

SCIENTIFIC REPORTS



OPEN

Computational identification and validation of alternative splicing in ZSF1 rat RNA-seq data, a preclinical model for type 2 diabetic nephropathy

Chi Zhang¹, Ken Dower², Baohong Zhang ¹, Robert V. Martinez², Lih-Ling Lin² & Shanrong Zhao¹

Obese ZSF1 rats exhibit spontaneous time-dependent diabetic nephropathy and are considered to be a highly relevant animal model of progressive human diabetic kidney disease. We previously identified gene expression changes between disease and control animals across six time points from 12 to 41 weeks. In this study, the same data were analysed at the isoform and exon levels to reveal additional disease mechanisms that may be governed by alternative splicing. Our analyses identified alternative splicing patterns in genes that may be implicated in disease pathogenesis (such as *Shc1*, *Serpinc1*, *Epb4.115*, and *Il-33*), which would have been overlooked in standard gene-level analysis. The alternatively spliced genes were enriched in pathways related to cell adhesion, cell-cell interactions/junctions, and cytoskeleton signalling, whereas the differentially expressed genes were enriched in pathways related to immune response, G protein-coupled receptor, and cAMP signalling. Our findings indicate that additional mechanistic insights can be gained from exon- and isoform-level data analyses over standard gene-level analysis. Considering alternative splicing is poorly conserved between rodents and humans, it is noted that this work is not translational, but the point holds true that additional insights can be gained from alternative splicing analysis of RNA-seq data.

Diabetic nephropathy (DN) accounts for about 44% of patients starting treatment for kidney failure each year, and is the leading cause for end-stage renal disease that requires dialysis or renal replacement therapy for patient survival¹. The underlying mechanisms of the disease progression are not fully understood. Although DN is generally considered to be a nonimmune-mediated degenerative disease, complement proteins and immunoglobulins can be detected in diseased glomeruli².

Several groups have demonstrated that the obese ZSF1 rat is one of relatively few rodent models that mimic the disease progression of human type 2 DN^{3,4}. Previously, we performed a high-resolution longitudinal study of age-matched lean and obese ZSF1 rats spanning from early DN to end-stage renal disease³. The analyses included metabolic endpoints, renal histology and ultrastructure, fibrosis biomarker evaluation, and gene-level RNA-seq analyses of glomerular-enriched tissue over the course of disease progression³.

RNA-seq is routinely used to profile transcriptome changes in disease versus normal tissues⁵. While most RNA-seq studies focus on expression at the gene level, it is now appreciated that the majority of mammalian genes can be alternatively spliced to generate multiple different isoforms that encode proteins with diverse, and even opposing, biological functions⁶. In the human Gencode annotation Release 25, the number of annotated transcripts is more than three times the number of annotated genes; for protein coding genes, this ratio increases to 7:1. It has also been reported that 50–60% of disease-causing mutations in the human gene mutation database affect splicing^{7,8}. These findings have resulted in a growing interest in understanding alternative splicing patterns

¹Precision Medicine, Early Clinical Development, Pfizer Worldwide Research and Development, Cambridge, MA, 02139, USA. ²Inflammation & Immunology Research Unit, Pfizer Worldwide Research and Development, Cambridge, MA, 02139, USA. Correspondence and requests for materials should be addressed to S.Z. (email: Shanrong.Zhao@pfizer.com)

Week	DE_gene	DE_tx_gene	DE_tx
12	282	320	340
20	227	243	256
24	299	332	360
29	395	462	499
34	1152	1247	1389
41	1524	1635	1855

Table 1. Numbers of differentially expressed genes (DE_gene) and differentially expressed transcripts (DE_tx) between obese and lean animals at each time point. DE_tx_gene indicates the number of genes that correspond to the transcripts in DE_tx.

in disease. Moreover, targeting splicing regulatory proteins and/or specific key altered splicing events has emerged as a therapeutic strategy, for example, in rare diseases⁹ and cancers¹⁰.

Several computational tools have been developed to analyse alternative spliced transcripts in RNA-seq data¹¹. These tools can be divided into two categories based on their functionality. Tools in the first category include Cufflinks¹², RSEM¹³, Kallisto¹⁴, and Salmon¹⁵, which can be applied to analyse known or annotated transcript isoforms. In this study, we chose Salmon, one of the top performers in speed and accuracy based on our internal evaluation¹⁶, for this type of analysis. Tools in the second category include MISO¹⁷, MAJIQ¹⁸, and rMATS¹⁹, which can be used to analyse RNA-seq data at the exon level and to detect known and novel splicing events. These latter methods are especially useful for organisms with limited annotation on isoforms. We chose rMATS for this type of analysis in the present study.

The longitudinal RNA-seq data of isolated glomeruli from lean (non-diseased) and obese (diseased) ZSF1 rats were comprehensively analysed at both the isoform and the exon levels. We report several alternative splicing patterns in obese versus lean animals in genes involved in metabolism, immune response, podocyte dedifferentiation, and apoptosis. Our findings provide an example of how isoform and exon analyses can identify molecular changes that may contribute to disease pathogenesis, which would be overlooked in standard gene-level analyses. In addition, we validated by qRT-PCR the existence of selected novel mRNA isoforms that were identified computationally.

