regulatory events than in those without

3.3.3 Sequencing variants beyond splice sites

Cis-regulatory variants could affect as well the regulatory elements beyond canonical splicing sites, such as exonic/intronic splicing enhancers/silencers. To identify the regulatory elements underlying these *cis*-divergent AS that we observed, we focused on those 243 *cis*-divergent events without sequence variants at the splicing sites. On average, about 12 variants were found within the exon/intron regions flanking each of these events. To determine the exact functional variant(s), we integrated published RNA-seq datasets from brain tissue of five mouse strains (C57BL/6NJ, CAST/EiJ, PWK/PhJ, WSB/EiJ and SPRET/EiJ) (Danecek et al., 2012). 5 events showed

consistent splicing patterns between brain tissues and fibroblast cell line for both C57BL/6J and SPRET/EiJ strains (|ΔPSI| <= 0.1, Appendix Table 2). By correlating the sequence variants with splicing patterns across different mouse strains, we could identify a total of 11 candidate variants potentially responsible for these events (see Appendix Table 2 for details). To confirm the relevance of our finding, we chose one divergent SE in Trim26 gene for further analysis. As shown in Figure 3.22, there were in total four sequence variants in the regions flanking the divergent SE, two of which followed the splicing pattern across different mouse strains, including one 9-nucleotide (nt) -insertion and one SNV (Appendix Table 2 and Fig 3.22). To assess which of the two variants contributed to the divergent splicing pattern, we investigated their effects using minigene reporter assays. Four different minigene constructs containing different combinations of these two variants were transfected into Hek293T and 3T3 cells: (1) "reference": containing no variant compared to C57BL/6J genome; (2) "insert only": containing only the SPRET/EiJ insertion variant; (3)"SNV only": containing only the SPRET/EiJ SNV variant; (4) "SNV & insert": containing both the SPRET/EiJ insertion and SNV variants (Fig 3.23A). As shown in Fig 3.23B and Fig 3.24, the splicing differences detected between "reference" and "SNV & insert" constructs were consistent with the splicing divergence observed between C57BL/6J and SPRET/EiJ strains, *i.e.* the PSI values from SPRET/EiJ allele were smaller than that from the C57BL/6J allele. Further comparison of "insert only" and "SNV only" constructs showed that, the insertion variant alone could lead to the enhanced SE observed in SPRET/EiJ allele.

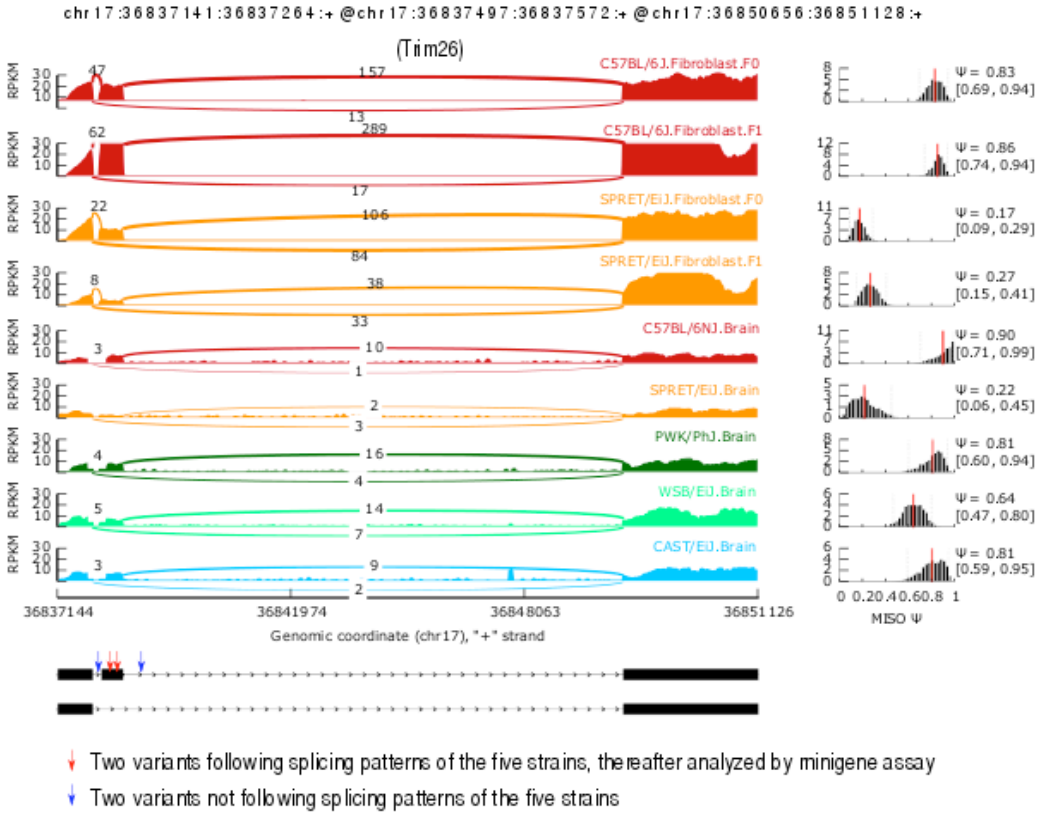


Fig 3.22 Sashimi plot for the splicing patterns of the SE event in Trim26 gene from fibroblast cell line as well as brain tissues of five mouse strains. The top four rows represented splicing patterns for C57BL/6J and SPRET/EiJ strains and their alleles in F1 hybrid. The bottom five rows represented splicing patterns for brains tissues of the five mouse strains. PWK/PhJ and CAST/EiJ had a similar splicing pattern as C57BL/6J, but different from SPRET/EiJ. 4 variants located in the flanking regions, two of which correlated with the species-specific splicing pattern.

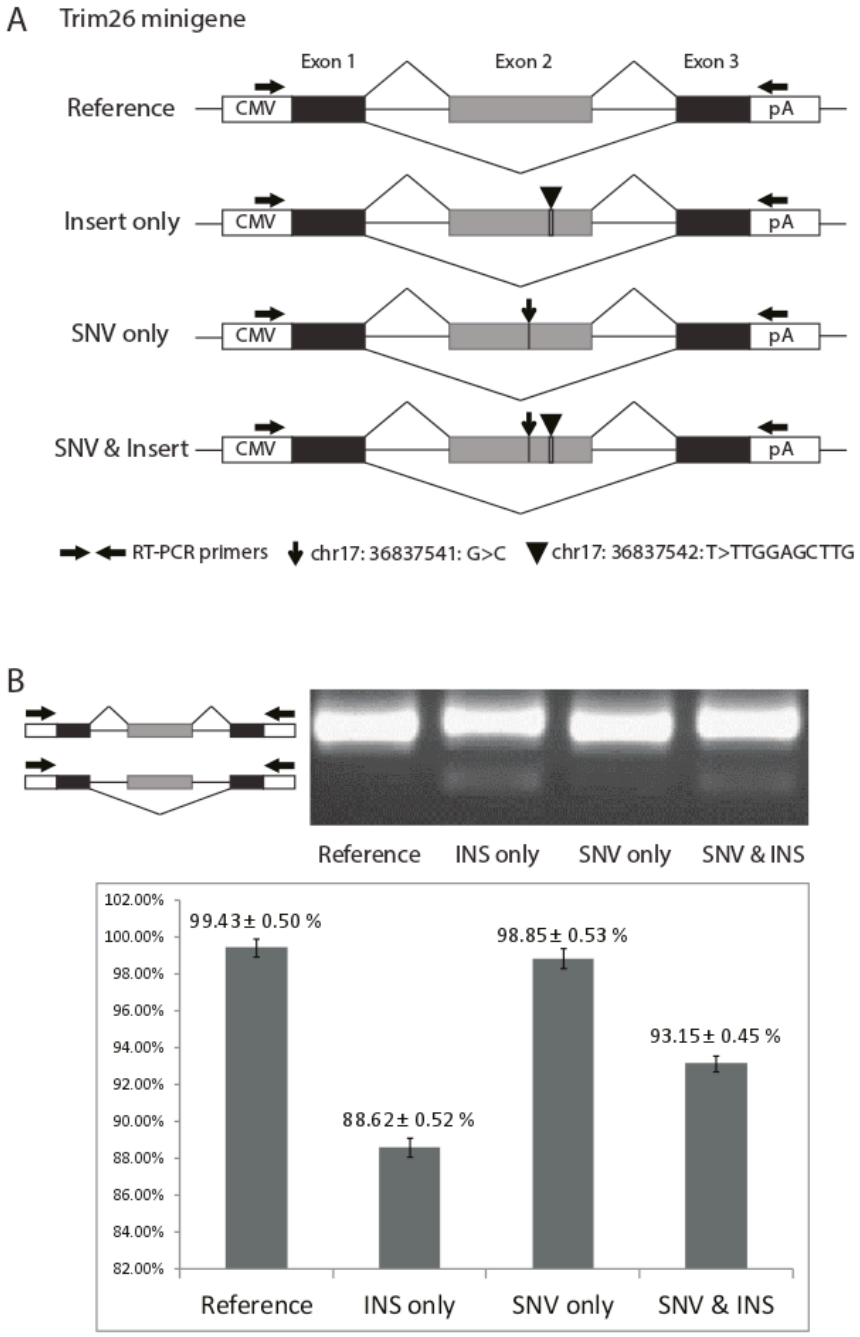


Fig 3.23 Minigene analysis for the *cis*-divergent SE event in Trim26 gene. A. Schematic diagrams of minigene constructs for validating the *cis*-divergent SE event identified in Trim26 gene. Two candidate variants, one SNV and one insertion (INS) were indicated. Four constructs were prepared in C57BL/6J background with no variant, only insertion, only SNV, and both insertion and SNV, respectively (See Materials and methods). B. Minigene assays of the four constructs transfected into HEK293T cells suggested only the insertion contributed to this divergent SE event. The gel image illustrated RT-PCR products from these constructs. The barplot below

the gel image represented the PSI values calculated from triplicates of RT-PCR products using Agilent Bioanalyzer 2000 system.

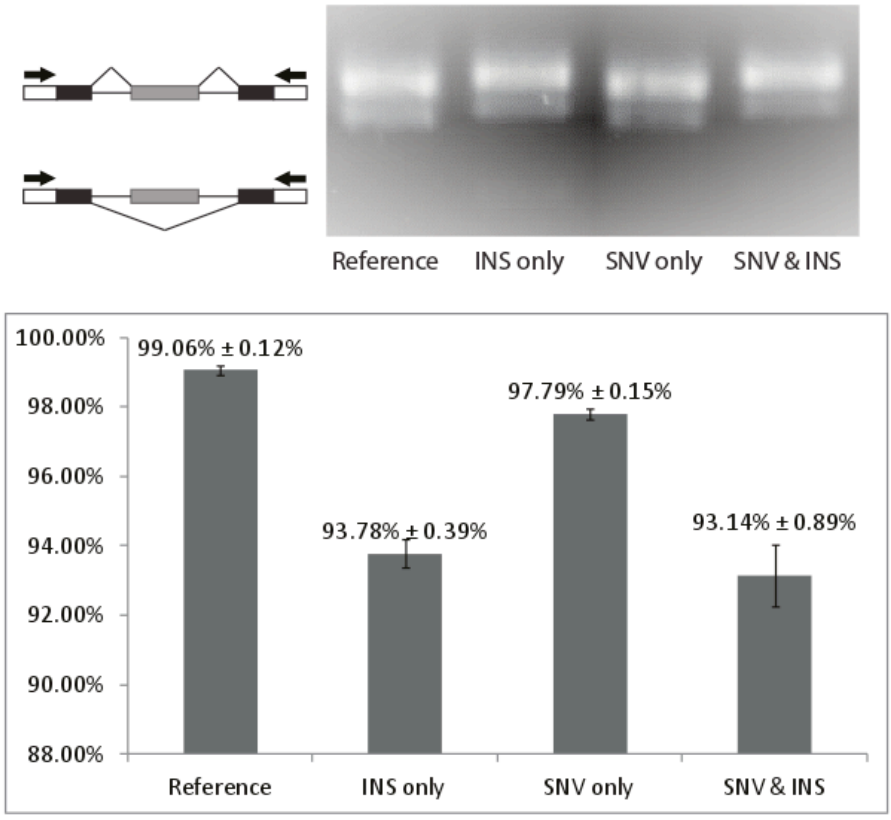


Figure 3.24 Minigene analysis for the *cis*-divergent SE event in Trim26 gene in NIH3T3 cells. Label is the same as in Fig 3.23B.

