Supporting Information

Shen et al. 10.1073/pnas.1419161111

SI Materials and Methods

Notations and Model. RNA-Seq reads are mapped to both the genome and splice junctions. For a skipped exon, the exon inclusion level (denoted as percent spliced in or ψ) is calculated by the number of reads uniquely mapped to the exon inclusion isoform or the exon skipping isoform. The number of reads mapped to the exon inclusion isoform is denoted by I. The number of reads mapped to the exon skipping isoform is denoted by S. The total number of reads mapped to the exon inclusion or exon skipping isoform is denoted by n, as $n = I + S$. In Fig. S1, the reads of the exon inclusion isoform and the exon skipping isoform are illustrated for a skipped exon. Similarly, we can count the isoform-specific reads corresponding to other types of alternative splicing events (Fig. S1). For the rest of *SI Materials and Methods*, we use skipped exon events to illustrate the rMATS model, although the same statistical framework can be applied to any other type of alternative splicing event. Our rMATS software also provides the option of using only the splice junction reads in the alternative splicing analysis or both the splice junction reads and exon body reads.

As the lengths of isoform-specific segments may differ between alternative isoforms (e.g., for exon inclusion vs. exon skipping isoforms), we need to normalize the isoform-specific read counts by the effective lengths of isoform-specific segments in the calculation of the exon inclusion levels. For a segment whose length is l, and the length of the read is r, the effective length of the segment is defined by the number of unique read intervals in this region, which is $l - r + 1$. Fig. S1 illustrates the calculation of effective lengths for different types of alternative splicing events.

For a skipped exon, we denote the effective length of the exon inclusion isoform as l_I and the effective length of the exon skipping isoform as l_S. Adjusted by the effective lengths of the isoform-specific segments, the exon inclusion level ψ can be estimated by $\hat{\psi} = (I/l_I)/(I/l_I + S/l_S)$. The proportion of reads from the exon inclusion isoform should be $p = l_I\psi/(l_I\psi + l_S(1-\psi))$. Assuming the reads from the exon inclusion isoform follow a binomial distribution, the total count of reads $n = I + S$, and the proportion of reads from the exon inclusion isoform is $p = l_I\psi/(l_I\psi + l_S(1-\psi))$; then

$$I|\psi \sim \text{Binomial}\left(n = I + S,\ p = \frac{l_I\psi}{l_I\psi + l_S(1-\psi)}\right).$$

This binomial model defines the relationship among the exon inclusion reads, the exon skipping reads, and the exon inclusion level in each individual sample, adjusted by the effective length of the exon inclusion or exon skipping isoform.

Statistical Model of rMATS for Unpaired Replicate Analysis.

Notations. When we compare exon inclusion levels between two sample groups, where the replicates are unpaired, multiple replicates within each sample group may have different exon inclusion levels, due to biological variation among replicates and/or technical variation in the RNA-Seq experiments. We can model the variation in the exon inclusion levels among replicates with a model where the logit of the individual exon inclusion levels within each sample group follows a normal distribution. Below we describe the notations and statistical models for the unpaired replicate analysis. Assuming we have a total of N alternatively spliced exons, for each exon $i = 1, \ldots, N$, there are M_1 replicates in sample group 1 and M_2 replicates in sample group 2, we denote $\psi_{i11}, \ldots, \psi_{i1k}, \ldots, \psi_{i1M_1}$, exon inclusion levels of exon i in sample group 1; $\psi_{i21}, \ldots, \psi_{i2k}, \ldots, \psi_{i2M_2}$, exon inclusion levels of exon i in sample group 2; ψ_{i1}, ψ_{i2}, mean of the exon inclusion levels of exon i in sample groups 1 and 2; $\sigma^2_{i1}, \sigma^2_{i2}$, variance of the exon inclusion levels of exon i in sample groups 1 and 2; $I_{i11}, \ldots, \mathrm{I}_{i1k}, \ldots, \mathrm{I}_{i1M_1}$, read counts of the exon inclusion isoform of exon i in sample group 1; $I_{i21}, \ldots, \mathrm{I}_{i2k}, \ldots, \mathrm{I}_{i2M_2}$, read counts of the exon inclusion isoform of exon i in sample group 2; $S_{i11}, \ldots, \mathrm{S}_{i1k}, \ldots, \mathrm{S}_{i1M_1}$, read counts of the exon skipping isoform of exon i in sample group 1; $S_{i21}, \ldots, \mathrm{S}_{i2k}, \ldots, \mathrm{S}_{i2M_2}$, read counts of the exon skipping isoform of exon i in sample group 2; and l_{iI}, l_{iS}, effective lengths of the exon inclusion and exon skipping isoforms of exon i.

Statistical model.

$$\text{logit}(\psi_{i11}), \ldots, \text{logit}(\psi_{i1k}), \ldots, \text{logit}(\psi_{i1M_1}) \sim \text{Normal}\left(\mu = \text{logit}(\psi_{i1}), \sigma^2 = \sigma^2_{i1}\right),$$

$$\text{logit}(\psi_{i21}), \ldots, \text{logit}(\psi_{i2k}), \ldots, \text{logit}(\psi_{i2M_2}) \sim \text{Normal}\left(\mu = \text{logit}(\psi_{i2}), \sigma^2 = \sigma^2_{i2}\right),$$

$$I_{i11}|\psi_{i11} \sim \text{Binomial}\left(n_{i11} = I_{i11} + S_{i11}, p_{i11} = \frac{l_{iI}\psi_{i11}}{l_{iI}\psi_{i11} + l_{iS}(1-\psi_{i11})}\right),$$

...

$$I_{i1k}|\psi_{i1k} \sim \text{Binomial}\left(n_{i1k} = I_{i1k} + S_{i1k}, p_{i1k} = \frac{l_{iI}\psi_{i1k}}{l_{iI}\psi_{i1k} + l_{iS}(1-\psi_{i1k})}\right),$$

...

$$I_{i1M_1}|\psi_{i1M_1} \sim \text{Binomial}\left(n_{i1M_1} = I_{i1M_1} + S_{i1M_1}, p_{i1M_1} = \frac{l_{iI}\psi_{i1M_1}}{l_{iI}\psi_{i1M_1} + l_{iS}(1-\psi_{i1M_1})}\right),$$

$$I_{i21}|\psi_{i21} \sim \text{Binomial}\left(n_{i21} = I_{i21} + S_{i21}, p_{i21} = \frac{l_{iI}\psi_{i21}}{l_{iI}\psi_{i21} + l_{iS}(1-\psi_{i21})}\right),$$

...

$$I_{i2k}|\psi_{i2k} \sim \text{Binomial}\left(n_{i2k} = I_{i2k} + S_{i2k}, p_{i2k} = \frac{l_{iI}\psi_{i2k}}{l_{iI}\psi_{i2k} + l_{iS}(1-\psi_{i2k})}\right),$$

...

$$I_{i2M_2}|\psi_{i2M_2} \sim \text{Binomial}\left(n_{i2M_2} = I_{i2M_2} + S_{i2M_2}, p_{i2M_2} = \frac{l_{iI}\psi_{i2M_2}}{l_{iI}\psi_{i2M_2} + l_{iS}(1-\psi_{i2M_2})}\right).$$

The numbers of replicates (M_1 and M_2) can differ between the two sample groups, when the replicates in the two groups are unpaired.