Results

RNA-seq analyses of ZSF1 glomeruli at the isoform level. The RNA-seq dataset was described in detail in a previous report³. Briefly, glomeruli were enriched from the kidneys of ZSF1 lean and obese animals at 12, 20, 24, 29, 34, and 41 weeks of age for RNA-seq library preparation. The study consisted of five lean and five obese RNA samples per age group with the exception of week 41, because one lean and two obese animals died over the study period. The metadata for all 57 RNA samples are listed in Supplementary Table S1.

The RNA-seq dataset was reanalysed using Salmon¹⁵ 0.8.2 and the Ensembl v87 rat transcriptome annotation. At each time point, the group average of transcript-level \log_2 TPM (\log_2 of Transcript Per Million) values was calculated for lean and obese animals. The scatter plots of lean versus obese groups from this analysis are shown in Supplementary Fig. S1. Transcripts with over 16-fold TPM differences in each comparison were identified and labelled. Next, differential expression analyses between lean and obese animals at each time point were performed at both the gene and transcript levels using the Bioconductor package edgeR²⁰. The numbers of differentially expressed genes and differentially expressed transcripts from this analysis are summarized in Table 1. In concordance with the gene-level analyses³, the number of differentially expressed transcripts increased with disease progression from weeks 12 to 41, and was particularly evident at weeks 34 and 41. The differentially expressed transcripts were then grouped by their corresponding genes to obtain the number of genes with one or more differentially expressed transcripts. At every time point, the number of genes with one or more differentially expressed transcripts was always higher than the corresponding number of differentially expressed genes. Thus, more differentially expressed genes were discovered at the isoform level than at the gene level.

We were particularly interested in genes that displayed significant changes at the transcript level but not at the gene level, because such changes would be overlooked in gene-level analysis. To identify such genes, we proceeded as follows. For each gene, the FDR (false discovery rate) at the gene level (FDR_{gene}) and the minimum FDR across all transcripts (\min_FDR_{trans}) were first calculated at all time points. Then, $(-\log_{10}\min_FDR_{trans}) - (-\log_{10}FDR_{gene})$ was calculated at each time point and the maximum value for each gene across all time points was used to rank all the genes. The top 200 genes were visually inspected, and three DN-associated genes, *Shc1*, *Serpinc1*, and *Epb4.115*, were examined in detail.

Shc1 (SHC adaptor protein 1) has two annotated isoforms in rat Ensembl v87: ENSRNOT00000028038 and ENSRNOT00000056652 because of alternative promoter usage (Fig. 1). ENSRNOT00000028038 can be translated to the p66^{Shc} protein isoform, which is predominantly found in epithelial cells²¹. ENSRNOT00000056652 can be translated to p52^{Shc} and p46^{Shc} protein isoforms through alternative translation start sites. Unlike p66^{Shc}, the p52^{Shc} and p46^{Shc} proteins have been detected in almost all cell types²¹. The three protein isoforms of *Shc1* reportedly have distinct biological functions. p66^{Shc} knockout mice were protected against vascular, cardiac, and renal damage, while overexpression of p66^{Shc} led to cytochrome c release and apoptosis^{22,23}. p52^{Shc}, the most well studied isoform, acts as a bridge between growth factor receptors and the adaptor Grb2, which binds guanine nucleotide exchange factor Sos²⁴ resulting in the activation of Ras and downstream MAPK signalling pathways. p46^{Shc} was recently reported to be a negative regulator of mitochondrial thiolase, which is involved in fatty acid metabolism²⁵.

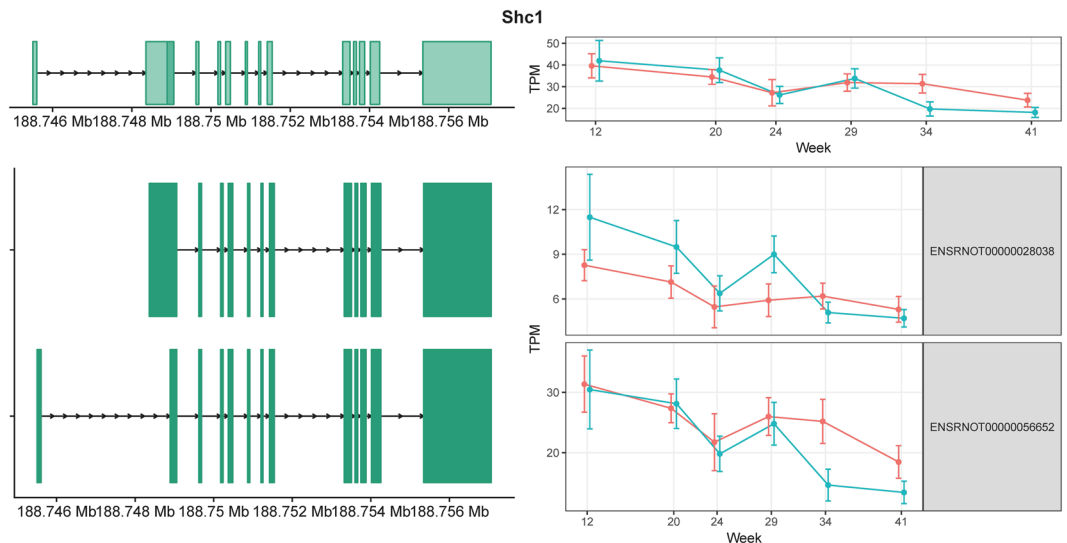


Figure 1. Exon structures and expression levels of *Shc1* and its two transcripts. The gene structure (top left) and isoform structures (bottom left) are shown with the chromosome coordinates on the left. *Shc1* is on the “+” strand. The exons are shown in green. The corresponding longitudinal TPM values estimated from the RNA-seq data are shown on the right. Lean and obese animals are indicated in red and blue, respectively. The error bars indicate the 95% confidence intervals.