4. Discussion

Change in AS, one of the major driving forces to shape phenotypic diversity during evolution, could arise from the divergence in *cis*-regulatory elements and/or *trans*-acting RBPs. To globally investigate the relative contribution of the two factors for the first time in a mammalian system, we applied RNA-seq to investigate splicing difference between C57BL/6J and SPRET/EiJ inbred mouse strains and allele-specific splicing pattern in their F1 hybrid. Our results clearly showed the predominant contribution of *cis*-regulatory variants across all the five types of AS.

4.1 QTL versus F1 hybrid

To identify the genetic variants with regulatory effects on gene expression, the most popular method is eQTL mapping, in which different genotypes are correlated with gene expression level in a large population with diverse genetic backgrounds (Lappalainen et al., 2013; Majewski and Pastinen, 2011; Pickrell et al., 2010). Recently this strategy has been extended to measure the genetic regulation on AS (sQTL) (Li et al., 2010). However, genome-wide eQTL/sQTL that tests the association between all SNPs against all expression/AS events is statistically underpowered, in particular for identifying *trans*-factors lying in distal regions. Therefore, the relative *cis*-/*trans*- contribution estimated using QTL methods could be biased towards higher *cis*- effect.

An alternative approach that could more directly address the effect of *cis*-/*trans*-divergence is to compare the allelic difference in F1 hybrid to the difference observed between two parental strains. In F1 hybrid nuclei, *cis*-regulatory alleles from both parental alleles are exposed to the same *trans*-acting factors. Thus the allele-specific splicing differences in hybrids directly reflect differences in pre-mRNA cis-regulatory elements. *Trans*-acting contributions can then be inferred by comparison of allele-specific differences in the hybrid to the splicing differences between the parental strains. This approach has been successfully used for studying *cis*-/*trans*-contribution in gene expression divergence in yeast, fly, mouse and plant (Coolon et al., 2014; Emerson et al., 2010; Goncalves et al., 2012; McManus et al., 2010; Springer and Stupar, 2007; Tirosh et al., 2009b; Wittkopp et al., 2004b, 2008). More recently, McManus *et al* used this strategy to address the *cis*-/*trans*- contribution to

AS evolution in *Drosophila* (McManus et al., 2014). In this study, we applied the same approach in mice to globally and directly address the contributions of *cis*-/*trans*-regulatory divergence.

4.2 SPRET/EiJ versus other mouse strains

Mouse models are important for research into many human diseases. An individual mouse strain differs from all other strains in a wide range of medically important characteristics. To identify all the underlying genetic variations between different strains, the genomes of 17 key mouse strains were decoded using next generation sequencing technologies (the Mouse Genome Project) (Keane et al., 2011). Among all these strains with high quality genome assembly, SPRET/EiJ has the largest number of sequence variants relative to C57BL/6J, including about 35.4 million SNVs and 4.5 million indels. In previous study, C57BL/6J, CAST/EiJ and their F1 hybrids were successfully used to dissect the relative contributions of *cis*-/*trans*-regulatory divergence in the evolution of gene expression (Goncalves et al., 2012). The sequence variants between C57BL/6J and SPRET/EiJ are about twice as many as those between C57BL/6J and CAST/EiJ. This large genomic divergence provides a large number of potential regulatory variants between the two strains (Table 4.1).

Despite ~1.5 million years of evolutionary divergence between C57BL/6J and CAST/EiJ, their interspecific crosses can still produce viable and fertile progeny (Dejager et al., 2009). More importantly, the allele-specific RNA transcripts in their F1 progeny can be distinguished using sequencing-based approach according to their high sequence divergence. In our study, about 60% of mapped 2*100nt reads could be unambiguously assigned to their parental alleles. Moreover, the allelic ΔPSI value correlated well with independent measurement using PacBio full-length sequencing of AS-spanning cDNA PCR products (R^2=0.92).

4.3 Mouse versus *Drosophila*

In cultured fibroblast cells, we identified 796 and 381 differentially regulated splicing events between the two parental strains and between the two alleles in F1 hybrid, respectively. By comparing the two datasets, we could attribute the splicing divergence between the two strains predominately to *cis*-regulatory variants for all

Table 4.1 An overview of the sequence and variants called from 17 mouse genomes.

Strain	Mapped data (GB)	Depth	%genome inaccessible	SNVs	Indels	Structural variants
C57BL/6NJ	77.29	29.29	13.21	9,844	22,228	431
129S1/SvImJ	71.91	27.25	15.3	4,458,004	886,136	29,153
129S5SvEv^{Brd}	50.27	19.05	15.17	4,383,799	810,310	25,340
129P2/Ola	115.52	43.78	14.47	4,694,529	1,028,629	32,227
A/J	70.39	26.68	15.9	4,198,324	823,688	28,691
AKR/J	107.16	40.61	14.86	4,331,384	966,002	30,742
BALB/cJ	65.72	24.9	15.09	3,920,925	831,193	25,702
C3H/HeJ	92.81	35.17	15.09	4403599	949,206	28,532
CBA/J	77.43	29.34	14.79	4,511,278	929,860	28,183
DBA/2J	65.11	24.67	15.09	4,468,071	868,611	28,346
LP/J	73.03	27.67	15.29	4,701,445	947,614	30,024
NOD/ShiLtJ	75.88	28.75	17.3	4,323,530	797,086	30,605
NZO/HILtJ	45.68	17.31	16.06	4,492,372	806,511	25,125
PWK/PhJ	66.99	25.38	19.26	17,202,436	2,635,885	90,125
CAST/EiJ	64.84	24.57	19.18	17,673,726	2,727,089	86,322
WSB/EiJ	48.19	18.26	16.23	6,045,573	1,197,006	35,066
SPRET/EiJ	70.41	26.68	23.26	35,441,735	4,456,243	157,306

This table is adapted from Keane et al. 2011.

five types of AS. Importantly, a similar analysis on the liver tissues from the same parental and F1 strains showed a same trend. To further exclude the possibility that our observation of predominant *cis*-contribution was a peculiarity of the two mouse strains used in this study, we re-analyzed published RNA-seq datasets generated from the liver of C57BL/6J, CAST/EiJ and their F1 hybrid (Goncalves et al., 2012). Although the absolute number of divergent events both between parental strains and between alleles in F1 hybrid that we could identify was much lower, the predominant contribution of *cis*-regulatory difference was still evident, implying the predominant

cis-contribution could be generalized to the evolution of AS in mouse.

Our observation was consistent with previous study of difference in exon skipping between human and mouse, in which 13 divergent SE events were mostly attributed to *cis*-regulatory variants (Barbosa-Morais et al., 2012). In contrast, a more recent study in *Drosophila* found that whereas RI, A3SS and A5SS were still primarily *cis*-directed, *trans*-effects played a dominant role in SE divergence. The authors of latter study attributed the inconsistence between their result and the result from human/mouse study to the different evolutionary distances, *i.e.* ~2.5 million years between different *Drosophila* strains versus ~75 million years between human and mouse (Cutter, 2008; McManus et al., 2014; Waterston et al., 2002). *Cis*-regulatory divergences could preferentially accumulate over evolutionary time, therefore contribute more substantially to the human/mouse comparison (Lemos et al., 2008; Wittkopp et al., 2008). However, in our study, the evolutionary distance between C57BL/6J and SPRET/EiJ strains is ~1.5 million years, similar as that in the *Drosophila* study. Thus, our results of consistent *cis*-dominant contribution excluded different evolutionary distances as a plausible explanation for inconsistent observations between *Drosophila* and mammals. Instead, a more plausible explanation for the discrepancy is genuine differences in mechanisms underlying evolutions of AS regulations between *Drosophila* and mammals. Previous studies have demonstrated the splicing evolutions differ from several perspectives between *Drosophila* and mouse (Khodor et al., 2012; Xiao et al., 2007). For instances, in mammals, the exon has been suggested as the primary evolutionary unit, while the intron was considered as the unit in *Drosophila* (Xiao et al., 2007). Moreover, the cotranscriptional splicing efficiency also differ dramatically between *Drosophila* and mouse (Khodor et al., 2012). Other explanations could also be: 1) the conclusion in the *Drosophila* study might be affected by a much lower number of divergent events identified there (Between *Drosophila melanogaster* and *Drosophila simulans*, 7 and 4 divergent SE were attributed to *cis*- and *trans*- divergence, whereas between *Drosophila melanogaster* and *Drosophila sechellia*, 2 and 3 divergent SE were attributed to *cis*- and *trans*- effects, respectively). 2) the study designs were different (whole animal for *Drosophila* versus distinct cell/tissue for mouse).

4.4 Tissue-specific alternative splicing

High throughput studies have shown that most of alternative splicing isoforms are differentially expressed among tissues, indicating that most AS events are regulated in a tissue-specific manner (Wang et al., 2008). Systematic analysis has identified several features of tissue-specific alternative exons. For example, they are more frame-preserving than constitutive exons (Xing and Lee, 2005), are enriched in predicted post-translational modification sites such as phosphorylation sites (Buljan et al., 2012; Merkin et al., 2012; Zhang et al., 2010), and are able to modulate interactions with their partner proteins thereby remodeling protein-protein interaction network (Ellis et al., 2012).

Tissue-specific AS events can be controlled by tissue-specific expression of splicing factors and the corresponding regulation of their target mRNA transcripts (Castle et al., 2008; David and Manley, 2008). Many tissue-specific splicing regulators have been identified. However, the total number of reported splicing factors is much less than transcription factors (Chen and Manley, 2009). Given the prevalence and importance of AS, there might be many more potential tissue-specific splicing factors to be discovered. Especially, recent large-scale cross-linking studies combined with quantitative mass spectrometry have identified hundreds of new RBPs, some of which could also be involved in the regulation of AS (Baltz et al., 2012; Castello et al., 2012).