Conceptually, if some features are removed, our model is equivalent to the generalized linear mixed model (GLMM) where ψ_{i1}, ψ_{i2} model the fixed effect of mean exon inclusion levels in the two sample groups and σ_{i1}, σ_{i2} model the random effect of exon inclusion level variation among individual replicates. However, there are two key distinctions between the rMATS model and a standard GLMM with a logit link function. First, to allow flexible definition and hypothesis testing of differential alternative splicing patterns, rMATS tests whether the difference in mean exon inclusion levels between the two sample groups exceeds a user-defined cutoff (i.e., $|\psi_{i1} - \psi_{i2}| > c$), instead of testing whether the sample group effect is nonzero (i.e., $|\psi_{i1} - \psi_{i2}| > 0$). Second, the length normalization ($p = l_I\psi/(l_I\psi + l_S(1-\psi))$) in the binomial distribution leads to a noncanonical link function. Because of these issues, we need to modify the standard GLMM Laplace approximations to fit our model. The modifications are described below.

Likelihood function. Before we describe the model fitting, we first describe the full-likelihood function of our model. The joint-likelihood function of our model is a combination of (*i*) the normal distribution modeling the variation of the replicate exon inclusion levels within sample group and (*ii*) the binomial distribution modeling the relationship of the exon inclusion reads, exon skipping reads, and the exon inclusion level in each individual replicate. Thus, the joint-likelihood function is composed of two components:

For each exon $i = 1, \ldots, N$,

$$\begin{aligned} L &= L_1 L_2 \\ L_1 &= \prod_{k=1}^{M_1} P(I_{i1k}|\psi_{i1k}, n_{i1k}) \prod_{k=1}^{M_2} P(I_{i2k}|\psi_{i2k}, n_{i2k}) \\ L_2 &= \prod_{k=1}^{M_1} P(\psi_{i1k}|\psi_{i1}, \sigma_{i1}) \prod_{k=1}^{M_2} P(\psi_{i2k}|\psi_{i2}, \sigma_{i2}). \end{aligned} \quad \textbf{[S1]}$$

The L_1 part of Eq. **S1** is from the binomial distribution:

$$\begin{aligned} \prod_{k=1}^{M_1} P(I_{i1k}|\psi_{i1k}, n_{i1k}) &= \prod_{k=1}^{M_1} \binom{I_{i1k}+S_{i1k}}{I_{i1k}} \\ &\times \exp\left(\sum_{k=1}^{M_1} I_{i1k} \log\left(\frac{l_{iI}\psi_{i1k}}{l_{iI}\psi_{i1k} + l_{iS}(1-\psi_{i1k})} \right) \right. \\ &\left. + S_{i1k} \log\left(\frac{l_{iS}(1-\psi_{i1k})}{l_{iI}\psi_{i1k} + l_{iS}(1-\psi_{i1k})} \right) \right), \\ \prod_{k=1}^{M_2} P(I_{i2k}|\psi_{i2k}, n_{i2k}) &= \prod_{k=1}^{M_2} \binom{I_{i2k}+S_{i2k}}{I_{i2k}} \\ &\times \exp\left(\sum_{k=1}^{M_2} I_{i2k} \log\left(\frac{l_{iI}\psi_{i2k}}{l_{iI}\psi_{i2k} + l_{iS}(1-\psi_{i2k})} \right) \right. \\ &\left. + S_{i2k} \log\left(\frac{l_{iS}(1-\psi_{i2k})}{l_{iI}\psi_{i2k} + l_{iS}(1-\psi_{i2k})} \right) \right). \end{aligned} \quad \textbf{[S2]}$$

The L_2 part of Eq. **S1** is from the normal distribution:

$$\begin{aligned} \prod_{k=1}^{M_1} P(\psi_{i1k}|\psi_{i1}, \sigma_{i1}) &= \exp\left(\sum_{k=1}^{M_1} -0.5\log(2\pi) - \log\sigma_{i1} \right. \\ &\left. - \frac{(\text{logit}(\psi_{i1k}) - \text{logit}(\psi_{i1}))^2}{2\sigma_{i1}^2} + \log(\psi_{i1k}(1-\psi_{i1k})) \right), \\ \prod_{k=1}^{M_2} P(\psi_{i2k}|\psi_{i2}, \sigma_{i2}) &= \exp\left(\sum_{k=1}^{M_2} -0.5\log(2\pi) - \log\sigma_{i2} \right. \\ &\left. - \frac{(\text{logit}(\psi_{i2k}) - \text{logit}(\psi_{i2}))^2}{2\sigma_{i2}^2} + \log(\psi_{i2k}(1-\psi_{i2k})) \right). \end{aligned} \quad \textbf{[S3]}$$

In Eq. **S3**, $\log(\psi_{i1k}(1-\psi_{i1k}))$ and $\log(\psi_{i2k}(1-\psi_{i2k}))$ are introduced by the logit transformation of the exon inclusion levels.

The joint-likelihood function (Eq. **S1**) is a combination of the two components in Eqs. **S2** and **S3**.

Laplace approximation of the marginal distribution of the mean and variance of exon inclusion levels. Because our goal is to test the difference of mean exon inclusion levels of exon *i* between two sample groups (i.e., $\psi_{i1} - \psi_{i2}$), we treat the individual exon inclusion levels ($\psi_{i11}, \ldots, \psi_{i1k}, \ldots, \psi_{i1M_1}$ and $\psi_{i21}, \ldots, \psi_{i2k}, \ldots, \psi_{i2M_2}$) as latent variables and derive the marginal distribution of the mean and variance of exon inclusion levels:

$$\begin{aligned} & f(\psi_{i1}, \sigma_{i1}, \psi_{i2}, \sigma_{i2}) \\ &= c \int f(\psi_{i1}, \sigma_{i1}, \psi_{i2}, \sigma_{i2}, \psi_{i11}, \ldots, \psi_{i1M_1}, \psi_{i21}, \ldots, \psi_{i2M_2}) \\ &\quad \times d\psi_{i11} \ldots d\psi_{i1M_1} d\psi_{i21} \ldots d\psi_{i2M_2} \\ &= c\left(\prod_{k=1}^{M_1} \int f(\psi_{i1}, \sigma_{i1}, \psi_{i1k}) d\psi_{i1k} \prod_{k=1}^{M_2} \int f(\psi_{i2}, \sigma_{i2}, \psi_{i2k}) d\psi_{i2k} \right). \end{aligned} \quad \textbf{[S4]}$$

