Gene xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Gene

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

Research paper

Allele-specific splicing effects on *DKKL1* and *ZNF419* transcripts in HeLa cells

Grace Martin [a,1], S. Duygu Selcuklu [a,1], Katherine Schouest [a], Victoria Nembaware [b], Peter C. McKeown [a], Cathal Seoighe [c], Charles Spillane [a,*]

[a] Genetics & Biotechnology Lab, School of Natural Sciences, College of Science, National University of Ireland, Galway, Ireland
[b] Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Private Bag, Rondebosch, 7700 Cape Town, South Africa
[c] School of Mathematics, Statistics and Applied Mathematics, National University of Ireland, Galway, Ireland

ARTICLE INFO

Article history:
Received 9 July 2016
Received in revised form 16 October 2016
Accepted 3 November 2016
Available online xxxx

Keywords:
Alternative splicing
Minigene splicing reporter assay
Single nucleotide polymorphisms
Allele-specific

ABSTRACT

Allele-specific splicing is the production of different RNA isoforms from different alleles of a gene. Altered splicing patterns such as exon skipping can have a dramatic effect on the final protein product yet have traditionally proven difficult to predict. We investigated the splicing effects of a set of nine single nucleotide polymorphisms (SNPs) which are predicted to have a direct impact on mRNA splicing, each in a different gene. Predictions were based on SNP location relative to splice junctions and intronic/exonic splicing elements, combined with an analysis of splice isoform expression data from public sources. Of the nine genes tested, six SNPs led to direct impacts on mRNA splicing as determined by the splicing reporter minigene assay and RT-PCR in human HeLa cells, of which four were allele-specific effects. These included previously unreported alternative splicing patterns in the genes *ZNF419* and *DKKL1*. Notably, the SNP in *ZNF419*, a transcription factor, leads to the deletion of a DNA-binding domain from the protein and is associated with an expression QTL, while the SNP in *DKKL1* leads to shortened transcripts predicted to produce a truncated protein. We conclude that the impact of SNP mutations on mRNA splicing, and its biological relevance, can be predicted by integrating SNP position with available data on relative isoform abundance in human cell lines.

1. Introduction

Alternative splicing is a common regulatory process in eukaryotic cells with roles in many essential biological processes (Kelemen et al., 2013). Since one primary transcript can produce multiple mature mRNA isoforms, a greater diversity of proteins to be encoded by the genome (Kornblihtt et al., 2013). Using deep sequencing, 92–94% of genes in the human genome were found to be alternatively spliced, with alternative mRNA isoforms often occurring at distinct relative frequencies in different tissues (Wang et al., 2008). Genome-wide association studies using large sample cohorts have revealed that such variations are often biologically significant (Xiong et al., 2015). Differential expression of transcript isoforms can be controlled by *cis*-regulatory genetic variants (Coulombe-Huntington et al., 2009; Fu and Ares, 2014). Mutations or polymorphisms in such *cis*-regulatory elements can lead to allele-specific alternative splicing where each of the two possible alleles produces a different mature mRNA (Nembaware et al., 2004; Li et al., 2012).

Understanding the control of alternative splicing is important, as aberrant splicing is implicated in a range of diseases including cystic fibrosis (Niksic et al., 1999), diseases of the skeletal system (Fan and Tang, 2013), lung cancer (Yae et al., 2012) and many neurodegenerative diseases (Mills and Janitz, 2012). Mutations that perturb splicing can have dramatic effects on the final protein product - for example, by producing truncated proteins. Such mutations can also affect protein dosage by generating defective mRNA copies with premature stop codons, which are typically targeted for degradation (Kelemen et al., 2013; Zhang et al., 2014). Several cases of single nucleotide polymorphisms (SNPs) have been reported to cause splicing defects associated with pathologies (Baralle et al., 2003). Indeed, it has been proposed that point mutations that perturb splicing may be the most common cause of hereditary disease (Lopez-Bigas et al., 2005; Ward and Cooper, 2010).