In this study, we showed the predominant contribution of *cis*-regulatory variants in two cell types – liver tissue and fibroblast cell line. By comparison of the allelic splicing differences between the two cell types of F1 hybrid, we also found some AS events were divergent between the two alleles in both tissues while others showed a tissue-specific manner. These results indicated that during evolution, some tissue-specific splicing motifs are disrupted or created by the sequencing variants between C57BL/6J and SPRET/EiJ. Thus, our F1 hybrid system can also be used to identify tissue-specific cis-regulatory elements as well as their trans-acting factors by associating the tissue-specific splicing pattern with tissue-specific gene expression pattern of RBPs.

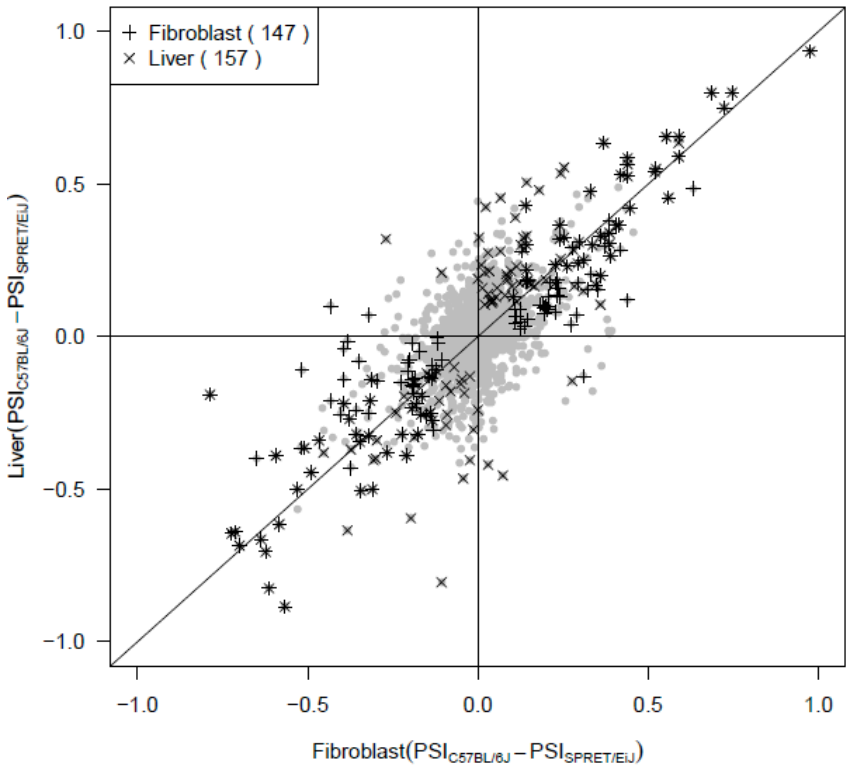


Fig 4.1 Scatterplot comparing allelic splicing difference in liver tissue (y axis) versus fibroblast cell line (x axis). Out of 2,594 AS events (grey dots) expressed in both cell types after filtering using mock F1 hybrid, 157 and 147 AS events were divergent in liver (indicated as "x") and fibroblast (indicated as "+"), respectively, 73 of which showed divergence in both.

4.5 Conclusion and perspective remarks

In summary, our study provided the first direct global investigation for the regulatory divergences of all five AS types in mouse. Our result demonstrated that, in mouse, all five AS types diverged under similar rates, and more importantly, clearly showed the predominant contribution of *cis* variants across divergences of all the five AS types. Comparing with the study from McManus *et al*, this indicates the regulatory differences between the AS evolutions of *Drosophila* and mouse. Furthermore, our F1 hybrid mice also provide a unique system for discovering novel AS regulatory elements. *Cis*-regulatory divergence result solely from sequence variants in pre-mRNA sequences, which could affect directly canonical splicing sites or exonic or intronic regulatory elements. Among the *cis*- divergent events identified in this

study, 36.2% contained sequence variants at the canonical splice sites, a proportion of which could substantially affect the strength of splicing sites. The remaining events without sequencing variants at splicing sites could be used to identify potential exonic/intronic regulatory elements, as demonstrated in this study. Using the same F1 hybrid mice, future datasets on the allelic splicing obtained from different tissues could be used to discover more novel regulatory elements.

5. References

Alekseyenko, A.V., Kim, N., and Lee, C.J. (2007). Global analysis of exon creation versus loss and the role of alternative splicing in 17 vertebrate genomes. RNA *13*, 661-670.

Altman, D.G., and Bland, J.M. (2003). Interaction revisited: the difference between two estimates. BMJ *326*, 219.

Ast, G. (2004). How did alternative splicing evolve? Nature Reviews Genetics *5*, 773-782.

Baltz, A.G., Munschauer, M., Schwanhausser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M.*, et al.* (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol Cell *46*, 674-690.

Barbosa-Morais, N.L., Irimia, M., Pan, Q., Xiong, H.Y., Gueroussov, S., Lee, L.J., Slobodeniuc, V., Kutter, C., Watt, S., Colak, R.*, et al.* (2012). The evolutionary landscape of alternative splicing in vertebrate species. Science *338*, 1587-1593.

Battle, A., Mostafavi, S., Zhu, X., Potash, J.B., Weissman, M.M., McCormick, C., Haudenschild, C.D., Beckman, K.B., Shi, J., Mei, R.*, et al.* (2014). Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. Genome Res *24*, 14-24.

Blekhman, R., Marioni, J.C., Zumbo, P., Stephens, M., and Gilad, Y. (2010). Sex-specific and lineage-specific alternative splicing in primates. Genome Res *20*, 180-189.

Blencowe, B.J. (2006). Alternative splicing: new insights from global analyses. Cell *126*, 37-47.

Braunschweig, U., Barbosa-Morais, N.L., Pan, Q., Nachman, E.N., Alipanahi, B., Gonatopoulos-Pournatzis, T., Frey, B., Irimia, M., and Blencowe, B.J. (2014). Widespread intron retention in mammals functionally tunes transcriptomes. Genome Res *24*, 1774-1786.

Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csardi, G., Harrigan, P., Weier,

M., Liechti, A., Aximu-Petri, A., Kircher, M., *et al.* (2011). The evolution of gene expression levels in mammalian organs. Nature *478*, 343-348.

Buljan, M., Chalancon, G., Eustermann, S., Wagner, G.P., Fuxreiter, M., Bateman, A., and Babu, M.M. (2012). Tissue-Specific Splicing of Disordered Segments that Embed Binding Motifs Rewires Protein Interaction Networks. Mol Cell *46*, 871-883.

Calarco, J.A., Xing, Y., Caceres, M., Calarco, J.P., Xiao, X., Pan, Q., Lee, C., Preuss, T.M., and Blencowe, B.J. (2007). Global analysis of alternative splicing differences between humans and chimpanzees. Gene Dev *21*, 2963-2975.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., *et al.* (2012). Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. Cell *149*, 1393-1406.

Castle, J.C., Zhang, C., Shah, J.K., Kulkarni, A.V., Kalsotra, A., Cooper, T.A., and Johnson, J.M. (2008). Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines. Nat Genet *40*, 1416-1425.

Chen, L., Tovar-Corona, J.M., and Urrutia, A.O. (2012). Alternative splicing: a potential source of functional innovation in the eukaryotic genome. Int J Evol Biol *2012*, 596274.

Chen, M., and Manley, J.L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat Rev Mol Cell Biol *10*, 741-754.

Cieply, B., and Carstens, R.P. (2015). Functional roles of alternative splicing factors in human disease. Wiley Interdiscip Rev RNA *6*, 311-326.

Clark, F., and Thanaraj, T.A. (2002). Categorization and characterization of transcript-confirmed constitutively and alternatively spliced introns and exons from human. Hum Mol Genet *11*, 451-464.

Coolon, J.D., McManus, C.J., Stevenson, K.R., Graveley, B.R., and Wittkopp, P.J. (2014). Tempo and mode of regulatory evolution in Drosophila. Genome Res *24*, 797-808.

Cutter, A.D. (2008). Divergence times in Caenorhabditis and Drosophila inferred

from direct estimates of the neutral mutation rate. Mol Biol Evol *25*, 778-786.

Danecek, P., Nellaker, C., McIntyre, R.E., Buendia-Buendia, J.E., Bumpstead, S., Ponting, C.P., Flint, J., Durbin, R., Keane, T.M., and Adams, D.J. (2012). High levels of RNA-editing site conservation amongst 15 laboratory mouse strains. Genome Biol *13*, 26.

David, C.J., and Manley, J.L. (2008). The search for alternative splicing regulators: new approaches offer a path to a splicing code. Genes Dev *22*, 279-285.

Dejager, L., Libert, C., and Montagutelli, X. (2009). Thirty years of Mus spretus: a promising future. Trends Genet *25*, 234-241.

Dodt, M., Roehr, J.T., Ahmed, R., and Dieterich, C. (2012). FLEXBAR-Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms. Biology (Basel) *1*, 895-905.

Dori-Bachash, M., Shema, E., and Tirosh, I. (2011). Coupled evolution of transcription and mRNA degradation. PLoS Biol *9*, e1001106.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B.*, et al.* (2009). Real-time DNA sequencing from single polymerase molecules. Science *323*, 133-138.

Ellis, J.D., Barrios-Rodiles, M., Colak, R., Irimia, M., Kim, T., Calarco, J.A., Wang, X., Pan, Q., O'Hanlon, D., Kim, P.M.*, et al.* (2012). Tissue-specific alternative splicing remodels protein-protein interaction networks. Mol Cell *46*, 884-892.

Emerson, J.J., Hsieh, L.C., Sung, H.M., Wang, T.Y., Huang, C.J., Lu, H.H., Lu, M.Y., Wu, S.H., and Li, W.H. (2010). Natural selection on cis and trans regulation in yeasts. Genome Res *20*, 826-836.

Fu, X.D., and Ares, M., Jr. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. Nat Rev Genet *15*, 689-701.

Gao, Q., Sun, W., You, X., Froehler, S., and Chen, W. (2013). A systematic evaluation of hybridization-based mouse exome capture system. BMC Genomics *14*, 492.

Gerstein, M.B., Lu, Z.J., Van Nostrand, E.L., Cheng, C., Arshinoff, B.I., Liu, T., Yip, K.Y., Robilotto, R., Rechtsteiner, A., Ikegami, K.*, et al.* (2010). Integrative analysis

of the Caenorhabditis elegans genome by the modENCODE project. Science *330*, 1775-1787.

Gibson, G., and Weir, B. (2005). The quantitative genetics of transcription. Trends Genet *21*, 616-623.

Goncalves, A., Leigh-Brown, S., Thybert, D., Stefflova, K., Turro, E., Flicek, P., Brazma, A., Odom, D.T., and Marioni, J.C. (2012). Extensive compensatory cis-trans regulation in the evolution of mouse gene expression. Genome Res *22*, 2376-2384.

Graveley, B.R. (2009). Alternative splicing: regulation without regulators. Nat Struct Mol Biol *16*, 13-15.

Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, L., Artieri, C.G., van Baren, M.J., Boley, N., Booth, B.W.*, et al.* (2011). The developmental transcriptome of Drosophila melanogaster. Nature *471*, 473-479.

Hartmann, B., Castelo, R., Minana, B., Peden, E., Blanchette, M., Rio, D.C., Singh, R., and Valcarcel, J. (2011). Distinct regulatory programs establish widespread sex-specific alternative splicing in Drosophila melanogaster. RNA *17*, 453-468.

Jangi, M., and Sharp, P.A. (2014). Building robust transcriptomes with master splicing factors. Cell *159*, 487-498.