Mice with low levels of $p46^{Shc}$ are lean and resistant to obesity²⁵. The expression levels for *Shc1* isoforms in our dataset, calculated by Salmon, are shown in Fig. 1. The comparison and statistical analysis results between obese and lean groups across different time points were summarized in Supplementary Table S2. The expression of the mRNA corresponding to $p66^{Shc}$ (ENSRNOT00000028038) was elevated in obese animals at 12, 20, and 29 weeks but not at later time points, which may be linked to obesity-induced oxidative stress during the early and intermediate stages of DN. In contrast, the expression levels of the mRNA corresponding to $p46^{Shc}/p52^{Shc}$ (ENSRNOT00000056652) were similar between the lean and obese groups at 12, 20, 24, and 29 weeks of age, but reduced in obese animals at 34 and 41 weeks of age. This reduction may be indicative of dysregulated fatty acid metabolism and/or cell proliferation in late stage DN.

Serpinc1 (serpin family C member 1) encodes antithrombin III (ATIII), which is a serine protease inhibitor in the coagulation cascade²⁶. Evidence suggests that ATIII deficiency is associated with kidney injury²⁶. There are three annotated isoforms (Fig. 2), but their biological functions are unknown. In our analysis, ENSRNOT00000003748 showed higher expression in obese animals than in lean animals at weeks 34 and 41, whereas the expression levels of ENSRNOT00000046512 and ENSRNOT00000076043 at these time points were depressed in obese animals (Supplementary Table S2). Therefore, at these later time points, the direction of the expression changes was opposite for different isoforms, resulting in an expression profile that appeared relatively unchanged in the gene-level analysis (Fig. 2). This result illustrates how an isoform-level analysis of RNA-seq data can unmask more accurately changes in mRNA abundance.

Epb4.115 (the erythrocyte protein band 4.1-like5) encodes a mesenchymal-specific protein that is induced during the process of epithelial-mesenchymal transition (EMT)²⁷. *Epb4.115* binds to p120 catenin, and causes E-cadherin internalization²⁸. It also binds to paxillin, promotes focal adhesion dynamics, and has been implicated in cell motility²⁸. The expression levels of the two annotated isoforms of *Epb4.115* are shown in Fig. 3. The comparison and statistical analysis results between obese and lean groups across different time points were summarized in Supplementary Table S2. ENSRNOT00000003428 had higher expression in the obese rats, while ENSRNOT00000042862 had higher expression in the lean controls. Although the biological functions of the isoforms of *Epb4.115* are not well studied, the consistent difference in isoform usage in obese and lean animals indicates the two isoforms may play different roles in EMT during DN progression²⁹.

RNA-seq analyses of known and novel splicing at the exon level. We next performed an exon-level analysis of the RNA-seq data to examine alternative splicing in an unbiased manner. Raw sequence reads were mapped to the Ensembl v87 annotated rat genome using STAR³⁰. To find as many novel junctions as possible, the STAR 2-PASS protocol was used. The rMATS package was then applied to discover both annotated and novel exons, and to calculate differences in exon inclusion between obese and lean animals. The total numbers of sequencing reads were very close for all RNA samples (see Supplementary Table S1); therefore, the rMATS reported read counts were not further normalized. By way of illustration, below we describe our findings for *Il-33* that displayed alternative splicing patterns.

Interleukin (IL)-33 is a member of the IL-1 cytokine family, which is widely expressed in epithelial and endothelial cells, and which mediates tissue inflammation and repair responses^{31–33}. Recent evidence from both clinical and preclinical studies indicates that IL-33 and its receptor ST2 (suppression of tumorigenicity 2) are involved in the pathogenesis of chronic kidney disease, systemic lupus nephritis, and acute kidney injury³⁴. A