In Eq. **S4**, c is a constant that is not changed by parameters; $f(\psi_{i1}, \sigma_{i1}, \psi_{i1k})$ and $f(\psi_{i2}, \sigma_{i2}, \psi_{i2k})$ are defined by the combination of normal and binomial distributions in Eqs. **S2** and **S3**:

$$\begin{aligned} & f(\psi_{i1}, \sigma_{i1}, \psi_{i1k}) \\ &= \exp\left(-\log\sigma_{i1} - \frac{0.5(\text{logit}(\psi_{i1k}) - \text{logit}(\psi_{i1}))^2}{\sigma_{i1}^2} \right. \\ &\quad + \log(\psi_{i1k}) + \log(1-\psi_{i1k}) + I_{i1k}\log\left(\frac{l_{iI}\psi_{i1k}}{l_{iI}\psi_{i1k} + l_{iS}(1-\psi_{i1k})} \right) \\ &\quad \left. + S_{i1k}\log\left(\frac{l_{iS}(1-\psi_{i1k})}{l_{iI}\psi_{i1k} + l_{iS}(1-\psi_{i1k})} \right) \right), \\ & f(\psi_{i2}, \sigma_{i2}, \psi_{i2k}) \\ &= \exp\left(-\log\sigma_{i2} - \frac{0.5(\text{logit}(\psi_{i2k}) - \text{logit}(\psi_{i2}))^2}{\sigma_{i2}^2} \right. \\ &\quad + \log(\psi_{i2k}) + \log(1-\psi_{i2k}) + I_{i2k}\log\left(\frac{l_{iI}\psi_{i2k}}{l_{iI}\psi_{i2k} + l_{iS}(1-\psi_{i2k})} \right) \\ &\quad \left. + S_{i2k}\log\left(\frac{l_{iS}(1-\psi_{i2k})}{l_{iI}\psi_{i2k} + l_{iS}(1-\psi_{i2k})} \right) \right). \end{aligned} \quad \textbf{[S5]}$$

Because of the lack of closed-form expressions, we use Laplace's method to approximate the integrals of Eq. **S5**:

$$\begin{aligned}
&\int f(\psi_{i1},\sigma_{i1},\psi_{i1k})d\psi_{i1k}\\
&\overset{f_1=\log f}{=} \int \exp(f_1(\psi_{i1},\sigma_{i1},\psi_{i1k}))d\psi_{i1k}\\
&= \int \exp\Bigg(f_1(\psi_{i1},\sigma_{i1},\hat{\psi}_{i1k}) + 0.5\frac{\partial^2 f_1(\psi_{i1},\sigma_{i1},\hat{\psi}_{i1k})}{\partial\psi_{i1k}^2}(\psi_{i1k}-\hat{\psi}_{i1k})^2\\
&\qquad + o\Big((\psi_{i1k}-\hat{\psi}_{i1k})^2\Big)\Bigg)d\psi_{i1k}\\
&\approx \sqrt{2\pi}\left(\left|\frac{\partial^2 f_1(\psi_{i1},\sigma_{i1},\hat{\psi}_{i1k})}{\partial\psi_{i1k}^2}\right|\right)^{-0.5}\exp(f_1(\psi_{i1},\sigma_{i1},\hat{\psi}_{i1k})),\\
&\int f(\psi_{i2},\sigma_{i2},\psi_{i2k})d\psi_{i2k}\\
&\overset{f_1=\log f}{\approx} \sqrt{2\pi}\left(\left|\frac{\partial^2 f_1(\psi_{i2},\sigma_{i2},\hat{\psi}_{i2k})}{\partial\psi_{i2k}^2}\right|\right)^{-0.5}\exp(f_1(\psi_{i2},\sigma_{i2},\hat{\psi}_{i2k})).
\end{aligned}\tag{S6}$$

In Eq. **S6**, Laplace's method approximates the distribution of ψ_{i1k} and ψ_{i2k} by a normal distribution, using the second-level derivative function in the Taylor series. The first-level derivative function of the Taylor series is equal to zero because $\hat{\psi}_{i1k}$ and $\hat{\psi}_{i2k}$ are the maximum-likelihood estimates based on the full-likelihood functions of Eq. **S1**, with fixed values of $\psi_{i1},\sigma_{i1},\psi_{i2},\sigma_{i2}$. The fixed values (denoted as $\hat{\psi}_{i1},\hat{\sigma}_{i1},\hat{\psi}_{i2},\hat{\sigma}_{i2}$ below) are estimated in *Optimization procedure for the MLE*:

$$\begin{aligned}
\hat{\psi}_{i1k} = \underset{\psi_{i1k}}{\arg\max}\Bigg(&\frac{-0.5(\mathrm{logit}(\psi_{i1k})-\mathrm{logit}(\hat{\psi}_{i1}))^2}{\hat{\sigma}_{i1}^2}+\log(\psi_{i1k})\\
&+\log(1-\psi_{i1k})+\ I_{i1k}\log\left(\frac{l_{iI}\psi_{i1k}}{l_{iI}\psi_{i1k}+l_{iS}(1-\psi_{i1k})}\right)\\
&+S_{i1k}\ \log\left(\frac{l_{iS}(1-\psi_{i1k})}{l_{iI}\psi_{i1k}+l_{iS}(1-\psi_{i1k})}\right)\Bigg),\\
\hat{\psi}_{i2k} = \underset{\psi_{i2k}}{\arg\max}\Bigg(&\frac{-0.5(\mathrm{logit}(\psi_{i2k})-\mathrm{logit}(\hat{\psi}_{i2}))^2}{\hat{\sigma}_{i2}^2}\\
&+\log(\psi_{i2k})+\log(1-\psi_{i2k})\\
&+I_{i2k}\log\left(\frac{l_{iI}\psi_{i2k}}{l_{iI}\psi_{i2k}+l_{iS}(1-\psi_{i2k})}\right)\\
&+S_{i2k}\ \log\left(\frac{l_{iS}(1-\psi_{i2k})}{l_{iI}\psi_{i2k}+l_{iS}(1-\psi_{i2k})}\right)\Bigg).
\end{aligned}\tag{S7}$$

In Eq. **S6**, the second-level derivative function is

$$\begin{aligned}
&\frac{\partial^2 f_1(\psi_{i1},\sigma_{i1},\hat{\psi}_{i1k})}{\partial\psi_{i1k}^2}\\
&=\frac{2\hat{\psi}_{i1k}-1}{\hat{\psi}_{i1k}^2(1-\hat{\psi}_{i1k})^2}\left(\frac{\mathrm{logit}\ \psi_{i1}-\mathrm{logit}\ \hat{\psi}_{i1k}-(2\hat{\psi}_{i1k}-1)^{-1}}{\sigma_{i1}^2}+1\right)\\
&\quad-\ I_{i1k}l_{iS}\frac{(2l_{iI}+l_{iS})\hat{\psi}_{i1k}+l_{iS}(1-\hat{\psi}_{i1k})}{\hat{\psi}_{i1k}^2(l_{iI}\hat{\psi}_{i1k}+l_{iS}(1-\hat{\psi}_{i1k}))^2}\\
&\quad-S_{i1k}l_{iI}\frac{(l_{iI}+2l_{iS})(1-\hat{\psi}_{i1k})+l_{iI}\hat{\psi}_{i1k}}{(1-\hat{\psi}_{i1k})^2(l_{iI}\hat{\psi}_{i1k}+l_{iS}(1-\hat{\psi}_{i1k}))^2}.
\end{aligned}\tag{S8}$$