Despite the biological and medical significance of splicing control, it remains challenging to predict how any given SNP or other mutation will affect the splicing pattern of a gene product. We previously performed a genome-screen for exonic SNPs that we predicted to directly affect splicing by changing splice junction strength or affecting the

Abbreviations: SNPs, single nucleotide polymorphisms; RT-PCR, reverse-transcriptase polymerase chain reaction; SNP, single nucleotide polymorphism; QTL, quantitative trait locus; EDB, extra domain B; MiHA, minor histocompatibility antigen; ESEs, exonic splice enhancers; sQTLs, splicing quantitative trait loci.

* Corresponding author.
E-mail address: charles.spillane@nuigalway.ie (C. Spillane).

[1] These authors contributed equally to this manuscript.

http://dx.doi.org/10.1016/j.gene.2016.11.004
0378-1119/

branch point upstream of the splice acceptor site (Nembaware et al., 2008). We also identified those SNPs that displayed a statistical association with alternative splice isoforms, as inferred from exon microarray data from HapMap cell lines (Huang et al., 2007) or from publicly available Expressed Sequence Tag databases. SNPs in nine genes were supported by such analyses. Here, we use a minigene splicing assay to experimentally validate the allele-specific splicing effects of these nine candidate SNPs. We demonstrate that four of these SNPs alter splicing patterns, including two that lead to previously unreported allele-specific splicing in the genes *ZNF419* and *DKKL1*. Finally, we use RT-PCR to directly confirm these splicing effects. Overall, we demonstrate that it is possible to predict SNPs that alter splicing in an allele-specific manner.

2. Material and methods

2.1. Minigene splicing assay using hybrid minigene reporter constructs

Single nucleotide polymorphisms (SNPs) in nine candidate genes were selected for further investigation from our previous bioinformatic pipeline (Table 1). Minigene splicing assays were performed by cloning genomic regions of interest into minigene vectors and identifying the alternatively spliced alleles from the variant segment, as previously described (Pagani et al., 2000; Gaildrat et al., 2010). The modified pTB-α-globin-fibronectin-EDB minigene (Baralle et al., 2003) was used as the expression vector. Fragments containing the exon of interest plus several hundred bases of flanking sequence, including the SNP, were PCR amplified from human genomic DNA. Forward and reverse DNA primers contained an *NdeI* restriction site to allow cloning into the modified minigene vector. In this vector, the extra domain B (EDB) exon of fibronectin is deleted, and the fragment of the gene of interest is inserted. All plasmids were sequenced across the insertion to determine which allele of the SNP was present (GATC, Germany) and to confirm that no SNPs had been introduced by PCR errors. The alternative allele was then produced by site directed mutagenesis, using the Quikchange site directed mutagenesis kit (Agilent Technologies, US) according to manufacturer's instructions. All gene models were accessed from Ensembl; while predictions of amino acid changes were done using SIFT (Ng and Henikoff, 2003) and predictions of altered protein domain content using InterProScan (Hunter et al., 2009).

2.2. Transfections and analysis of hybrid minigene expression

HeLa cells (cervical carcinoma cell line) were seeded at a density of 2×10^5 cells per well in 6 well plates. On the following day, the cells were transfected with 1 μg of either minigene plasmid or empty vector plasmid as mock control using the Lipofectamine 2000 reagent according to the manufacturer's protocol (Invitrogen, US). Twenty four hour post-transfection, total RNA was extracted from the cells using Trizol reagent (Invitrogen, US). Total RNA was reverse transcribed with random hexamer primers using QuantiTect Reverse Transcription Kit (Qiagen, US). PCR amplification was performed using cDNA templates and products visualised by 2% w/v agarose gel electrophoresis. The amplification products were gel-extracted and confirmed by sequencing as before.

2.3. Detection of expression patterns from tissue panel datasets

Microarray datasets derived from 'normal' (i.e. non-cancer) tissue panels; Su_Normal_2 (Su et al., 2004) and Shyamsundar_Normal (Shyamsundar et al., 2005) were obtained from the Oncomine™ database available at www.oncomine.org. *DKKL1* expression (log-2 median-centered intensity) was retrieved from Su_Normal dataset with 22 different tissues and *ZNF419* expression was retrieved from Shyamsundar_Normal dataset with 26 different tissues that displayed high expression in at least one of the tissues tested.