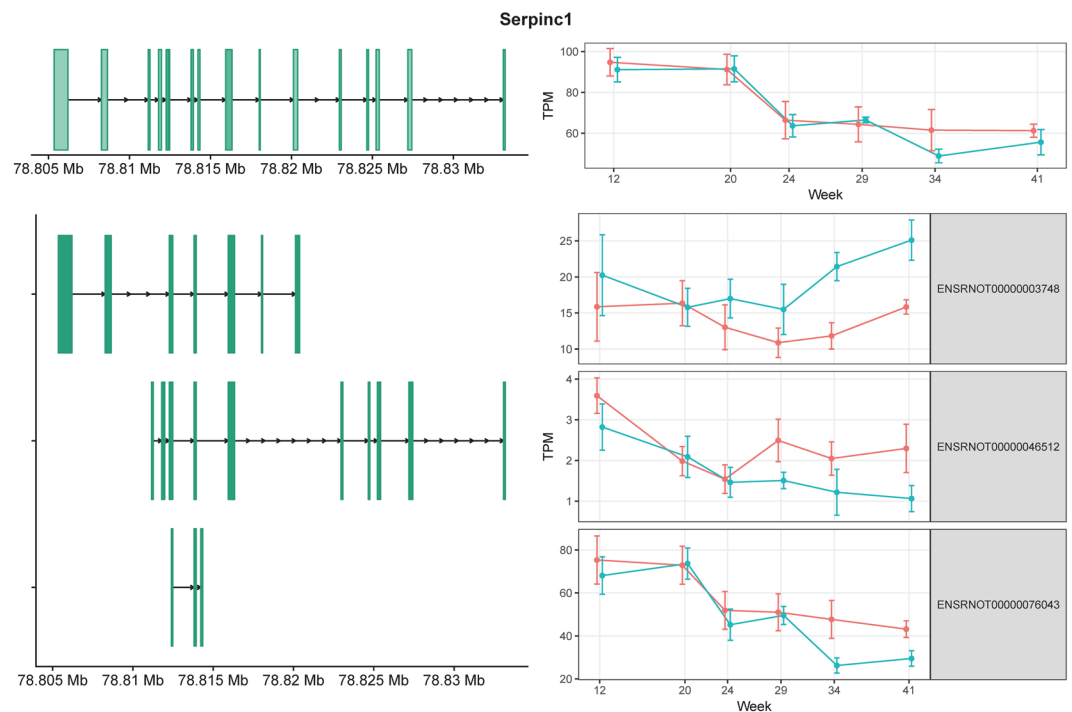


Figure 2. Exon structures and expression levels of *Serpinc1* and its three transcripts. *Serpinc1* is on the “+” strand. The exons are shown in green. The corresponding longitudinal TPM values estimated from the RNA-seq data are shown on the right. Lean and obese animals are indicated in red and blue, respectively. The error bars indicate the 95% confidence intervals.

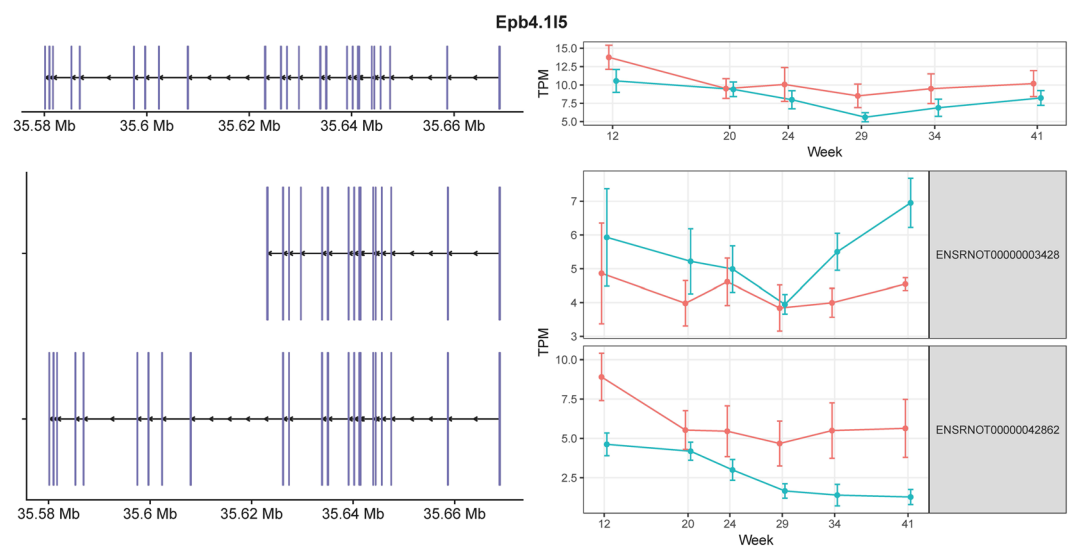


Figure 3. Exon structures and expression levels of *Epb4.115* and its two transcripts. *Epb4.115* is located on the “-” strand. The exons are shown in purple. The corresponding longitudinal TPM values estimated from the RNA-seq data are shown on the right. Lean and obese animals are indicated in red and blue, respectively. The error bars indicate the 95% confidence intervals.

novel skipped exon in *Il-33* was discovered in our analysis, and is illustrated in Fig. 4A. The exon-level RNA-seq analyses and qRT-PCR (Fig. 4B) both confirmed the presence of this novel exon. Moreover, the expression of the included isoform was higher in obese rats than in lean controls at week 41, whereas, for the excluded isoform, there was less of a difference between obese and lean animals at week 41 by both RNA-seq and qRT-PCR (Fig. 4). The comparison and statistical analysis results for gene expression, exon inclusion and exclusion between obese and lean groups across different time points were summarized in Supplementary Table S3. The qRT-PCR results matched well with the RNA-seq data in most cases, except for the excluded form in obese animals at week 34.