Jelen, N., Ule, J., Zivin, M., and Darnell, R.B. (2007). Evolution of Nova-dependent splicing regulation in the brain. PLoS Genet *3*, 1838-1847.

Kafasla, P., Mickleburgh, I., Llorian, M., Coelho, M., Gooding, C., Cherny, D., Joshi, A., Kotik-Kogan, O., Curry, S., Eperon, I.C.*, et al.* (2012). Defining the roles and interactions of PTB. Biochem Soc Trans *40*, 815-820.

Kalsotra, A., and Cooper, T.A. (2011). Functional consequences of developmentally regulated alternative splicing. Nat Rev Genet *12*, 715-729.

Katz, Y., Wang, E.T., Airoldi, E.M., and Burge, C.B. (2010). Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nat Methods *7*, 1009-1015.

Keane, T.M., Goodstadt, L., Danecek, P., White, M.A., Wong, K., Yalcin, B., Heger, A., Agam, A., Slater, G., Goodson, M.*, et al.* (2011). Mouse genomic variation and its effect on phenotypes and gene regulation. Nature *477*, 289-294.

Keren, H., Lev-Maor, G., and Ast, G. (2010). Alternative splicing and evolution: diversification, exon definition and function. Nat Rev Genet *11*, 345-355.

Khan, Z., Bloom, J.S., Amini, S., Singh, M., Perlman, D.H., Caudy, A.A., and Kruglyak, L. (2012). Quantitative measurement of allele-specific protein expression in a diploid yeast hybrid by LC-MS. Mol Syst Biol *8*, 602.

Khodor, Y.L., Menet, J.S., Tolan, M., and Rosbash, M. (2012). Cotranscriptional splicing efficiency differs dramatically between Drosophila and mouse. RNA *18*, 2174-2186.

Kim, E., Goren, A., and Ast, G. (2008). Alternative splicing: current perspectives. Bioessays *30*, 38-47.

Koren, E., Lev-Maor, G., and Ast, G. (2007). The emergence of alternative 3' and 5' splice site exons from constitutive exons. PLoS Comput Biol *3*, e95.

Kornblihtt, A.R., Schor, I.E., Allo, M., Dujardin, G., Petrillo, E., and Munoz, M.J. (2013). Alternative splicing: a pivotal step between eukaryotic transcription and translation. Nat Rev Mol Cell Biol *14*, 153-165.

Lappalainen, T., Sammeth, M., Friedlander, M.R., t Hoen, P.A., Monlong, J., Rivas, M.A., Gonzalez-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G.*, et al.* (2013). Transcriptome and genome sequencing uncovers functional variation in humans. Nature *501*, 506-511.

Lee, J.A., Tang, Z.Z., and Black, D.L. (2009). An inducible change in Fox-1/A2BP1 splicing modulates the alternative splicing of downstream neuronal target exons. Genes Dev *23*, 2284-2293.

Lemos, B., Araripe, L.O., Fontanillas, P., and Hartl, D.L. (2008). Dominance and the evolutionary accumulation of cis- and trans-effects on gene expression. Proc Natl Acad Sci U S A *105*, 14471-14476.

Li, Y., Breitling, R., Snoek, L.B., van der Velde, K.J., Swertz, M.A., Riksen, J., Jansen, R.C., and Kammenga, J.E. (2010). Global genetic robustness of the alternative splicing machinery in Caenorhabditis elegans. Genetics *186*, 405-410.

Lin, L., Shen, S., Jiang, P., Sato, S., Davidson, B.L., and Xing, Y. (2010). Evolution of alternative splicing in primate brain transcriptomes. Hum Mol Genet *19*,

2958-2973.

Lu, Z.X., Jiang, P., and Xing, Y. (2012). Genetic variation of pre-mRNA alternative splicing in human populations. Wiley Interdiscip Rev RNA *3*, 581-592.

Lynch, V.J., and Wagner, G.P. (2008). Resurrecting the role of transcription factor change in developmental evolution. Evolution *62*, 2131-2154.

Majewski, J., and Pastinen, T. (2011). The study of eQTL variations by RNA-seq: from SNPs to phenotypes. Trends Genet *27*, 72-79.

Matlin, A.J., Clark, F., and Smith, C.W. (2005). Understanding alternative splicing: towards a cellular code. Nat Rev Mol Cell Biol *6*, 386-398.

McManus, C.J., Coolon, J.D., Duff, M.O., Eipper-Mains, J., Graveley, B.R., and Wittkopp, P.J. (2010). Regulatory divergence in Drosophila revealed by mRNA-seq. Genome Res *20*, 816-825.

McManus, C.J., Coolon, J.D., Eipper-Mains, J., Wittkopp, P.J., and Graveley, B.R. (2014). Evolution of splicing regulatory networks in Drosophila. Genome Res *24*, 786-796.

Meireles-Filho, A.C., and Stark, A. (2009). Comparative genomics of gene regulation-conservation and divergence of cis-regulatory information. Curr Opin Genet Dev *19*, 565-570.

Merkin, J., Russell, C., Chen, P., and Burge, C.B. (2012). Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. Science *338*, 1593-1599.

Moore, M.J., and Proudfoot, N.J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. Cell *136*, 688-700.

Necsulea, A., and Kaessmann, H. (2014). Evolutionary dynamics of coding and non-coding transcriptomes. Nat Rev Genet *15*, 734-748.

Nilsen, T.W., and Graveley, B.R. (2010). Expansion of the eukaryotic proteome by alternative splicing. Nature *463*, 457-463.

Ongen, H., and Dermitzakis, E.T. (2015). Alternative Splicing QTLs in European and African Populations. Am J Hum Genet *97*, 567-575.

Pan, Q., Bakowski, M.A., Morris, Q., Zhang, W., Frey, B.J., Hughes, T.R., and Blencowe, B.J. (2005). Alternative splicing of conserved exons is frequently

species-specific in human and mouse. Trends in Genetics *21*, 73-77.

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.

Pickrell, J.K., Marioni, J.C., Pai, A.A., Degner, J.F., Engelhardt, B.E., Nkadori, E., Veyrieras, J.B., Stephens, M., Gilad, Y., and Pritchard, J.K. (2010). Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature *464*, 768-772.

Pillmann, H., Hatje, K., Odronitz, F., Hammesfahr, B., and Kollmar, M. (2011). Predicting mutually exclusive spliced exons based on exon length, splice site and reading frame conservation, and exon sequence homology. BMC Bioinformatics *12*, 270.

Pohl, M., Bortfeldt, R.H., Grutzmann, K., and Schuster, S. (2013). Alternative splicing of mutually exclusive exons--a review. Biosystems *114*, 31-38.

Proudfoot, N.J. (2011). Ending the message: poly(A) signals then and now. Genes Dev *25*, 1770-1782.

Ramani, A.K., Calarco, J.A., Pan, Q., Mavandadi, S., Wang, Y., Nelson, A.C., Lee, L.J., Morris, Q., Blencowe, B.J., Zhen, M.*, et al.* (2011). Genome-wide analysis of alternative splicing in Caenorhabditis elegans. Genome Res *21*, 342-348.

Rozowsky, J., Abyzov, A., Wang, J., Alves, P., Raha, D., Harmanci, A., Leng, J., Bjornson, R., Kong, Y., Kitabayashi, N.*, et al.* (2011). AlleleSeq: analysis of allele-specific expression and binding in a network framework. Mol Syst Biol *7*, 522.

Sakabe, N.J., and de Souza, S.J. (2007). Sequence features responsible for intron retention in human. BMC Genomics *8*, 59.

Sorek, R., and Ast, G. (2003). Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. Genome Res *13*, 1631-1637.

Springer, N.M., and Stupar, R.M. (2007). Allele-specific expression patterns reveal biases and embryo-specific parent-of-origin effects in hybrid maize. Plant Cell *19*, 2391-2402.

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.

Sterne-Weiler, T., Martinez-Nunez, R.T., Howard, J.M., Cvitovik, I., Katzman, S., Tariq, M.A., Pourmand, N., and Sanford, J.R. (2013). Frac-seq reveals isoform-specific recruitment to polyribosomes. Genome Res *23*, 1615-1623.

Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. Proc Natl Acad Sci U S A *100*, 9440-9445.

Sun, W., You, X., Gogol-Doring, A., He, H., Kise, Y., Sohn, M., Chen, T., Klebes, A., Schmucker, D., and Chen, W. (2013). Ultra-deep profiling of alternatively spliced Drosophila Dscam isoforms by circularization-assisted multi-segment sequencing. EMBO J *32*, 2029-2038.

Tirosh, I., Reikhav, S., Levy, A.A., and Barkai, N. (2009a). A Yeast Hybrid Provides Insight into the Evolution of Gene Expression Regulation. Science *324*, 659-662.

Tirosh, I., Reikhav, S., Levy, A.A., and Barkai, N. (2009b). A yeast hybrid provides insight into the evolution of gene expression regulation. Science *324*, 659-662.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

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.

Wang, Z., and Burge, C.B. (2008). Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. RNA *14*, 802-813.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet *10*, 57-63.

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P.*, et al.* (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520-562.

Wittkopp, P.J., Haerum, B.K., and Clark, A.G. (2004a). Evolutionary changes in cis and trans gene regulation. Nature *430*, 85-88.

Wittkopp, P.J., Haerum, B.K., and Clark, A.G. (2004b). Evolutionary changes in cis and trans gene regulation. Nature *430*, 85-88.

Wittkopp, P.J., Haerum, B.K., and Clark, A.G. (2008). Regulatory changes underlying expression differences within and between Drosophila species. Nat Genet *40*, 346-350.

Wittkopp, P.J., and Kalay, G. (2012). Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat Rev Genet *13*, 59-69.

Xiao, X., Wang, Z., Jang, M., and Burge, C.B. (2007). Coevolutionary networks of splicing cis-regulatory elements. Proc Natl Acad Sci U S A *104*, 18583-18588.

Xing, Y., and Lee, C.J. (2005). Protein modularity of alternatively spliced exons is associated with tissue-specific regulation of alternative splicing. Plos Genetics *1*, 323-328.

Yap, K., and Makeyev, E.V. (2013). Regulation of gene expression in mammalian nervous system through alternative pre-mRNA splicing coupled with RNA quality control mechanisms. Mol Cell Neurosci *56*, 420-428.

Yu, Y., Maroney, P.A., Denker, J.A., Zhang, X.H., Dybkov, O., Luhrmann, R., Jankowsky, E., Chasin, L.A., and Nilsen, T.W. (2008). Dynamic regulation of alternative splicing by silencers that modulate 5' splice site competition. Cell *135*, 1224-1236.