3. Results & discussion

3.1. Validation of SNPs causal for altered splicing patterns

To functionally test the nine SNPs predicted to lead to altered splicing patterns in vivo, we developed hybrid minigene constructs using the minigene splicing reporter vector (Fig. 1) containing the exons in which the nine candidate SNPs were located. Of these nine SNPs, four (in genes *PARP2, TCL6, ZNF419* and *DKKL1*) were located close to splice junctions and were predicted to affect the strength of the splice donor or the acceptor site. The remaining five SNPs (in genes *ATP5SL*, *GLO1*, *RBM47*, *SUPT5H* and *TCN6*) were located further from the splice junction, within exonic splice enhancers (Table 1). For four of these genes (*ATP5SL, DKKL1, PARP2* and *ZNF419*), the candidate SNPs displayed qualitative differences in mRNA splicing between the two alleles tested (Fig. 2). For the remaining five SNPs, we identified two that led to the production of alternatively spliced forms: however, in these genes (*SUPT5H* and *TCL6*), the same alternatively spliced form was produced by both alleles tested (Suppl. Fig. S1A–B). For the remaining three candidates, no evidence of alternative splicing could be found for either isoform. We note that this should be considered as a lower bound on the number of accurately predicted SNP candidates, as we only tested for qualitative effects on splicing and not for effects on the relative frequencies of the transcript isoforms.

3.2. Polymorphisms in PARP2 and ATP5SL are demonstrated to be associated with allele-specific splicing in human cells

Consistent with a previous report (Coulombe-Huntington et al., 2009), we could confirm by the minigene splicing reporter assay that the SNPs rs2297616 and rs1043413 (C/G) were associated with allele-specific splicing of *PARP2* (Fig. 2A) and *ATP5SL* (Fig. 2B), respectively. The C/T SNP at rs2297616 in *PARP2* is intronic and was previously

Table 1
Candidate genes and SNPs used in this study.

Gene	Ensembl ID	Functional indication[a]	SNP ID/location	Ancestral/derived allele[b]	Prediction[c]
ATP5SL	ENSG0000105341	N/A	rs1043413/exon 5	C/G	Exon 5 skipped
DKKL1	ENSG0000104901	Spermatogenesis (Kohn et al., 2010; Yan et al., 2015)	rs2303759/exon 4	T/G	Exon 4 skipped
GLO1	ENSG000012476	Cancer (Santarius et al., 2010)	rs4746 (rs2736654)/exon 4	A/C	Exon 4 skipped
PARP2	ENSG0000129484	Spermiogenesis (Meyer-Ficca et al., 2011)	rs2297616/intron 2–3	C/T	Exon 2 alternative donor site
RBM47	ENSG0000163694	N/A	rs2307046/exon 4	C/T	Exon 4 skipped
SUPT5H	ENSG0000196235	Telomerase reverse transcriptase promoter-binding protein (Chen et al., 2015)	rs1932/exon 3	C/T	Exon 2 alternative acceptor site
TCL6	ENSG0000187621	Leukemogenesis (Saitou et al., 2000)	rs2296310/intron 2–3	C/A	Exon 3 skipped
TCN2	ENSG0000185339	VitB(12) delivery (Stanislawska-Sachadyn et al., 2010)	rs1081198/exon 6	T/G	Exon 6 skipped
ZNF419	ENSG0000105136	Immune response (Broen et al., 2011)	rs11672136/intron 1–2	T/A	Exon 2 skipped

N/A indicates no available evidence for function.
[a] Based on reference cited.
[b] Based on RefSNP (NCBI).
[c] Predicted splicing effect of the SNP on the gene of interest.

Please cite this article as: Martin, G., et al., Allele-specific splicing effects on *DKKL1* and *ZNF419* transcripts in HeLa cells, Gene (2016), http://dx.doi.org/10.1016/j.gene.2016.11.004