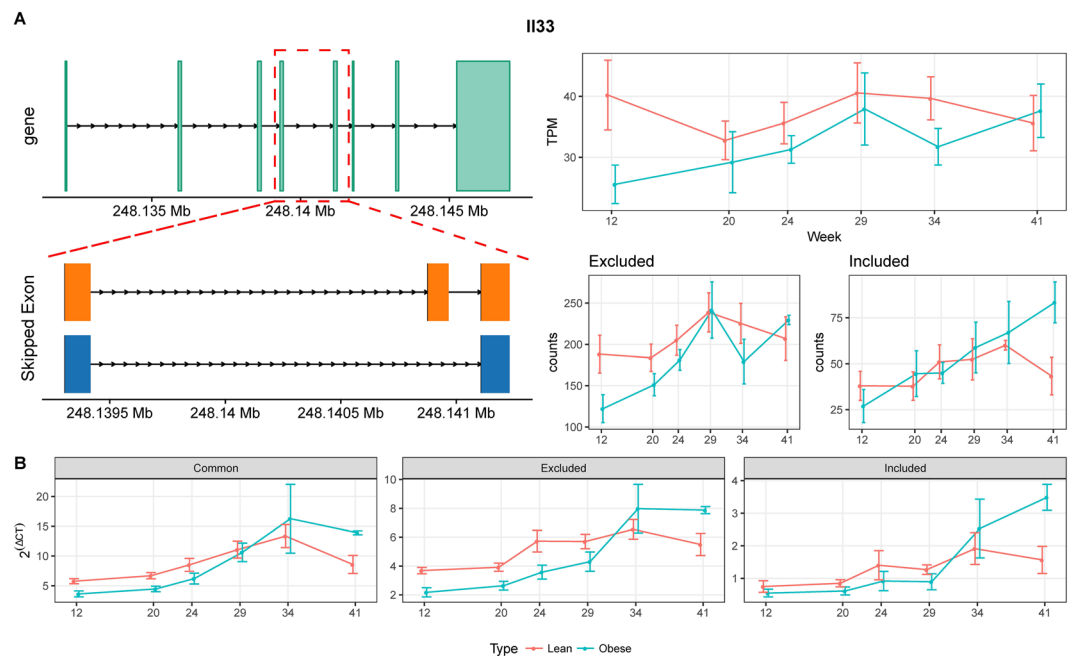


Figure 4. Differential splicing of *Il-33*. **(A)** Gene structure (top left) and the skipped exon event at the 3' end of *Il-33* (bottom left). The gene-level TPM values and the corresponding read counts supporting exon inclusion and exclusion are shown on the right. **(B)** Expression values obtained by qRT-PCR. The expression values are $2^{\Delta\Delta Ct}$ normalized to the housekeeping gene *Gusb*. All the error bars indicate the 95% confidence intervals.

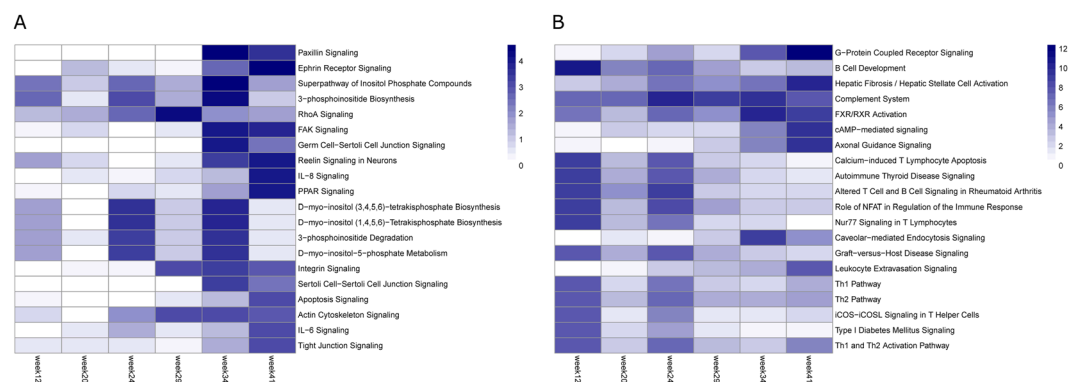


Figure 5. Top 20 enriched pathways associated with the differentially spliced genes **(A)** and differentially expressed genes **(B)**. The top 20 canonical pathways were selected according to the smallest P values reported by IPA across all time points, and coloured based on the corresponding $-\log_{10}(P \text{ value})$.

Whether the included or excluded isoforms play different roles is not clear, and additional research is needed to uncover the function of this novel excluded isoform of *Il-33* identified from the exon-level RNA-seq data analysis.

After stringent filtering (details in methods section), the lists of genes with statistically significant differences at the exon inclusion level were uploaded to Ingenuity Pathway (IPA) for enrichment analysis. A total of 513 alternatively spliced genes mapped to the IPA database. The enriched canonical pathways in the exon-level analyses were ranked by the P values reported by IPA across all time points, and the top 20 pathways are shown in Fig. 5A. For comparison, the top 20 enriched pathways from a gene-level differential expression analysis, as opposed to a differentially spliced gene analysis, are shown in Fig. 5B. Different pathways were enriched between these two analyses. The alternatively spliced genes were enriched in pathways related to cell adhesion (paxillin signalling, FAK signalling, and integrin signalling), cell-cell interactions/junctions (ephrin receptor signalling, tight junction signalling, and germ cell-Sertoli cell junction signalling), cytoskeleton signalling, and inositol phosphate signalling, whereas the differentially expressed genes were enriched in pathways related to the immune response, G protein-coupled receptor, and cAMP signalling.