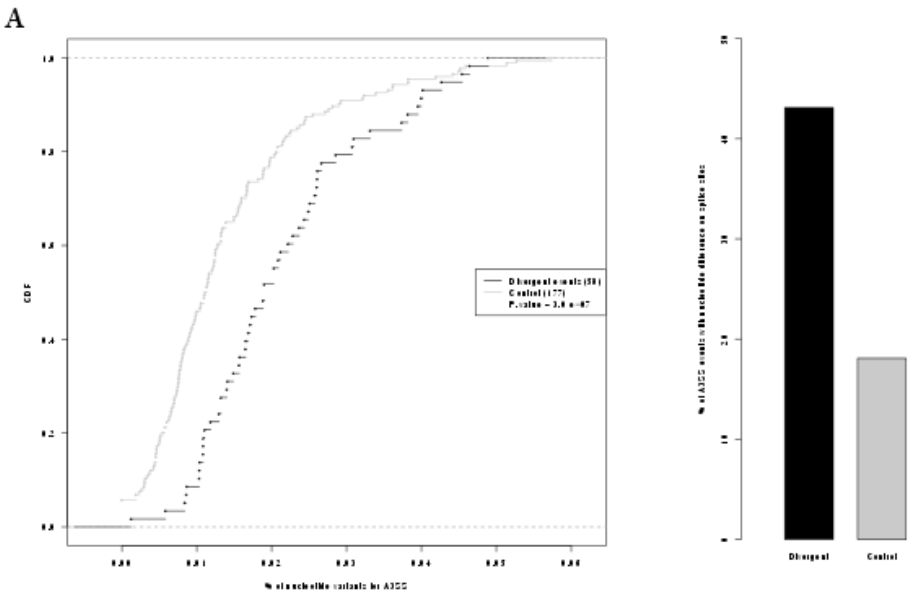
Zhang, C., Frias, M.A., Mele, A., Ruggiu, M., Eom, T., Marney, C.B., Wang, H., Licatalosi, D.D., Fak, J.J., and Darnell, R.B. (2010). Integrative modeling defines the Nova splicing-regulatory network and its combinatorial controls. Science *329*, 439-443.

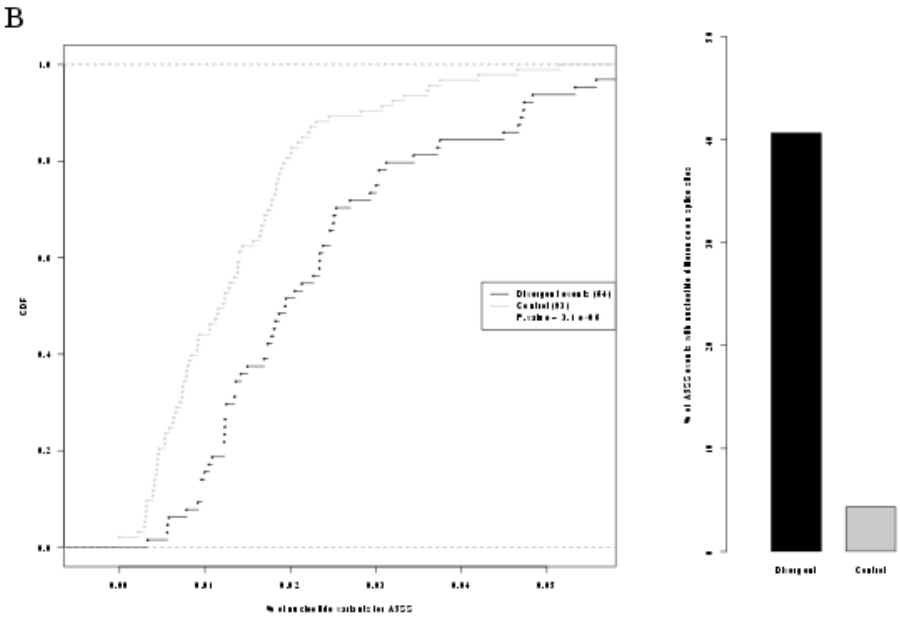
Zhang, X., Joehanes, R., Chen, B.H., Huan, T., Ying, S., Munson, P.J., Johnson, A.D., Levy, D., and O'Donnell, C.J. (2015). Identification of common genetic variants controlling transcript isoform variation in human whole blood. Nat Genet *47*, 345-352.

Zhao, K., Lu, Z.X., Park, J.W., Zhou, Q., and Xing, Y. (2013). GLiMMPS: robust statistical model for regulatory variation of alternative splicing using RNA-seq data. Genome Biol *14*, R74.

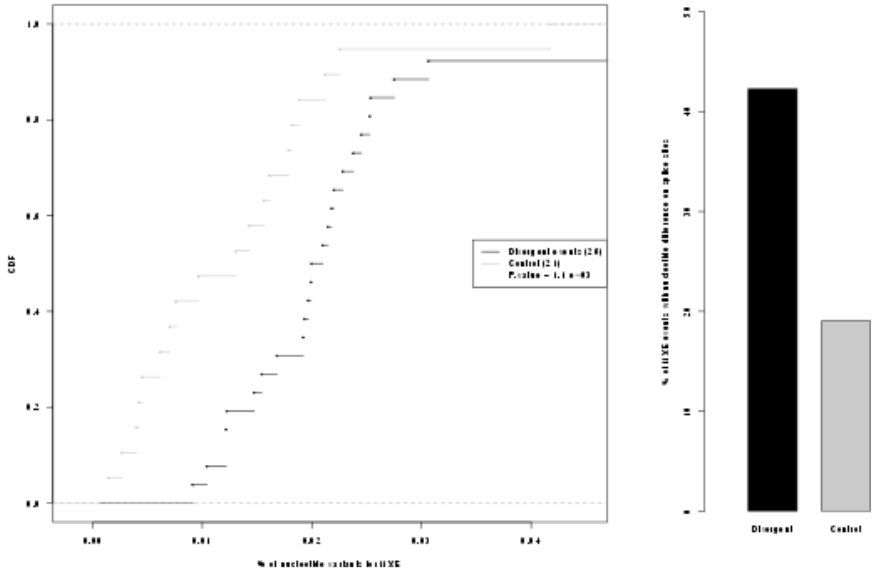
6. Appendix

Appendix figures

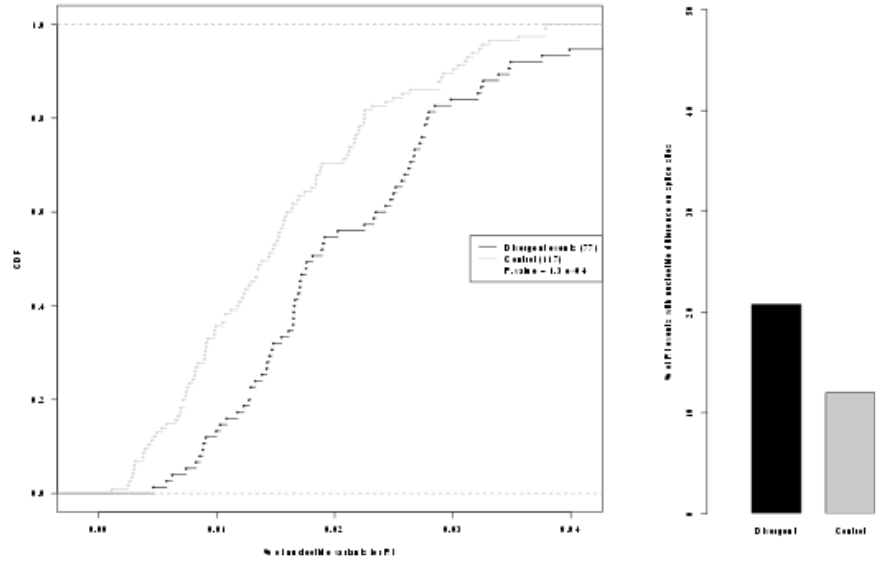




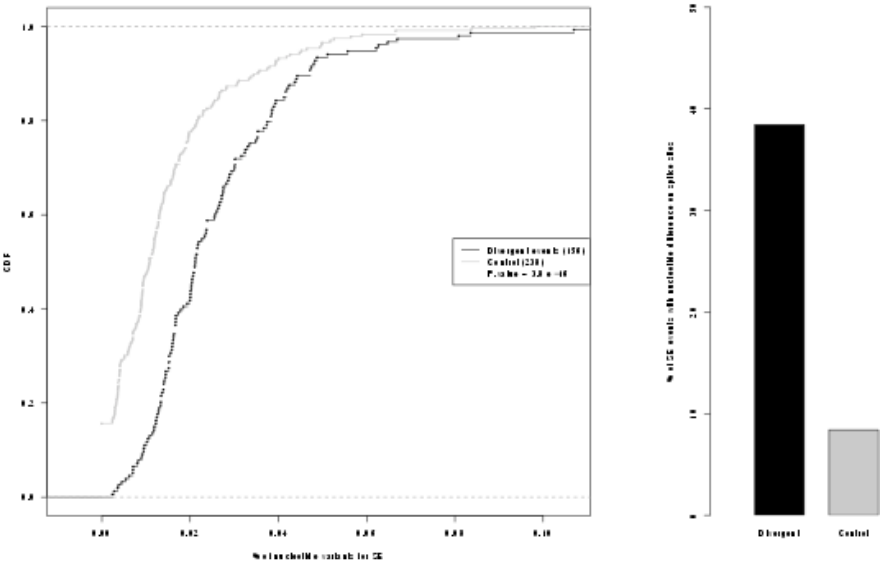
C


CDF
Diverged events (240)
Control (211)
P.value = 1.1 e-03
Diverged
Control

D


CDF
Diverged events (77)
Control (117)
Diverged
Control

E


CDF
Diverged events (150)
Control (238)
Diverged
Control

Appendix figure 1. Genomic features that correlate with *cis*-regulatory alternative splicing divergence for each AS type separately. CDF of frequencies of nucleotide variants in the AS flanking regions for the events with *cis*-regulatory divergence (black) and controls (gray) (left); and percentages of the events with significant *cis*-regulatory divergence (black) and controls events (grey) that had sequence divergence at the exact splice sites (right) for A3SS (A), A5SS (B), MXE (C), RI (D) and SE (E), respectively.

Appendix tables

Appendix table 1. List of RT-PCR primers and summary of PacBio sequencing results for the 20 selected AS events

Event ID Event type Contribution ForwardPrimer ReversePrimer #C57BL/6J_Long #C57BL/6J_Short #SPRET/EiJ_Long #SPRET/EiJ_Short #C57BL/6J allele_Long #C57BL/6J allele_Short #SPRET/EiJ allele_Long #SPRET/EiJ allele_Short

chr2:165287784:165287838:-@chr2:165283535|165283600:165283249:- A3SS Cis AGAGCTGCCTTCCACACAAA GCCCACCTCGAGCTTCC 975 1694 2485 7 327 699 493 5

chr17:23673608:23674068:+@chr17:23674535|23674547:23674675:+ A3SS Cis AGAACCAGCCACCACTGGGGTAACTCATGTAGGCTGAGTCGGGAGA 11207 9 13 10119 3924 111 748 2866

chr7:44467980:44468159|44468190:+@chr7:44468807:44468969:+ A5SS Cis GGGAGCTGTAGTCCCTTTCG TTGTTGAGAGCATGGACAGC 1705 633 2518 0 1227 433 1176 3

chr8:84689247:84689349|84689376:+@chr8:84689632:84689908:+ A5SS Cis AGAACCACTAATGCGAGAAGCATGCAC CGGCTTAGCCACAGGATCG 540 1650 9043 614 122 523 2170 105

chr11:78536260:78536073|78536086:-@chr11:78530028:78530347:- A5SS Cis ACGGCAGGAGGTAGGCTTAGAGCCGCAGGCCAGAGGTG 2472 4545 5 6487 1405 2630 7 1638

chr19:10203943:10203688|10203804:-@chr19:10199132:10201097:- A5SS Cis GCCAGCTCGCCTGACTTC GCTAGCTGCTTAAGGCTCGT 451 312 1570 0 261 212 287 1

chr11:3539112:3539292:-@chr11:3537671:3537743:-@chr11:3537555:3537605:-@chr11:3537323:3537436:- MXE Cis TAAAGCCTCGTCTGCCATCT CTGCCGGTCTCTACTGCATC 95 602 2258 76 40 367 935 179

chr12:103436563:103436682:+@chr12:103437390:103437524:+@chr12:103437676:103437705:+@chr12:103439401:103439562:+ MXE Cis