Similarly,

$$\begin{aligned}
&\int f(\psi_{i2},\sigma_{i2},\psi_{i2k})d\psi_{i2k}\\
&\qquad\approx\sqrt{2\pi}\left(\left|\frac{\partial^2 f_1(\psi_{i2},\sigma_{i2},\hat{\psi}_{i2k})}{\partial\psi_{i2k}^2}\right|\right)^{-0.5}\exp(f_1(\psi_{i2},\sigma_{i2},\hat{\psi}_{i2k})),\\
&\frac{\partial^2 f_1(\psi_{i2},\sigma_{i2},\hat{\psi}_{i2k})}{\partial\psi_{i2k}^2}\\
&=\frac{2\hat{\psi}_{i2k}-1}{\hat{\psi}_{i2k}^2(1-\hat{\psi}_{i2k})^2}\left(\frac{\mathrm{logit}\ \psi_{i2}-\mathrm{logit}\ \hat{\psi}_{i2k}-(2\hat{\psi}_{i2k}-1)^{-1}}{\sigma_{i2}^2}+1\right)\\
&\quad-\ I_{i2k}l_{iS}\frac{(2l_{iI}+l_{iS})\hat{\psi}_{i2k}+l_{iS}(1-\hat{\psi}_{i2k})}{\hat{\psi}_{i2k}^2(l_{iI}\hat{\psi}_{i2k}+l_{iS}(1-\hat{\psi}_{i2k}))^2}\\
&\quad-S_{i2k}l_{iI}\frac{(l_{iI}+2l_{iS})(1-\hat{\psi}_{i2k})+l_{iI}\hat{\psi}_{i2k}}{(1-\hat{\psi}_{i2k})^2(l_{iI}\hat{\psi}_{i2k}+l_{iS}(1-\hat{\psi}_{i2k}))^2}.
\end{aligned}\tag{S9}$$

Likelihood-ratio test of splicing difference. In the previous section, we describe the approximation for the marginal distribution of the mean and variance of exon inclusion levels. Based on the marginal distribution, we can calculate the P value of splicing difference for each exon by the likelihood-ratio test. Recall that rMATS tests whether the difference in mean exon inclusion levels between the two sample groups exceeds a user-defined cutoff (i.e., $|\psi_{i1}-\psi_{i2}|\ >c$). For each exon i, the null hypothesis is that the difference of the mean exon inclusion levels is smaller than or equal to the user-defined cutoff c (i.e., $|\Delta\psi| = |\psi_{i1}-\psi_{i2}|\le c$), whereas the alternative hypothesis is $|\psi_{i1}-\psi_{i2}|>c$.

If the maximum-likelihood estimations (MLEs) of ψ_{i1}, ψ_{i2} have a difference smaller than or equal to the user-defined cutoff (i.e., $|\psi_{i1}-\psi_{i2}|\le c$), we set the P value to be 1. Otherwise, we compare the likelihood under the constraint of the null hypothesis and the likelihood from the unconstrained MLE. The constraint of the null hypothesis leads to a likelihood ratio whose probability distribution does not have a closed-form expression. However, note that when the MLEs of ψ_{i1}, ψ_{i2} have a difference greater than the user-defined cutoff c, the MLEs of ψ_{i1}, ψ_{i2} under the constraint of $|\psi_{i1}-\psi_{i2}|\le c$ always fall on the boundary of $|\psi_{i1}-\psi_{i2}|=c$. We can instead calculate a more conservative P value by comparing the null hypothesis on the boundary $H_0:|\psi_{i1}-\psi_{i2}|=c$, vs. the alternative hypothesis $H_1:|\psi_{i1}-\psi_{i2}|>c$. With such a null hypothesis, the likelihood-ratio test statistic asymptotically follows a χ^2 distribution with 1 df,

$$-2\big(\log L_{|\psi_{i1}-\psi_{i2}|=c}-\log L\big)\sim\chi_1^2,\tag{S10}$$

in which $\log L_{|\psi_{i1}-\psi_{i2}|=c}$ is the log likelihood under the constraint that $|\psi_{i1}-\psi_{i2}|\ =c$ and log L is the log likelihood from the unconstrained MLE. We calculate the MLEs of ψ_{i1}, ψ_{i2} and σ_{i1}, σ_{i2} based on the marginal distribution of Eqs. **S6**, **S8**, and **S9**. The optimization procedure to calculate the MLE is described in the next section.

Optimization procedure for the MLE. In this section, we describe the optimization procedure to calculate the MLE of the mean exon inclusion levels ψ_{i1}, ψ_{i2} and the variance σ_{i1}, σ_{i2}, based on the marginal distribution of ψ_{i1}, ψ_{i2} and σ_{i1}, σ_{i2}.

In the marginal distribution (Eq. **S4**), the function has a closed form if all of the latent variables ψ_{i1k} and ψ_{i2k} are fixed. However, the estimated values of the latent variables $\hat{\psi}_{i1k}$ and $\hat{\psi}_{i2k}$ are the MLEs of the full-likelihood function with fixed values of mean inclusion levels ψ_{i1}, ψ_{i2} and variance σ_{i1}, σ_{i2}. Therefore, in the Laplace approximation, we use an iterative optimization procedure for the MLE calculation.

The initial estimated values of the latent variables ψ_{i1k} and ψ_{i2k} are derived from the individual binomial distributions of each replicate (Eq. **S2**):

$$\hat{\psi}_{i1k}^{(1)} = \frac{I_{i1k} l_{iS}}{I_{i1k} l_{iS} + S_{i1k} l_{iI}}$$

and

$$\hat{\psi}_{i2k}^{(1)} = \frac{I_{i2k} l_{iS}}{I_{i2k} l_{iS} + S_{i2k} l_{iI}}.$$

In each round (t) of the iterative optimization procedure, we first estimate the MLE of the marginal distribution (Eq. **S4**), based on the estimated values of the latent variables $\hat{\psi}_{i1k}^{(t)}$ and $\hat{\psi}_{i2k}^{(t)}$ and the Laplace approximation of the integrals of the full likelihood (Eqs. **S6**, **S8**, and **S9**):

$$\begin{aligned}
&\left(\hat{\psi}_{i1}^{(t)}, \hat{\psi}_{i2}^{(t)}, \hat{\sigma}_{i1}^{(t)}, \hat{\sigma}_{i2}^{(t)}\right) \\
&= \underset{\psi_{i1},\psi_{i2},\sigma_{i1},\sigma_{i2}}{\arg\max} \left(\sum_{k=1}^{M_1} \left(f_1\left(\psi_{i1}, \sigma_{i1}, \hat{\psi}_{i1k}^{(t)}\right) - 0.5 \log \left| \frac{\partial^2 f_1\left(\psi_{i1}, \sigma_{i1}, \hat{\psi}_{i1k}^{(t)}\right)}{\partial \psi_{i1k}^2} \right| \right) \right. \\
&\left. + \sum_{k=1}^{M_2} \left(f_1\left(\psi_{i2}, \sigma_{i2}, \hat{\psi}_{i2k}^{(t)}\right) - 0.5 \log \left| \frac{\partial^2 f_1\left(\psi_{i2}, \sigma_{i2}, \hat{\psi}_{i2k}^{(t)}\right)}{\partial \psi_{i2k}^2} \right| \right) \right).
\end{aligned} \qquad \textbf{[S11]}$$

In [**S11**], the function f_1 is the log of the function f in Eq. **S5**. The second-level derivative function is described in Eqs. **S8** and **S9**.