Podocytes are specialized epithelial cells that are essential to the glomerular filtration barrier³⁵. Our previous study showed there were no significant changes in podocyte numbers per glomerulus between lean and obese animals, despite a significant decline in the abundance of podocyte-specific mRNAs³. This result, combined with observed podocyte foot process effacement by transmission electron microscopy, suggested that podocyte

epithelial-mesenchymal transition (EMT) may be occurring in obese animals as they approached end-stage renal disease³⁶. Consistent with this hypothesis, in the pathway analysis of genes showing differences in alternative splicing patterns, enriched pathways included cytoskeleton reorganization, changes in cell morphology, and mobility (Fig. 5A). Importantly, these pathways are heavily involved during the process of EMT^{37,38}. There is also clinical and preclinical evidence that suggests a role for alternative splicing in cytoskeleton reorganization during EMT³⁹. Therefore, the pathway enrichment of differentially spliced genes reveals potential additional mechanisms not revealed by the gene-level analyses. By combining both differential splicing and standard expression analyses, additional hypotheses with respect to mechanisms of disease pathogenesis can be formulated in this manner.

Discussion

Recently, alignment-free tools, such as Kallisto¹⁴ and Salmon¹⁵, have been developed for transcript quantification. However, the accuracy of the quantification relies heavily on accurate isoform annotation, which is far from complete even for well-annotated organisms such as human and mouse. Exon-level analyses using tools such as MISO¹⁷, MAJIQ¹⁸, and rMATS¹⁹ can reveal novel isoforms and are especially attractive for less well-annotated species, such as rat.

Here we applied Salmon and rMATS to analyse RNA-seq data at the isoform- and exon-levels, respectively, in a rat model of type 2 DN. These analyses identified alternative splicing patterns in genes that may be implicated in disease pathogenesis (such as *Shc1*, *Serpinc1*, *Epb4.1l5*, and *Il-33*), which would have been overlooked in standard gene-level analysis. The novel splice form in *Il-33* identified from our exon-level analysis was validated by qRT-PCR. In general, there was very good concordance between the RNA-seq data and the qRT-PCR analysis, although there were situations in which the results did not precisely align. We speculate there are three main reasons for the discrepancy. First, qRT-PCR can suffer from primer specificity and PCR artifacts⁴⁰. Second, RNA-seq is vulnerable to the general biases and errors inherent in next-generation sequencing upon which it is based. The fragments are not uniformly sampled and sequenced due to preferential sites of fragmentation, and variable primer and tag nucleotide composition effects^{41,42}. Third, the exon and isoform annotations may be incomplete or inaccurate. Our findings point to the necessity for orthogonal, experimental approaches to validate novel splicing events that are identified through bioinformatics approaches.

Alternative exons, especially minor-isoform ones, were shown to be less conserved among species than constitutive exons⁴³. For a particular gene in rat, its alternative splicing pattern in different conditions is not necessarily the same as its splicing pattern in its orthologous gene in human. The results reported in this work are not translational. However, our point holds true that much more insights can be gained from alternative splicing analysis of RNA-seq data.

RNA-seq has become the *de facto* standard in transcriptome profiling, and most studies have focused only on gene-level analyses. However, mRNAs and proteins, rather than genes, are the actual functional units in living cells. In this study, we explored the role of alternative splicing during DN progression in glomeruli-enriched tissues of ZSF1 rats and demonstrated that broader and deeper biological insights can be obtained if RNA-seq data are analysed at the transcript and exon levels. Subtle differences, such as alternative isoform usage and exon inclusion/exclusion, can be missed by gene-level RNA-seq data analysis. Therefore, we believe that significant mechanistic insights can be gained from the simultaneous analysis of RNA-seq data at the gene, transcript, and exon levels.

Materials and Methods

RNA-seq isoform analyses. Raw sequencing reads were processed by Salmon¹⁵ 0.8.2 using Ensembl v87 rat annotation with default options. The resulting isoform-level counts tables were analysed using the R/Bioconductor package edgeR²⁰ 3.16.5 for normalization and differential expression between lean and obese groups at each time point. The quasi-likelihood F-test glmQLFit was used instead of a likelihood ratio test because of the relatively small numbers of replicates at each time point.

RNA-seq exon-level analyses. Raw sequencing reads were first mapped to the rat genome using STAR³⁰ 2.5.2a with the 2-PASS option. The resulting 2-PASS bam files were processed by rMATS¹⁹ for exon inclusion level differences between lean and obese groups at each time point. Both exon-exon junction reads and the reads that mapped to the exons were used in the analyses. The resulting alternative splicing tables in five categories (alternative 5' splice site, alternative 3' splice site, mutually exclusive exons, retained introns, and skipped exons) were merged.

Ingenuity pathway analyses. Pathway enrichment analyses were performed using IPA for both differentially expressed genes and differentially spliced genes. The Ensembl ID and FDR were used as input for pathway enrichment analyses. For the gene-level enrichment analyses, the following filters were applied: average read counts >2, log₂fold_change >2, and FDR <0.05. For the exon-level analyses, the filters were: FDR <0.05, abs (inclusion level difference) >0.05, percent sample expressed >80%, and average read counts >10.

qRT-PCR validation. Custom primer probe sequences labelled with 6-carboxyfluorescein (FAM) were ordered from Integrated DNA Technologies (IDT, Coralville, IA). The nucleotide sequences of these primers and probes, and their relative positions with respect to detection of alternative splice forms, are provided in Supplementary Table S4. The same total RNA samples that were subjected to RNA-seq were used for the qRT-PCR analysis. All qRT-PCR reactions were performed using a Taqman RNA-to-Ct 1-Step Kit (Thermo Fisher; catalogue number 4392656) following the manufacturer's protocol, on an Applied Biosystems Quantstudio 12 K Flex Real-Time PCR System (Thermo Fisher). Ct values for each primer/probe set were normalized to the Ct values for the housekeeping gene *Gusb* for each sample to calculate 2^{-Δct} values.