AGAACCAGCTCATCCTGTTAACACAG ACATCATCTTAGCTGCTAGA 648 965 11 3624 339 590 9 3248

chr10:80393280-80393095:-@chr10:80393018-80392539:-RI Cis

AGAACCAGGCTGCCTCCTACAGGTCT AGGCACCACTGGACAAGG 623 4152 2116 2446 349 2200 371 800

chr3:5563165:5563272:-@chr3:5562669:5562732:-@chr3:5560188:5561760:- SE Cis

CCAGGGCCTTATTCAGTTCA CAACAGTTGATGTCCACAGGA 2096 2976 197 9599 1214 1814 114 4100

chr4:62408552:62408623:-@chr4:62404456:62404552:-@chr4:62402724:62402915:- SE Cis

GTGAAATGGCAGGAGAGGAA GTTGAGGTTCTGGGGAACTG 7915 12 2573 806 3741 8 562 172

chr5:32746183:32746411:-@chr5:32745469:32745682:-@chr5:32742210:32742475:- SE Cis

GCTGGTGGCAATGCTTCTA CGGGAGTGACTGAGCTGATA 545 2850 33 3798 309 1627 18 1716

chr6:83481347:83481462:-@chr6:83480843:83480942:-@chr6:83480214:83480436:- SE Cis

AGAACCAGGCAGTACAAAGACAGCACA GGAGAGACTCTACCAGAGGGATA 393 257 659 7 352 171 523 8

chr7:98359358:98359467:-@chr7:98353477:98353616:-@chr7:98350668:98353130:- SE Cis

ACACGGATCAGGCTGAAACT CAGGTCCCATAGGAGCAGAC 987 2134 0 2146 643 1212 0 2158

chr17:36837141:36837264:+@chr17:36837497:36837572:+@chr17:36850656:36851128:+ SE Cis GCTCCAGGGAGAGGAGTGAC CTTCCTCCAGGCTCCTCAA 4355 86 153 1439 2961 32 46 169

chr18:60828365:60828559:-@chr18:60827340:60827447:-@chr18:60822804:60822990:- SE Cis AGAACCATGAGGCCTTCTTTGTGGTGT CAGGTCAGAGCTGCCTCAG

1178 2080 2952 205 554 791 998 55

chr12:80171729:80171875:-@chr12:80170686:80170766:-@chr12:80170183:80170248:-@chr12:80168858:80169016:- MXE Trans GGTTGGGGTCTACAATGCTC CATCAGCCAGGAACAGATGA 4802 2505 9388 120 3519 490 1689 319

chr7:30312237-30312388:+@chr7:30312591-30312707:+ RI Trans TCTCCCAGCTAACCATGTCC CTGATGGTGCTGACAGTTGG 704 228 1878 191 368 122 317 73

chr9:106473833-106473681:-@chr9:106473519-106473433:- RI Trans AGAACCACCTTCTCGGTTCACTTTCCA AAGTGGAAGAGGTGCCACAC 206 10643 670 2599 319 6421 318 1646

chr6:84063321:84063435:+@chr6:84064471:84064563:+@chr6:84064876:84065087:+ SE Trans AGAACCACCACCACCTGCTTCTCTAGC CTTTTTGCCCCTGGATAGTG 2758 520 937 2201 715 81 637 343

Appendix table 2. Splicing patterns of 143 *cis*-divergent events in brain tissues of the five mouse strains.

Event ID C57BL/6J (Fibroblast) SPRET/EiJ (Fibroblast) C57BL/6NJ (Brain) PWK/PhJ (Brain) WSB/EiJ (Brain) CAST/EiJ (Brain) SPRET/EiJ(Brain)
abs(ΔPSI(Fibroblast-Brain))<=0.1?C57BL/6NJ
abs(ΔPSI(Fibroblast-Brain))<=0.1?SPRET/EiJ
abs(ΔPSI(C57BL/6NJ-SPRET/EiJ))>0.1?Fibroblast
abs(ΔPSI(C57BL/6NJ-SPRET/EiJ))>0.1?Liver

chr6:42709959:42710071:-@chr6:42709529:42709637:-@chr6:42686325:42686958:- 0.13 0 0.19 0.06 0.06 0.11 0.01 TRUE TRUE TRUE TRUE

chr7:80893341-80892976:-@chr7:80891909-80890724:- 0.25 0.02 0.26 0.09 0.28 0.08 0.04 TRUE TRUE TRUE TRUE

chr13:58160152:58160238:+@chr13:58162838:58162881:+@chr13:58163433:58164693:+ 0.07 0.39 0.1 0.06 0.26 0.17 0.34 TRUE TRUE TRUE TRUE

chr17:34837019:34837216|34837241:+@chr17:34837393:34837754:+ 0.41 0.61 0.5 0.52 0.66 0.81 0.65 TRUE TRUE TRUE TRUE

chr17:36837141:36837264:+@chr17:36837497:36837572:+@chr17:36850656:36851128:+ 0.87 0.13 0.9 0.81 0.64 0.81 0.22 TRUE TRUE TRUE TRUE

chr19:6338826:6338961:+@chr19:6339339:6339503:+@chr19:6339687:6340894:+ 0.83 0.74 0.83 0.87 0.8 0.97 0.83 TRUE TRUE FALSE FALSE

chr1:161036820-161037164:+@chr1:161037288-161037315:+ 0.44 0.89 0 0.89 0.7 0.83 0 FALSE FALSE TRUE FALSE

chr1:75195801-75195948:+@chr1:75196153-75196381:+ 0.28 0.14 0 0.64 0.52 0 0.35 FALSE FALSE TRUE TRUE

chr2:12924041:12924198:+@chr2:12974369:12974453:+@chr2:12978160:12978617:+ 0.6 0.92 0.32 0 0.82 0 0 FALSE FALSE TRUE TRUE

chr2:155956558:155956657|155956662:+@chr2:155961872:155962154:+0.49 0.68 0 0.63 0.72 0.73 0.51 FALSE FALSE TRUE TRUE

chr2:32363198:32363298|32363656:+@chr2:32363805:32363979:+ 0.88 0.26 0.64 0.68 0.55 0.82 0 FALSE FALSE TRUE TRUE

chr2:32363805:32363979:+@chr2:32365803:32365874:+@chr2:32367303:32367532:+ 0.66 0.08 0.85 0.88 0.76 0.88 0.79 FALSE FALSE TRUE FALSE

chr4:155854566-155854800:+@chr4:155854886-155855025:+ 0.35 0.11 0.64 0.66 0.5 0.65 0 FALSE FALSE TRUE TRUE

chr5:32611171:32611375|32611837:+@chr5:32640331:32640603:+ 0.22 0.08 0.41 0.36 0.31 0.33 0.35 FALSE FALSE TRUE FALSE

chr7:45134026-45133882:-@chr7:45132412-45132228:- 0.11 0.35 0 0.66 0.7 0 0 FALSE FALSE TRUE FALSE

chr7:45549804:45549965:+@chr7:45553245:45553294:+@chr7:45553569:45553681:+@chr7:45553765:45554229:+ 0.72 0.61 0 0 0.83 0 0 FALSE FALSE TRUE FALSE

chr8:84689247:84689376:+@chr8:84689632|84689708:84689908:+ 0.82 0.57 0 0 0.89 0.67 0.35 FALSE FALSE TRUE TRUE

chr9:15311527-15311695:+@chr9:15311909-15312104:+ 0.57 0.92 0 0.83 0.9 0.9 0 FALSE FALSE TRUE FALSE

chr9:58233153:58233247:-@chr9:58230491:58230628:-@chr9:58229828:58230086:- 0.57 0.83 0 0.52 0.25 0.17 0.4 FALSE FALSE TRUE TRUE

chr10:79858752-79858824:+@chr10:79859072-79859271:+ 0.81 0.98 0.6 0.63 0.75 0.76 0.46 FALSE FALSE TRUE TRUE

chr10:79858952-79858981:+@chr10:79859089-79859271:+ 0.76 0.83 0 0.57 0.7 0.72 0.54 FALSE FALSE FALSE TRUE

chr11:121229309:121229039|121229213:-@chr11:121228328:121228470:- 0.68 0.3 0 0.72 0.8 0.71 0.64 FALSE FALSE TRUE TRUE

chr11:121229309:121229059|121229213:-@chr11:121228328:121228470:- 0.53 0.23 0 0.74 0.78 0.76 0.65 FALSE FALSE TRUE TRUE

chr11:58323270:58323365:-@chr11:58322600:58322729:-@chr11:58319324:58319590:- 0.8 0.69 0.61 0.45 0.63 0.65 0.45 FALSE FALSE TRUE TRUE

chr11:77472877-77472725:-@chr11:77472612-77472449:-0.12 0.11 0.31 0.39 0.23 0.1 0.34 FALSE FALSE FALSE FALSE

chr12:103436563:103436682:+@chr12:103437390:103437524:+@chr12:103437676:103437705:+@chr12:103439401:103439562:+ 0.44 0 0.78 0.49 0.18 0.54 0.17 FALSE FALSE TRUE TRUE

chr13:59474712-59474532:-@chr13:59473817-59473688:-0.02 0 0.25 0.41 0.02 0.09 0.12 FALSE FALSE FALSE TRUE

chr14:51885151-51885304:+@chr14:51885388-51885522:+ 0.21 0.12 0 0.79 0.7 0.79 0 FALSE FALSE FALSE FALSE

chr16:78576657:78576191|78576391:-@chr16:78575921:78576054:- 0.35 0.82 0 0 0.75 0 0 FALSE FALSE TRUE FALSE

chr16:78576688:78576191|78576391:-@chr16:78575921:78576054:- 0.37 0.83 0 0 0.77 0 0 FALSE FALSE TRUE FALSE

chr17:24895055-24894821:-@chr17:24894707-24894571:-0.42 0.18 0 0.04 0.17 0.1 0 FALSE FALSE TRUE FALSE

chr18:60828365:60828559:-@chr18:60827340:60827447:-@chr18:60822804:60822990:- 0.54 0.94 0.67 0.4 0.29 0.26 0.68 FALSE FALSE TRUE FALSE

chr2:120032114-120032465:+@chr2:120034478-120034670:+ 0.09 0.54 0.84 0.85 0.85 0.75 0 FALSE FALSE TRUE TRUE

chr2:120032114-120032465:+@chr2:120034574-120034670:+ 0.06 0.36 0.75 0.85 0.79 0.7 0 FALSE FALSE TRUE TRUE

chr2:158360599-158359182:-@chr2:158357263-158357198:- 0.54 0.47 0.82 0.88 0.21 0.57 0