The next step of the iterative optimization procedure updates the estimation of the latent variables $\hat{\psi}_{i1k}^{(t+1)}$ and $\hat{\psi}_{i2k}^{(t+1)}$ based on the full likelihood (Eq. **S1**) and the latest MLE of $\hat{\psi}_{i1}^{(t)}$, $\hat{\psi}_{i2}^{(t)}$, $\hat{\sigma}_{i1}^{(t)}$, $\hat{\sigma}_{i2}^{(t)}$. Given the mean and variance of exon inclusion levels $\hat{\psi}_{i1}^{(t)}$, $\hat{\psi}_{i2}^{(t)}$, $\hat{\sigma}_{i1}^{(t)}$, $\hat{\sigma}_{i2}^{(t)}$, the exon inclusion level $\hat{\psi}_{i1k}^{(t+1)}$ or $\hat{\psi}_{i2k}^{(t+1)}$ can be estimated separately for each individual sample. As described in Eq. **S7**, for each replicate $k = 1 \ldots M_1$ in the sample group 1,

$$\begin{aligned}
\hat{\psi}_{i1k}^{(t+1)} = \underset{\psi_{i1k}}{\arg\max} \Bigg(& \frac{-0.5\left(\operatorname{logit}(\psi_{i1k}) - \operatorname{logit}\left(\hat{\psi}_{i1}^{(t)}\right)\right)^2}{\left(\hat{\sigma}_{i1}^{(t)}\right)^2} \\
& + \log(\psi_{i1k}) + \log(1 - \psi_{i1k}) \\
& + I_{i1k} \log\left(\frac{l_{iI}\psi_{i1k}}{l_{iI}\psi_{i1k} + l_{iS}(1 - \psi_{i1k})}\right) \\
& + S_{i1k} \log\left(\frac{l_{iS}(1 - \psi_{i1k})}{l_{iI}\psi_{i1k} + l_{iS}(1 - \psi_{i1k})}\right) \Bigg).
\end{aligned}$$

And for each replicate $k = 1 \ldots M_2$ in the sample group 2,

$$\begin{aligned}
\hat{\psi}_{i2k}^{(t+1)} = \underset{\psi_{i2k}}{\arg\max} \Bigg(& \frac{-0.5\left(\operatorname{logit}(\psi_{i2k}) - \operatorname{logit}\left(\hat{\psi}_{i2}^{(t)}\right)\right)^2}{\left(\hat{\sigma}_{i2}^{(t)}\right)^2} \\
& + \log(\psi_{i2k}) + \log(1 - \psi_{i2k}) \\
& + I_{i2k} \log\left(\frac{l_{iI}\psi_{i2k}}{l_{iI}\psi_{i2k} + l_{iS}(1 - \psi_{i2k})}\right) \\
& + S_{i2k} \log\left(\frac{l_{iS}(1 - \psi_{i2k})}{l_{iI}\psi_{i2k} + l_{iS}(1 - \psi_{i2k})}\right) \Bigg).
\end{aligned}$$

This optimization procedure iterates for multiple rounds until the difference in log likelihood between two consecutive iterations is smaller than 10^{-4}. On average, the iterative procedure takes 6.4 iterations to converge on the RNA-Seq data from the prostate cancer cell lines described in this article. The maximum number of iterations is 37.

The optimization procedure uses the L-BFGS-B algorithm (1) to optimize the likelihood function with the parameter ψ constrained within 0–1 and the parameter σ within 0 to infinite.

The constrained MLE under the null hypothesis is estimated with the same procedure, except for an additional constraint $|\psi_{i1} - \psi_{i2}| = c$. Specifically, we replace the parameter ψ_{i2} with either $\psi_{i1} - c$ or $\psi_{i1} + c$ and select the best MLE under the two scenarios.

Statistical Model of rMATS for Paired Replicate Analysis. In certain studies, replicates are paired between sample groups. One example is the comparison of matched cancer-normal tissue pairs across multiple cancer patients. We have extended the rMATS model to handle paired replicates. Although the notations for the paired model and the unpaired model are almost identical, in the paired model we use a bivariate normal distribution with the correlation parameter ρ_i to model the correlation within matched pairs for exon i.

Statistical model. Assuming we have a total of N alternatively spliced exons, and for each exon $i = 1, \ldots, N$, there are M matched replicates in sample groups 1 and 2, we have

$$\begin{bmatrix} \operatorname{logit}(\psi_{i11}) \\ \operatorname{logit}(\psi_{i21}) \end{bmatrix}, \ldots, \begin{bmatrix} \operatorname{logit}(\psi_{i1k}) \\ \operatorname{logit}(\psi_{i2k}) \end{bmatrix}, \ldots, \begin{bmatrix} \operatorname{logit}(\psi_{i1M}) \\ \operatorname{logit}(\psi_{i2M}) \end{bmatrix}$$

$$\sim \text{Normal}\left(\mu = \begin{bmatrix} \operatorname{logit}(\psi_{i1}) \\ \operatorname{logit}(\psi_{i2}) \end{bmatrix}, \ \Sigma_i = \begin{bmatrix} \sigma_{i1}^2 & \rho_i\sigma_{i1}\sigma_{i2} \\ \rho_i\sigma_{i1}\sigma_{i2} & \sigma_{i2}^2 \end{bmatrix} \right),$$

$$I_{i11}|\psi_{i11} \sim \text{Binomial}\left(n_{i11} = I_{i11} + S_{i11}, p_{i11} = \frac{l_{iI}\psi_{i11}}{l_{iI}\psi_{i11} + l_{iS}(1 - \psi_{i11})} \right),$$

...

$$I_{i1k}|\psi_{i1k} \sim \text{Binomial}\left(n_{i1k} = I_{i1k} + S_{i1k}, p_{i1k} = \frac{l_{iI}\psi_{i1k}}{l_{iI}\psi_{i1k} + l_{iS}(1 - \psi_{i1k})} \right),$$

...

$$\begin{aligned} I_{i1M}|\psi_{i1M} \sim \text{Binomial}\Bigg(& n_{i1M} = I_{i1M} + S_{i1M}, p_{i1M} \\ & = \frac{l_{iI}\psi_{i1M}}{l_{iI}\psi_{i1M} + l_{iS}(1 - \psi_{i1M})} \Bigg), \end{aligned}$$

$$I_{i21}|\psi_{i21} \sim \text{Binomial}\left(n_{i21} = I_{i21} + S_{i21}, p_{i21} = \frac{l_{iI}\psi_{i21}}{l_{iI}\psi_{i21} + l_{iS}(1 - \psi_{i21})} \right),$$

...

$$I_{i2k}|\psi_{i2k} \sim \text{Binomial}\left(n_{i2k} = I_{i2k} + S_{i2k}, p_{i2k} = \frac{l_{iI}\psi_{i2k}}{l_{iI}\psi_{i2k} + l_{iS}(1 - \psi_{i2k})} \right),$$

...

$$\begin{aligned} I_{i2M}|\psi_{i2M} \sim \text{Binomial}\Bigg(& n_{i2M} = I_{i2M} + S_{i2M}, p_{i2M} \\ & = \frac{l_{iI}\psi_{i2M}}{l_{iI}\psi_{i2M} + l_{iS}(1 - \psi_{i2M})} \Bigg). \end{aligned}$$