Availability of data. The RNA-seq data in SRP094779 were downloaded from Sequence Read Archive.

References

- Collins, A. J. *et al.* United States Renal Data System 2011 Annual Data Report: Atlas of chronic kidney disease & end-stage renal disease in the United States. *Am. J. Kidney Dis.* **59**(A7), e1–420, <https://doi.org/10.1053/j.ajkd.2011.11.015> (2012).
- Xiao, X. *et al.* Cellular and humoral immune responses in the early stages of diabetic nephropathy in NOD mice. *J. Autoimmun.* **32**, 85–93, <https://doi.org/10.1016/j.jaut.2008.12.003> (2009).
- Dower, K. *et al.* High resolution molecular and histological analysis of renal disease progression in ZSF1 fa/faCP rats, a model of type 2 diabetic nephropathy. *PLoS One* **12**, e0181861, <https://doi.org/10.1371/journal.pone.0181861> (2017).
- Rafikova, O., Salah, E. M. & Tofovic, S. P. Renal and metabolic effects of tempol in obese ZSF1 rats—distinct role for superoxide and hydrogen peroxide in diabetic renal injury. *Metabolism* **57**, 1434–1444, <https://doi.org/10.1016/j.metabol.2008.05.014> (2008).
- Ramakrishnan, V. R. *et al.* RNA sequencing and pathway analysis identify tumor necrosis factor alpha driven small proline-rich protein dysregulation in chronic rhinosinusitis. *Am. J. of Rhinol. Allergy* **31**, 283–288, <https://doi.org/10.2500/ajra.2017.31.4457> (2017).
- Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476, <https://doi.org/10.1038/nature07509> (2008).
- Lopez-Bigas, N., Audit, B., Ouzounis, C., Parra, G. & Guigo, R. Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett.* **579**, 1900–1903, <https://doi.org/10.1016/j.febslet.2005.02.047> (2005).
- Wang, G. S. & Cooper, T. A. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat. Rev. Genet.* **8**, 749–761, <https://doi.org/10.1038/nrg2164> (2007).
- Hoy, S. M. Nusinersen: First Global Approval. *Drugs* **77**, 473–479, <https://doi.org/10.1007/s40265-017-0711-7> (2017).
- Lee, S. C. & Abdel-Wahab, O. Therapeutic targeting of splicing in cancer. *Nat. Med.* **22**, 976–986, <https://doi.org/10.1038/nm.4165> (2016).
- Zhang, C., Zhang, B., Vincent, M. S. & Zhao, S. Bioinformatics tools for RNA-seq gene and isoform quantification. *Next Generat. Sequenc. & Applic.* **03**, 140, <https://doi.org/10.4172/2469-9853.1000140> (2016).
- Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515, <https://doi.org/10.1038/nbt.1621> (2010).
- Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323, <https://doi.org/10.1186/1471-2105-12-323> (2011).
- Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527, <https://doi.org/10.1038/nbt.3519> (2016).
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419, <https://doi.org/10.1038/nmeth.4197> (2017).
- Zhang, C., Zhang, B., Lin, L. L. & Zhao, S. Evaluation and comparison of computational tools for RNA-seq isoform quantification. *BMC Genomics* **18**, 583, <https://doi.org/10.1186/s12864-017-4002-1> (2017).
- Katz, Y., Wang, E. T., Airolidi, E. M. & Burge, C. B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* **7**, 1009–1015, <https://doi.org/10.1038/nmeth.1528> (2010).
- Vaquero-Garcia, J. *et al.* A new view of transcriptome complexity and regulation through the lens of local splicing variations. *Elife* **5**, e11752, <https://doi.org/10.7554/eLife.11752> (2016).
- Shen, S. *et al.* rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. USA* **111**, E5593–5601, <https://doi.org/10.1073/pnas.1419161111> (2014).
- Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140, <https://doi.org/10.1093/bioinformatics/btp616> (2010).
- Ventura, A., Luzzi, L., Pacini, S., Baldari, C. T. & Pelicci, P. G. The p66Shc longevity gene is silenced through epigenetic modifications of an alternative promoter. *J. Biol. Chem.* **277**, 22370–22376, <https://doi.org/10.1074/jbc.M200280200> (2002).
- De Marchi, E., Baldassari, F., Bononi, A., Wieckowski, M. R. & Pinton, P. Oxidative stress in cardiovascular diseases and obesity: role of p66Shc and protein kinase C. *Oxid. Med. Cell Longev.* **2013**, 564961, <https://doi.org/10.1155/2013/564961> (2013).
- Napoli, C. *et al.* Deletion of the p66Shc longevity gene reduces systemic and tissue oxidative stress, vascular cell apoptosis, and early atherosclerosis in mice fed a high-fat diet. *Proc. Natl. Acad. Sci. USA* **100**, 2112–2116, <https://doi.org/10.1073/pnas.0336359100> (2003).
- Ravichandran, K. S. Signaling via Shc family adapter proteins. *Oncogene* **20**, 6322–6330, <https://doi.org/10.1038/sj.onc.1204776> (2001).
- Tomilov, A. *et al.* p46Shc inhibits thiolase and lipid oxidation in mitochondria. *J. Biol. Chem.* **291**, 12575–12585, <https://doi.org/10.1074/jbc.M115.695577> (2016).
- Lu, Z., Wang, F. & Liang, M. SerpinC1/Antithrombin III in kidney-related diseases. *Clin. Sci. (Lond.)* **131**, 823–831, <https://doi.org/10.1042/CS20160669> (2017).
- Hashimoto, A. *et al.* ZEB1 induces EPB41L5 in the cancer mesenchymal program that drives ARF6-based invasion, metastasis and drug resistance. *Oncogenesis* **5**, e259, <https://doi.org/10.1038/oncsis.2016.60> (2016).
- Hirano, M., Hashimoto, S., Yonemura, S., Sabe, H. & Aizawa, S. EPB41L5 functions to post-transcriptionally regulate cadherin and integrin during epithelial-mesenchymal transition. *J. Cell. Biol.* **182**, 1217–1230, <https://doi.org/10.1083/jcb.200712086> (2008).
- Daimon, T. *et al.* Prognostic significance of erythrocyte protein band 4.1-like5 expression in upper urinary tract urothelial carcinoma. *Urol. Oncol.* **35**, 543.e517–543.e524, <https://doi.org/10.1016/j.urolonc.2017.04.008> (2017).
- Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21, <https://doi.org/10.1093/bioinformatics/bts635> (2013).
- Milovanovic, M. *et al.* IL-33/ST2 axis in inflammation and immunopathology. *Immunol. Res.* **52**, 89–99, <https://doi.org/10.1007/s12026-012-8283-9> (2012).
- Pichery, M. *et al.* Endogenous IL-33 is highly expressed in mouse epithelial barrier tissues, lymphoid organs, brain, embryos, and inflamed tissues: *in situ* analysis using a novel IL-33-LacZ gene trap reporter strain. *J. Immunol.* **188**, 3488–3495, <https://doi.org/10.4049/jimmunol.1101977> (2012).
- Cayrol, C. & Girard, J. P. IL-33: an alarmin cytokine with crucial roles in innate immunity, inflammation and allergy. *Curr. Opin. Immunol.* **31**, 31–37, <https://doi.org/10.1016/j.coi.2014.09.004> (2014).
- Chen, W. Y., Li, L. C. & Yang, J. L. Emerging roles of IL-33/ST2 axis in renal diseases. *Int. J. Mol. Sci.* **18**, <https://doi.org/10.3390/ijms18040783> (2017).
- Greka, A. & Mundel, P. Cell biology and pathology of podocytes. *Annu. Rev. Physiol.* **74**, 299–323, <https://doi.org/10.1146/annurev-physiol-020911-153238> (2012).
- Liu, Y. New insights into epithelial-mesenchymal transition in kidney fibrosis. *J. Am. Soc. Nephrol.* **21**, 212–222, <https://doi.org/10.1681/ASN.2008121226> (2010).
- Chanan-Khan, A. *et al.* Bortezomib in combination with pegylated liposomal doxorubicin and thalidomide is an effective steroid independent salvage regimen for patients with relapsed or refractory multiple myeloma: results of a phase II clinical trial. *Leuk. Lymphoma* **50**, 1096–1101, <https://doi.org/10.1080/10428190902912460> (2009).
- Bottinger, E. P. & Bitzer, M. TGF-beta signaling in renal disease. *J. Am. Soc. Nephrol.* **13**, 2600–2610 (2002).