FALSE FALSE FALSE TRUE

chr2:164832873:164833196:+@chr2:164833724:164833850:+@chr2:164834030:164834208:+ 0.86 0.99 0.1 0.3 0.07 0 0 FALSE FALSE TRUE FALSE

chr2:164832873:164833196:+@chr2:164833773:164833850:+@chr2:164834030:164834208:+ 0.89 0.99 0.11 0.33 0.08 0 0 FALSE FALSE FALSE TRUE

chr7:66069424-66069526:+@chr7:66070590-66070665:+ 0.34 0.2 0.5 0.54 0.49 0.49 0.54 FALSE FALSE TRUE FALSE

chr13:114825211-114825265:+@chr13:114826158-114826281:+ 0.14 0.05 0.58 0.49 0.52 0.68 0.39 FALSE FALSE FALSE TRUE

chr17:25876661-25876739:+@chr17:25876950-25877169:+ 0.3 0.06 0.85 0.87 0.83 0.75 0.75 FALSE FALSE TRUE FALSE

chr2:30953752:30954017:+@chr2:30955768|30955778:30955953:+ 0.06 0.19 0.55 0.36 0.24 0.17 0.24 FALSE TRUE TRUE TRUE

chr2:94010752:94010263|94010683:-@chr2:94008424:94008592:- 0.23 0.03 0 0.57 0.64 0.57 0 FALSE TRUE TRUE FALSE

chr5:93270794-93271044:+@chr5:93271201-93271279:+ 0.08 0.01 0.36 0.14 0.03 0.29 0.09 FALSE TRUE FALSE TRUE

chr6:34176778:34177054:-@chr6:34172628:34172755:-@chr6:34170461:34170574:- 0.11 0 0.25 0.36 0.27 0.05 0.06 FALSE TRUE TRUE TRUE

chr7:80319341-80319246:-@chr7:80319026-80318859:- 0.36 0.06 0 0 0.79 0 0 FALSE TRUE TRUE FALSE

chr9:15317593-15317459:-@chr9:15317113-15316915:- 0.22 0.02 0 0.73 0.79 0 0 FALSE TRUE TRUE FALSE

chr11:58323270:58323365:-@chr11:58322569:58322729:-@chr11:58319324:58319590:- 0.77 0.45 0.57 0.44 0.64 0.51 0.39 FALSE TRUE TRUE TRUE

chr4:141020489-141020274:-@chr4:141019932-141019708:- 0.21 0.03 0.42 0.05 0.2 0.23 0 FALSE TRUE TRUE TRUE

chr7:98359358:98359467:-@chr7:98353477:98353616:-@chr7:98350668:98353130:- 0.31 0.01 0.63 0.43 0.25 0.42 0 FALSE TRUE TRUE TRUE

chr2:25355721-25355558:-@chr2:25355028-25354893:- 0.1 0.07 0 0.7 0.6 0.63 0.5 TRUE FALSE FALSE TRUE

chr3:121171592:121171695:-@chr3:121159747:121159976:-@chr3:121158768:121159255:- 0.11 0.01 0.15 0.05 0.03 0.05 0.16 TRUE FALSE FALSE FALSE

chr5:23809784-23809894:+@chr5:23809969-23810194:+ 0.08 0.01 0 0.31 0.35 0.57 0.29 TRUE FALSE FALSE TRUE

chr5:3803165:3803483:+@chr5:3806419|3806465:3806568:+ 0.43 0.23 0.45 0.7 0.35 0.68 0.58 TRUE FALSE TRUE TRUE

chr11:5837561:5838016:-@chr11:5837289:5837478:-@chr11:5836194:5836374:- 0.08 0.23 0 0.49 0.14 0 0 TRUE FALSE TRUE FALSE

chr17:34952136:34952173:-@chr17:34951885:34951925:-@chr17:34950698:34950847:-@chr17:34950236:34950475:- 0.78 0.98 0.71 0.83 0.8 0.7 0.73 TRUE FALSE TRUE FALSE

chr7:137387371:137387433:-@chr7:137376955:137377117:-@chr7:137375569:137376351:- 0.25 0.02 0.31 0.32 0.48 0.16 0.34 TRUE FALSE TRUE FALSE

chr7:67647410:67647845:+@chr7:67662327:67662526:+@chr7:67667163:67667286:+ 0.46 0.9 0.51 0.47 0.62 0.59 0 TRUE FALSE TRUE TRUE

chr9:105494893:105495133:-@chr9:105494482|105494597:105494415:- 0.78 0.38 0.87 0.82 0.84 0.83 0.68 TRUE FALSE TRUE TRUE

chr14:70351280:70351424:-@chr14:70331269:70331366:-@chr14:70318662:70318940:- 0.63 0.16 0.7 0.75 0.85 0.76 0.44 TRUE FALSE TRUE TRUE

chr5:140670467:140670567:-@chr5:140669022:140669120:-@chr5:140666008:140666208:- 0.42 0.02 0.43 0.64 0 0 0 TRUE TRUE TRUE TRUE

chr8:84689247:84689349|84689376:+@chr8:84689632:84689908:+ 0.08 0.75 0 0.07 0.78 0.77 0.75 TRUE TRUE TRUE TRUE

chr14:18271142:18271323:+@chr14:18276696:18276806:+@chr14:18276826:18276968:+@chr14:18277844:18278101:+ 0.13 0.06 0.2 0.36 0.08 0.14 0 TRUE TRUE FALSE TRUE

chr3:33800191:33800436|33800674:+@chr3:33800956:33801080:+ 0.1 0.03 0.03 0.02 0.04 0.04 0 TRUE TRUE FALSE FALSE

chr3:33800195:33800436:+@chr3:33800582:33800674:+@chr3:33800956:33801080:+ 0.11 0.04 0.05 0.03 0.02 0.04 0 TRUE TRUE FALSE FALSE

chr6:48840996:48841374:-@chr6:48840361:48840453:-@chr6:48838172:48838365:- 0.35 0.6 0.49 0.8 0.32 0.51 0.75 FALSE FALSE TRUE TRUE

chr7:80321181-80320862:-@chr7:80320698-80320302:- 0.69 0.39 0 0.63 0.73 0.72 0.16 FALSE FALSE TRUE TRUE

chr8:25721662-25721786:+@chr8:25722430-25722600:+ 0.24 0.18 0 0.88 0.8 0.82 0.65 FALSE FALSE FALSE TRUE

chr8:25721662-25721786:+@chr8:25722472-25722600:+ 0.59 0.34 0 0.75 0.7 0.74 0.58 FALSE FALSE TRUE TRUE

chr9:15330754:15330910:-@chr9:15330505:15330648:-@chr9:15327796:15327951:- 0.58 0.88 0.72 0.85 0.87 0.77 0.75 FALSE FALSE TRUE FALSE

chr11:101421865:101421954:-@chr11:101419675:101419728:-@chr11:101418814:101419186:- 0.63 0.99 0 0.81 0.36 0 0.85 FALSE FALSE TRUE TRUE

chr1:131698416:131698631:+@chr1:131705596:131705721:+@chr1:131711620:131713464:+ 0.51 0.97 0.38 0.86 0.67 0.86 0.79 FALSE FALSE TRUE TRUE

chr1:155779938:155779439|155779718:-@chr1:155778155:155778851:- 0.43 0.91 0.68 0.41 0.56 0.51 0.71 FALSE FALSE TRUE FALSE

chr2:121169810:121171149:-@chr2:121169199:121169403:-@chr2:121166044:121166718:-

0.66 0.57 0.25 0.26 0.28 0.49 0.24 FALSE FALSE FALSE FALSE

chr2:148672615:148672835:+@chr2:148674464:148674598:+@chr2:148675282:148676026:+ 0.09 0.1 0.26 0.14 0.44 0.1 0.32 FALSE FALSE FALSE FALSE

chr2:165807929:165807519|165807675:-@chr2:165805187:165805377:- 0.13 0.05 0.87 0.95 0.91 0.89 0.92 FALSE FALSE FALSE FALSE

chr2:26387671-26387556:-@chr2:26387311-26386655:- 0.14 0.05 0.44 0.37 0.39 0.5 0.27 FALSE FALSE FALSE TRUE

chr2:27018888-27018979:+@chr2:27020021-27021089:+ 0.51 0.43 0.93 0.59 0.92 0.91 0.93 FALSE FALSE FALSE FALSE

chr4:154023891:154023594|154023782:-@chr4:154020470:154023453:- 0.09 0.04 0.25 0.14 0.31 0.22 0.48 FALSE FALSE FALSE TRUE

chr4:33043857:33043986:+@chr4:33045080:33045199:+@chr4:33049682:33052364:+ 0.82 0.5 0.59 0.37 0.74 0.54 0.28 FALSE FALSE TRUE TRUE

chr5:100806489-100805898:-@chr5:100805451-100805192:- 0.41 0.1 0.76 0.58 0.51 0.7 0.26 FALSE FALSE TRUE TRUE

chr5:112369455-112369551:+@chr5:112369946-112370744:+ 0.1 0.23 0.33 0.3 0.24 0.46 0.69 FALSE FALSE TRUE TRUE

chr6:108065045:108065239:+@chr6:108072967:108073119:+@chr6:108075706:108077127:+ 0.05 0.25 0.17 0.05 0.03 0.11 0.09 FALSE FALSE TRUE FALSE

chr6:133105239:133105605:+@chr6:133106363:133106530:+@chr6:133107864:133107982:+ 0.66 0.35 0.06 0.33 0.08 0.32 0.16 FALSE FALSE TRUE FALSE

chr6:90664022:90664071:-@chr6:90661025|90662955:90659598:- 0.94 0.99 0.75 0.71 0.92 0.83 0.7 FALSE FALSE FALSE FALSE

chr7:141191467-141191342:-@chr7:141191194-141189934:- 0.65 0.88 0.46 0.64 0.72 0.67 0.76 FALSE FALSE TRUE TRUE

chr7:19080042:19081301:+@chr7:19081919:19081993:+@chr7:19082276:19082775:+ 0.33 0.35 0.53 0.65 0.55 0.63 0.62 FALSE FALSE FALSE FALSE

chr7:97579366:97579497:-@chr7:97575588:97575691:-@chr7:97565151:97565300:-@chr7:97550331:97550741:- 0.71 0.66 0.92 0.86 0.89 0.51 0.89 FALSE FALSE FALSE FALSE

chr7:98361242:98361288:-@chr7:98359358:98359467:-@chr7:98350668:98353130:- 0.11 0.21 0.22 0.47 0.27 0.45 0.4 FALSE FALSE FALSE TRUE

chr11:102284961:102285088:+@chr11:102285664:102285718:+@chr11:102287308:102287871:+ 0.63 0.42 0.9 0.68 0.83 0.71 0.87 FALSE FALSE TRUE FALSE

chr11:58330718:58330792|58330802:+@chr11:58331254:58331863:+ 0.24 0.62 0.57 0.23 0.54 0.25 0.73 FALSE FALSE TRUE TRUE

chr11:69903008:69903452|69903719:+@chr11:69903865:69904163:+ 0.63 0.32 0.17 0.16 0.07 0.21 0.04 FALSE FALSE TRUE TRUE

chr11:71019611:71019723:+@chr11:71020006:71020189:+@chr11:71024550:71024711:+@chr11:71025110:71027134:+ 0.19 0.54 0.05 0.17 0.03 0.06 0.12 FALSE FALSE TRUE FALSE

chr15:79247265-79247315:+@chr15:79248152-79248296:+ 0.11 0.06 0.7 0.57 0.72 0.5 0.62 FALSE FALSE FALSE FALSE

chr17:28322646:28322771:+@chr17:28326094|28326152:28326567:+ 0.17 0.02 0.54 0.26 0.48 0.21 0.13 FALSE FALSE TRUE TRUE

chr18:60604490:60602113|60602233:-@chr18:60593990:60596488:- 0.69 0.87 0.97 0.98 0.99 0.93 0.99 FALSE FALSE TRUE FALSE

chr19:61118227:61118350:-@chr19:61117615:61117727:-@chr19:61114984:61117288:- 0.39 0.03 0.08 0.11 0.03 0.05 0.16 FALSE FALSE TRUE FALSE

chr2:24963023:24963156:+@chr2:24963456:24963543:+@chr2:24965560:24965654:+@chr2:24966502:24966674:+ 0.64 0.47 0.38 0.23 0.21 0.16 0.47 FALSE TRUE TRUE FALSE

chr7:80321181-80320302:-@chr7:80320213-80320092:- 0.66 0.18 0 0.25 0.44 0.31 0.19