Likelihood-ratio test of splicing difference. In the paired analysis, the marginal distribution of mean exon inclusion levels can also be approximated by Laplace's method, by treating the individual exon inclusion levels ($\psi_{i11}, \ldots, \psi_{i1k}, \ldots, \psi_{i1M}$ and $\psi_{i21}, \ldots, \psi_{i2k}, \ldots, \psi_{i2M}$) as latent variables. However, the mean exon inclusion levels ψ_{i1}, ψ_{i2} cannot be separated in the marginal distribution because of the bivariate normal distribution. Therefore, we take the integral of the pair of variables ψ_{i1k}, ψ_{i2k} together to derive the marginal distribution:

$$\begin{aligned}
&f(\psi_{i1},\psi_{i2},\sigma_{i1},\sigma_{i2},\rho_i)\\
&\quad = c\int f(\psi_{i1},\psi_{i2},\sigma_{i1},\sigma_{i2},\rho_i,\psi_{i11}, \ldots, \psi_{i1M},\psi_{i21}, \ldots, \psi_{i2M})\\
&\qquad \times d\psi_{i11}\ldots d\psi_{i1M}d\psi_{i21}\ldots d\psi_{i2M}\\
&\quad = c\prod_{k=1}^{M}\int f(\psi_{i1},\psi_{i2},\sigma_{i1},\sigma_{i2},\rho_i,\psi_{i1k},\psi_{i2k})d\psi_{i1k}d\psi_{i2k}.
\end{aligned} \qquad \textbf{[S12]}$$

$$\begin{aligned}
&\int f(\psi_{i1},\psi_{i2},\sigma_{i1},\sigma_{i2},\rho_i,\psi_{i1k},\psi_{i2k})d\psi_{i1k}d\psi_{i2k}\\
&\overset{f_1=\log f}{=} \int \exp(f_1(\psi_{i1},\psi_{i2},\sigma_{i1},\sigma_{i2},\rho_i,\psi_{i1k},\psi_{i2k}))d\psi_{i1k}d\psi_{i2k}\\
&= \int \exp\Bigg(f_1(\psi_{i1},\psi_{i2},\sigma_{i1},\sigma_{i2},\rho_i,\psi_{i1k},\psi_{i2k})\\
&\qquad + 0.5\begin{bmatrix}\psi_{i1k}-\hat{\psi}_{i1k}\\ \psi_{i2k}-\hat{\psi}_{i2k}\end{bmatrix}'\sum\nolimits_{ik}^{1}\begin{bmatrix}\psi_{i1k}-\hat{\psi}_{i1k}\\ \psi_{i2k}-\hat{\psi}_{i2k}\end{bmatrix}\\
&\qquad + o\left((\psi_{i1k}-\hat{\psi}_{i1k})^2\right) + o\left((\psi_{i2k}-\hat{\psi}_{i2k})^2\right)\Bigg)d\psi_{i1k}d\psi_{i2k}\\
&\approx 2\pi\left(\left|\sum\nolimits_{ik}^{1}\right|\right)^{-0.5}\exp(f_1(\psi_{i1},\psi_{i2},\sigma_{i1},\sigma_{i2},\rho_i,\hat{\psi}_{i1k},\hat{\psi}_{i2k})).
\end{aligned} \qquad \textbf{[S14]}$$

In Eq. **S14**, $\sum_{ik}^{1}$ is the Hessian matrix of the log-likelihood function f_1:

$$\sum\nolimits_{ik}^{1} = \begin{bmatrix}
\dfrac{\partial^2 f_1(\psi_{i1},\psi_{i2},\sigma_{i1},\sigma_{i2},\rho_i,\hat{\psi}_{i1k},\hat{\psi}_{i2k})}{\partial\psi_{i1k}^2} & \dfrac{\partial^2 f_1(\psi_{i1},\psi_{i2},\sigma_{i1},\sigma_{i2},\rho_i,\hat{\psi}_{i1k},\hat{\psi}_{i2k})}{\partial\psi_{i1k}\partial\psi_{i2k}}\\
\dfrac{\partial^2 f_1(\psi_{i1},\psi_{i2},\sigma_{i1},\sigma_{i2},\rho_i,\hat{\psi}_{i1k},\hat{\psi}_{i2k})}{\partial\psi_{i1k}\partial\psi_{i2k}} & \dfrac{\partial^2 f_1(\psi_{i1},\psi_{i2},\sigma_{i1},\sigma_{i2},\rho_i,\hat{\psi}_{i1k},\hat{\psi}_{i2k})}{\partial\psi_{i2k}^2}
\end{bmatrix}.$$

Compared with the marginal distribution of the unpaired analysis (Eq. **S4**), the integral of variables ψ_{i1k}, ψ_{i2k} cannot be separated in the paired marginal distribution.

In Eq. **S12**,

$$\begin{aligned}
&f(\psi_{i1},\psi_{i2},\sigma_{i1},\sigma_{i2},\rho_i,\psi_{i1k},\psi_{i2k})\\
&\quad = \exp\Bigg(-0.5\log\left|\sum\nolimits_i\right| - 0.5\begin{bmatrix}\operatorname{logit}(\psi_{i1k})-\operatorname{logit}(\psi_{i1})\\ \operatorname{logit}(\psi_{i2k})-\operatorname{logit}(\psi_{i2})\end{bmatrix}'\\
&\qquad \times\sum\nolimits_i^{-1}\begin{bmatrix}\operatorname{logit}(\psi_{i1k})-\operatorname{logit}(\psi_{i1})\\ \operatorname{logit}(\psi_{i2k})-\operatorname{logit}(\psi_{i2})\end{bmatrix}\\
&+\log(\psi_{i1k})+\log(1-\psi_{i1k})+I_{i1k}\log\left(\frac{l_{iI}\psi_{i1k}}{l_{iI}\psi_{i1k}+l_{iS}(1-\psi_{i1k})}\right)\\
&+S_{i1k}\log\left(\frac{l_{iS}(1-\psi_{i1k})}{l_{iI}\psi_{i1k}+l_{iS}(1-\psi_{i1k})}\right)\\
&+\log(\psi_{i2k})+\log(1-\psi_{i2k})+I_{i2k}\log\left(\frac{l_{iI}\psi_{i2k}}{l_{iI}\psi_{i2k}+l_{iS}(1-\psi_{i2k})}\right)\\
&+S_{i2k}\log\left(\frac{l_{iS}(1-\psi_{i2k})}{l_{iI}\psi_{i2k}+l_{iS}(1-\psi_{i2k})}\right)\Bigg).
\end{aligned} \qquad \textbf{[S13]}$$

In [**S13**], $\sum_i = \begin{bmatrix}\sigma_{i1}^2 & \rho_i\sigma_{i1}\sigma_{i2}\\ \rho_i\sigma_{i1}\sigma_{i2} & \sigma_{i2}^2\end{bmatrix}$ is the covariance matrix. Because of the integral of two variables ψ_{i1k}, ψ_{i2k} in Eq. **S12**, Laplace's method approximates the joint posterior distribution of ψ_{i1k}, ψ_{i2k} with a bivariate normal in the paired analysis, instead of using the univariate normal distributions as in the unpaired analysis:

The first-level derivative function of the Taylor series is equal to zero because $\hat{\psi}_{i1k}$ and $\hat{\psi}_{i2k}$ are the maximum-likelihood estimates based on the full-likelihood functions of Eq. **S13**, with fixed values of $\psi_{i1},\psi_{i2},\sigma_{i1},\sigma_{i2},\rho_i$:

$$\begin{aligned}
(\hat{\psi}_{i1k},\hat{\psi}_{i2k}) = \underset{\psi_{i1k},\psi_{i2k}}{\arg\max}\Bigg(&-0.5\begin{bmatrix}\operatorname{logit}(\psi_{i1k})-\operatorname{logit}(\hat{\psi}_{i1})\\ \operatorname{logit}(\psi_{i2k})-\operatorname{logit}(\hat{\psi}_{i2})\end{bmatrix}'\\
&\times\hat{\sum}\nolimits_i^{-1}\begin{bmatrix}\operatorname{logit}(\psi_{i1k})-\operatorname{logit}(\hat{\psi}_{i1})\\ \operatorname{logit}(\psi_{i2k})-\operatorname{logit}(\hat{\psi}_{i2})\end{bmatrix}\\
&+\log(\psi_{i1k})+\log(1-\psi_{i1k})+I_{i1k}\log\left(\frac{l_{iI}\psi_{i1k}}{l_{iI}\psi_{i1k}+l_{iS}(1-\psi_{i1k})}\right)\\
&+S_{i1k}\log\left(\frac{l_{iS}(1-\psi_{i1k})}{l_{iI}\psi_{i1k}+l_{iS}(1-\psi_{i1k})}\right)\\
&+\log(\psi_{i2k})+\log(1-\psi_{i2k})+I_{i2k}\log\left(\frac{l_{iI}\psi_{i2k}}{l_{iI}\psi_{i2k}+l_{iS}(1-\psi_{i2k})}\right)\\
&+S_{i2k}\log\left(\frac{l_{iS}(1-\psi_{i2k})}{l_{iI}\psi_{i2k}+l_{iS}(1-\psi_{i2k})}\right)\Bigg).
\end{aligned}$$

The likelihood-ratio test of the paired analysis is based on the same hypotheses as in the unpaired analysis. For each exon i, the null hypothesis is that the difference of the mean exon inclusion levels is smaller than or equal to a user-defined cutoff c (i.e., $|\psi_{i1}-\psi_{i2}|\le c$), whereas the alternative hypothesis is $|\psi_{i1}-\psi_{i2}|>c$. The MLE and likelihood-ratio test statistics are estimated based on the same iterative optimization procedure on the Laplace approximation of the marginal distribution function of the mean exon inclusion levels ψ_{i1},ψ_{i2}, the variance σ_{i1},σ_{i2}, and the correlation parameter ρ_i (Eq. **S14**).

1. Zhu C, Byrd RH, Lu P, Nocedal J (1997) *ACM Trans Math Softw* 23:550–560.

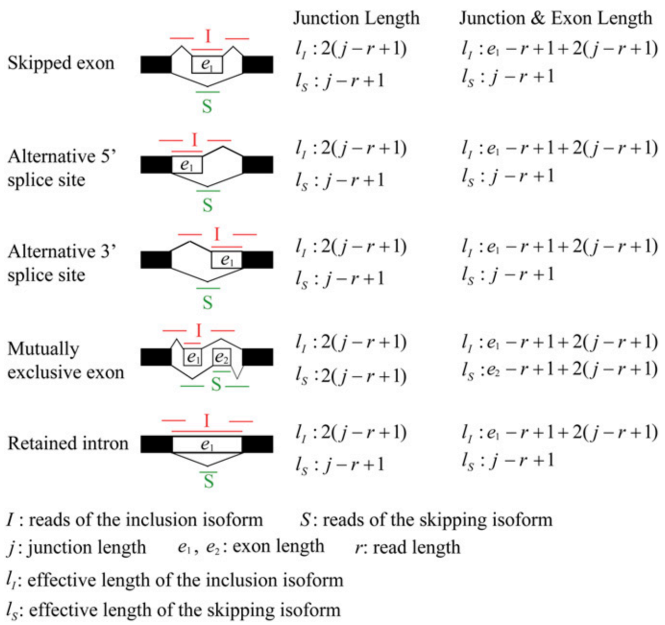
Junction Length
Junction & Exon Length
Skipped exon
I
e_1
S
$l_I : 2(j-r+1)$
$l_S : j-r+1$
$l_I : e_1 - r + 1 + 2(j-r+1)$
$l_S : j-r+1$
Alternative 5’ splice site
I
e_1
S
$l_I : 2(j-r+1)$
$l_S : j-r+1$
$l_I : e_1 - r + 1 + 2(j-r+1)$
$l_S : j-r+1$
Alternative 3’ splice site
I
e_1
S
$l_I : 2(j-r+1)$
$l_S : j-r+1$
$l_I : e_1 - r + 1 + 2(j-r+1)$
$l_S : j-r+1$
Mutually exclusive exon
I
e_1
e_2
S
$l_I : 2(j-r+1)$
$l_S : 2(j-r+1)$
$l_I : e_1 - r + 1 + 2(j-r+1)$
$l_S : e_2 - r + 1 + 2(j-r+1)$
Retained intron
I
e_1
S
$l_I : 2(j-r+1)$
$l_S : j-r+1$
$l_I : e_1 - r + 1 + 2(j-r+1)$
$l_S : j-r+1$
I : reads of the inclusion isoform
S: reads of the skipping isoform
j: junction length
e_1, e_2: exon length
r: read length
l_I: effective length of the inclusion isoform
l_S: effective length of the skipping isoform

Fig. S1. The schematic diagrams illustrating the read counts and effective lengths of different categories of alternative splicing events. The alternative splicing events of skipped exons, alternative 5′ splice sites, alternative 3′ splice sites, and retained introns have two splice junctions for the inclusion isoform and one splice junction for the skipping isoform. The mutually exclusive exons have two splice junctions for the inclusion isoform of the first exon and two splice junctions for the skipping isoform of the first exon (i.e., the inclusion isoform of the second exon). The exon body reads are RNA-Seq reads mapped to the genomic regions of the target exons. The rMATS model allows users to use either the splice junction counts plus the exon body counts or the splice junction counts alone as the input.

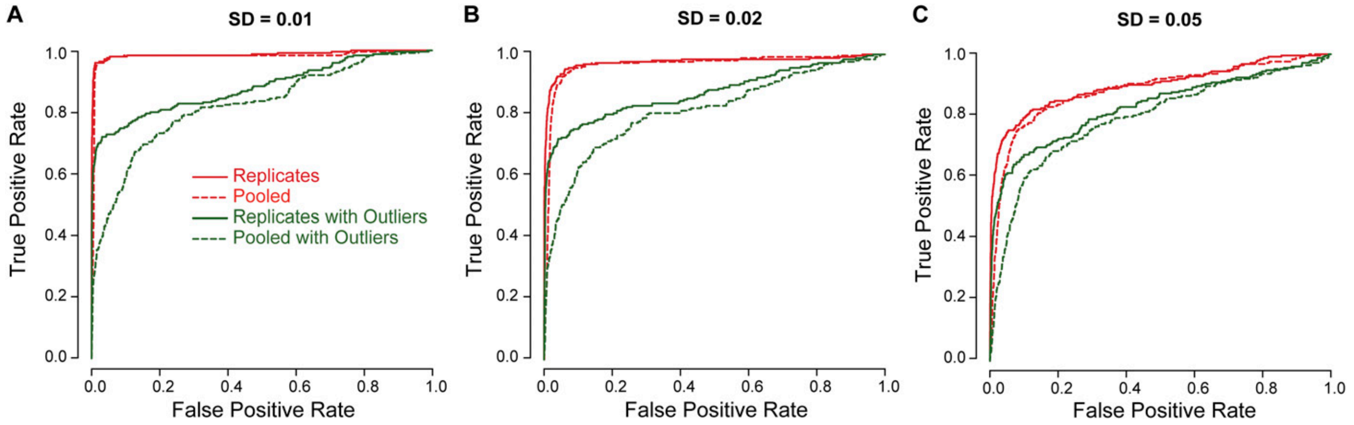

Fig. S2. (*A*–*C*) Simulation studies to assess the performance of rMATS and the importance of replicates where 10% of the exons were differentially spliced and the rest were not differentially spliced.