39. Valcourt, U., Kowanzetz, M., Niimi, H., Heldin, C. H. & Moustakas, A. TGF-beta and the Smad signaling pathway support transcriptomic reprogramming during epithelial-mesenchymal cell transition. *Mol. Biol. Cell* **16**, 1987–2002, <https://doi.org/10.1091/mbc.E04-08-0658> (2005).
40. Smith, C. J. & Osborn, A. M. Advantages and limitations of quantitative PCR (Q-PCR)-based approaches in microbial ecology. *FEMS Microbiol. Ecol.* **67**, 6–20, <https://doi.org/10.1111/j.1574-6941.2008.00629.x> (2009).
41. McIntyre, L. M. *et al.* RNA-seq: technical variability and sampling. *BMC Genomics* **12**, 293, <https://doi.org/10.1186/1471-2164-12-293> (2011).
42. Hansen, K. D., Brenner, S. E. & Dudoit, S. Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Res.* **38**, e131, <https://doi.org/10.1093/nar/gkq224> (2010).
43. Nurtudinov, R. N., Neverov, A. D., Favorov, A. V. & Mironov, A. A., Gelfand, M. S. Conserved and species-specific alternative splicing in mammalian genomes. *BMC Evol Biol.* **7**, 249, <https://doi.org/10.1186/1471-2148-7-249> (2007).

Acknowledgements

We thank Margaret Biswas, PhD, from Edanz Group (www.edanzediting.com/ac) for editing a draft of this manuscript.

Author Contributions

C.Z. and S.Z. conceived the study, designed and implemented the computational workflow, performed the isoform- and exon-level analyses, and drafted the manuscript. K.D. performed the qRT-PCR experiments. K.D., B.Z., R.M. and L.L. participated in writing the manuscript. All authors approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-26035-x>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018