FALSE TRUE TRUE TRUE

chr1:16665210:16665372|16665505:+@chr1:16667668:16667773:+ 0.98 0.85 0.84 0.84 0.95 0.92 0.84 FALSE TRUE TRUE FALSE

chr1:67038872-67038776:-@chr1:67038591-67038495:- 0.29 0.65 0.83 0.63 0.69 0.68 0.71 FALSE TRUE TRUE TRUE

chr2:119216209:119216307:+@chr2:119216432:119216541:+@chr2:119216629:119217050:+ 0.42 0.92 0.87 0.91 0.9 0.83 0.89 FALSE TRUE TRUE FALSE

chr2:165287784:165287838:-@chr2:165283535|165283600:165283249:- 0.3 0.96 0.59 0.91 0.91 0.9 0.86 FALSE TRUE TRUE TRUE

chr2:165287784:165287838:-@chr2:165283535|165283603:165283249:- 0.38 0.97 0.6 0.93 0.95 0.9 0.92 FALSE TRUE TRUE TRUE

chr2:83724397:83724919:+@chr2:83736510|83736513:83736640:+ 0.51 0.6 0.63 0.22 0.63 0.49 0.54 FALSE TRUE FALSE FALSE

chr3:5563165:5563272:-@chr3:5561760|5561764:5560188:- 0.73 0.69 0.88 0.84 0.96 0.59 0.74 FALSE TRUE FALSE TRUE

chr6:50455825:50456170:-@chr6:50428329:50428420:-@chr6:50370123:50370358:- 0.92 0.19 0.78 0.89 0.75 0.74 0.11 FALSE TRUE TRUE TRUE

chr6:86956259:86956572:+@chr6:86959032:86959274:+@chr6:86963993:86964225:+ 0.49 0.91 0.7 0.91 0.92 0.95 0.97 FALSE TRUE TRUE TRUE

chr7:127802236-127802398:+@chr7:127802786-127803802:+ 0.12 0.27 0.24 0.34 0.34 0.29 0.29 FALSE TRUE TRUE FALSE

chr9:109059908:109059189|109059818:-@chr9:109057932:109058928:- 0.31 0.13 0.49 0.27 0.26 0.29 0.17 FALSE TRUE TRUE TRUE

chr9:13246979:13247021:+@chr9:13249259:13249342:+@chr9:13252071:13252798:+ 0.25 0 0.46 0.05 0.31 0.18 0.1 FALSE TRUE TRUE TRUE

chr10:7598707:7598848:+@chr10:7602829:7602954:+@chr10:7604174:7604383:+ 0.73 0.95 0.95 0.89 0.89 0.94 0.86 FALSE TRUE TRUE FALSE

chr10:81622679-81622408:-@chr10:81622078-81621786:- 0.65 0.83 0.96 0.98 0.98 0.94 0.92 FALSE TRUE TRUE FALSE

chr11:3229869:3230101:+@chr11:3233988:3234059:+@chr11:3234462:3234645:+ 0.55 0.89 0.66 0.49 0.28 0.79 0.86 FALSE TRUE TRUE TRUE

chr11:45895060:45895129:+@chr11:45902211:45902285:+@chr11:45906181:45906467:+ 0.78 0.98 0.94 0.95 0.97 0.95 0.95 FALSE TRUE TRUE FALSE

chr11:67052535:67052618:-@chr11:67051681:67051778:-@chr11:67041481:67042107:- 0.23 0.96 0.87 0.89 0.85 0.89 0.89 FALSE TRUE TRUE FALSE

chr14:105304731:105304836:+@chr14:105307886|105308116:105309298:+ 0.71 0.97 0.92 0.98 0.96 0.95 0.99 FALSE TRUE TRUE FALSE

chr16:14190087:14190232:+@chr16:14191590|14191718:14192918:+ 0.16 0.19 0.03 0.07 0.1 0.21 0.24 FALSE TRUE FALSE TRUE

chr17:34031548:34031690:-@chr17:34030734|34031262:34030566:- 0.97 0.87 0.83 0.9 0.95 0.94 0.92 FALSE TRUE FALSE FALSE

chr18:34609949-34609805:-@chr18:34608627-34608418:- 0.27 0.1 0.38 0.39 0.18 0.16 0.08 FALSE TRUE TRUE TRUE

chr19:61118227:61118350:-@chr19:61117625:61117727:-@chr19:61114984:61117288:- 0.4 0.03 0.12 0.1 0.03 0.03 0.13 FALSE TRUE TRUE FALSE

chr1:106029752-106031233:+@chr1:106032096-106034079:+ 0.52 0.23 0.44 0.45 0.47 0.45 0.37 TRUE FALSE TRUE FALSE

chr2:30415312:30415748:-@chr2:30413860:30413996:-@chr2:30412983:30413246:- 0.09 0.29 0.05 0.02 0.04 0.02 0.05 TRUE FALSE TRUE FALSE

chr2:32363805:32363979:+@chr2:32364219:32364334:+@chr2:32365803:32365874:+@chr2:32367303:32367532:+ 0.49 0.86 0.41 0.51 0.61 0.53 0.69 TRUE FALSE TRUE TRUE

chr3:96557957:96558485:+@chr3:96558889|96558892:96558964:+ 0.63 0.76 0.59 0.56 0.53 0.59 0.43 TRUE FALSE TRUE TRUE

chr7:19715429-19715358:-@chr7:19715259-19714981:- 0.08 0.38 0.08 0.07 0.06 0.11 0.27 TRUE FALSE TRUE TRUE

chr8:71469199:71469325:+@chr8:71472037:71472246:+@chr8:71474488:71476098:+ 0.12 0.06 0.2 0.17 0.17 0.2 0.34 TRUE FALSE FALSE TRUE

chr8:83937649:83937777:+@chr8:83937916:83937933:+@chr8:83938350:83941954:+ 0.89 0.82 0.94 0.98 0.98 0.98 0.99 TRUE FALSE FALSE FALSE

chr9:50895961-50895344:-@chr9:50894081-50892801:- 0.28 0.2 0.38 0.17 0.24 0.39 0.49 TRUE FALSE FALSE TRUE

chr11:62648664:62648816|62648856:+@chr11:62664670:62666359:+ 0.13 0.43 0.22 0.25 0.29 0.26 0.14 TRUE FALSE TRUE FALSE

chr12:106070525:106070615:+@chr12:106073619:106073678:+@chr12:106075007:106077410:+ 0.47 0.21 0.55 0.62 0.26 0.27 0.48 TRUE FALSE TRUE FALSE

chr13:34162977:34163249:+@chr13:34166061:34166136:+@chr13:34177981:34178172:+ 0.51 0.78 0.52 0.28 0.61 0.76 0.53 TRUE FALSE TRUE FALSE

chr17:31521569:31521692:+@chr17:31527310:31528401:+@chr17:31531110:31531325:+ 0.08 0.24 0.16 0.1 0.41 0.14 0.42 TRUE FALSE TRUE TRUE

chr17:47687288-47687188:-@chr17:47687062-47686740:-0.06 0.16 0.04 0.33 0.17 0.08 0.02 TRUE FALSE FALSE FALSE

chr19:4288592:4288658:-@chr19:4288412:4288507:-@chr19:4287823:4287985:- 0.86 0.99 0.88 0.76 0.9 0.87 0.83 TRUE FALSE TRUE FALSE

chr15:62107148:62107465:+@chr15:62149239:62149352:+@chr15:62175467:62175548:+@chr15:62242174:62242301:+ 0.21 0.04 0.19 0.38 0.2 0.39 0.1 TRUE TRUE TRUE FALSE

chr3:90392105-90391993:-@chr3:90391911-90391380:- 0.76 0.95 0.84 0.75 0.76 0.87 0.88

TRUE TRUE TRUE FALSE

chr4:117864405-117864642:+@chr4:117864754-117864888:+ 0.15 0.05 0.07 0.04 0.01 0.04 0.05 TRUE TRUE FALSE FALSE

chr6:13607830:13608063:-@chr6:13607501:13607577:-@chr6:13605787:13605959:-@chr6:13602238:13603487:- 0.13 0.01 0.13 0.09 0.07 0.1 0.07 TRUE TRUE TRUE FALSE

chr6:71505734:71505849:+@chr6:71508686|71508869:71510881:+ 0.04 0.01 0.05 0.05 0.04 0.04 0.03 TRUE TRUE FALSE FALSE

chr11:4704157:4704344:-@chr11:4703904:4703989:-@chr11:4701968:4702220:- 0.07 0 0.04 0 0.1 0 0 TRUE TRUE FALSE FALSE

chr12:44307919:44308114|44308147:+@chr12:44311325:44313435:+ 0.04 0.02 0.02 0.02 0.02 0.02 0.02 TRUE TRUE FALSE FALSE

chr14:18271142:18271323:+@chr14:18276696:18276806:+@chr14:18276826:18276968:+@chr14:18278304:18278545:+ 0.07 0.04 0.05 0.01 0.03 0.01 0.02 TRUE TRUE FALSE FALSE

chr7:97565151:97565300:-@chr7:97558131:97558226:-@chr7:97550331:97550741:- 0.79 0.81 0.73 0.88 0.81 0.91 0.9 TRUE TRUE FALSE TRUE

Publications

1. **Gao, Q.***, Sun, W.*, Ballegeer, M., Libert, C., and Chen, W. (2015). *Predominant contribution of cis-regulatory divergence in the evolution of mouse alternative splicing*. Mol Syst Biol 11, 816.

2. **Gao, Q.**, Sun, W., You, X., Froehler, S., and Chen, W. (2013). *A systematic evaluation of hybridization-based mouse exome capture system*. BMC Genomics *14*, 492.

3. Yuan, Z.*, **Gao, Q.***, He, Y., Zhang, X., Li, F., Zhao, J., and Xue, F. (2012). *Detection for gene-gene co-association via kernel canonical correlation analysis.* BMC Genet 13, 83.

4. **Gao, Q.**, He, Y., Yuan, Z., Zhao, J., Zhang, B., and Xue, F. (2011). *Gene- or region-based association study via kernel principal component analysis.* BMC Genet *12*, 75.

Curriculum Vitae

For reasons of data protection, the Curriculum vitae is not published in the online version.