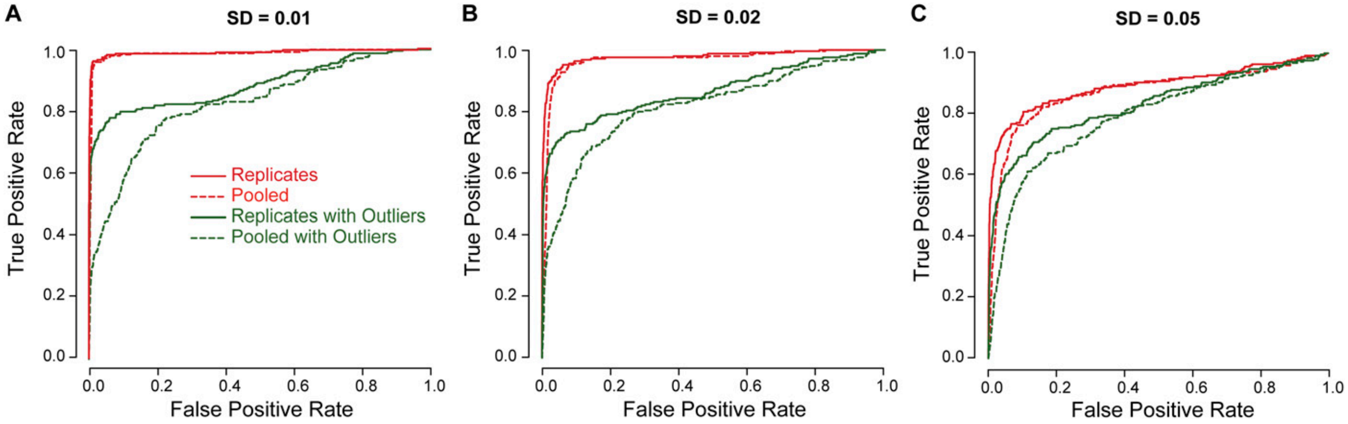

Fig. S3. (*A*–*C*) Simulation studies to assess the performance of rMATS and the importance of replicates where 20% of the exons were differentially spliced and the rest were not differentially spliced.

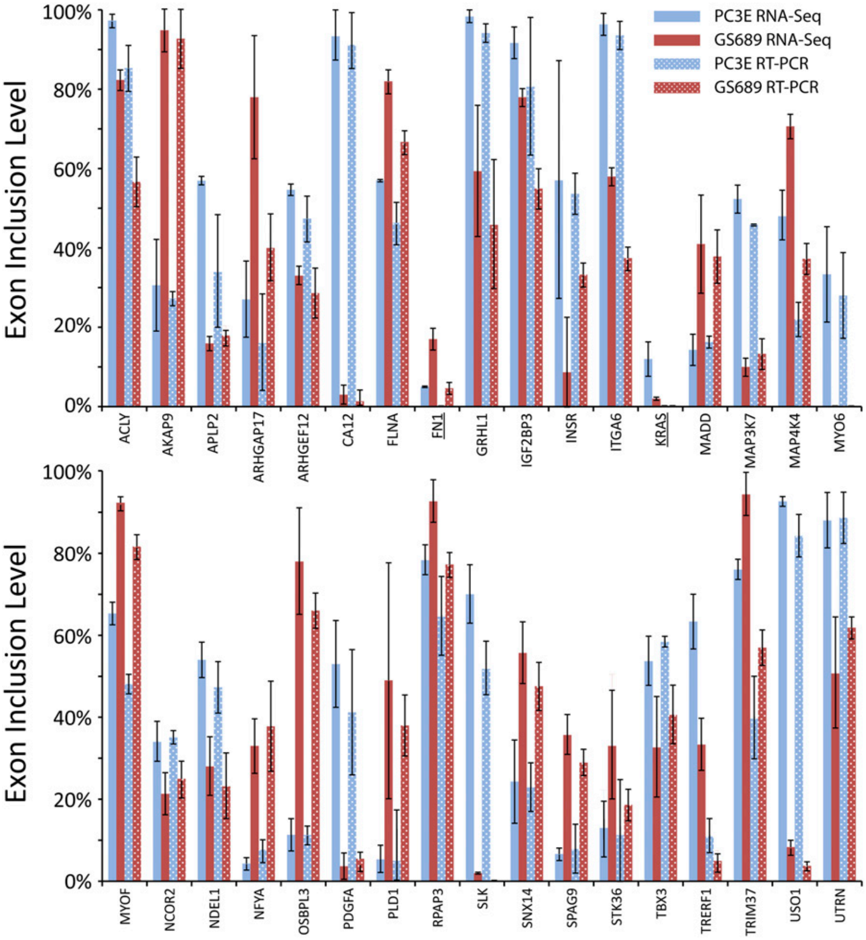

Fig. S4. RNA-Seq data and RT-PCR validation of 34 exons. For each exon, the first two bars represent the mean exon inclusion levels and 95% confidence intervals estimated by RNA-Seq, and the last two bars represent the mean exon inclusion levels and 95% confidence intervals measured by RT-PCR (PC3E, blue; GS689, red). Based on the RT-PCR data, 32 of 34 exons (i.e., 94%) were validated to have differential alternative splicing between the PC3E and GS689 cell lines (i.e., >5% difference in mean exon inclusion levels between the two sample groups). The two nonvalidated exons are underlined (FN1 and KRAS).

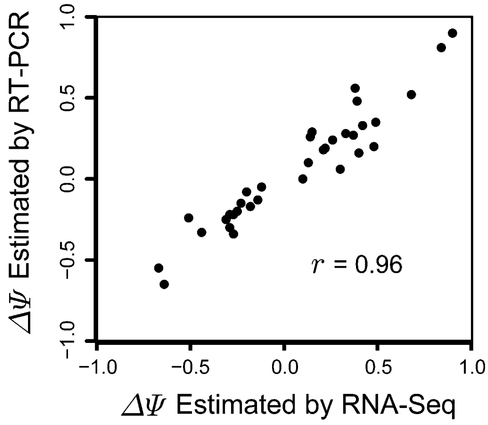

Fig. S5. RNA-Seq estimates of exon inclusion levels are highly correlated with RT-PCR measurements. The difference of the mean exon inclusion levels between the two cell lines (PC3E – GS689) is denoted as $\Delta\psi$. The scatter plot shows the $\Delta\psi$ values of 34 exons estimated by RNA-Seq (*x* axis) and by RT-PCR (*y* axis).

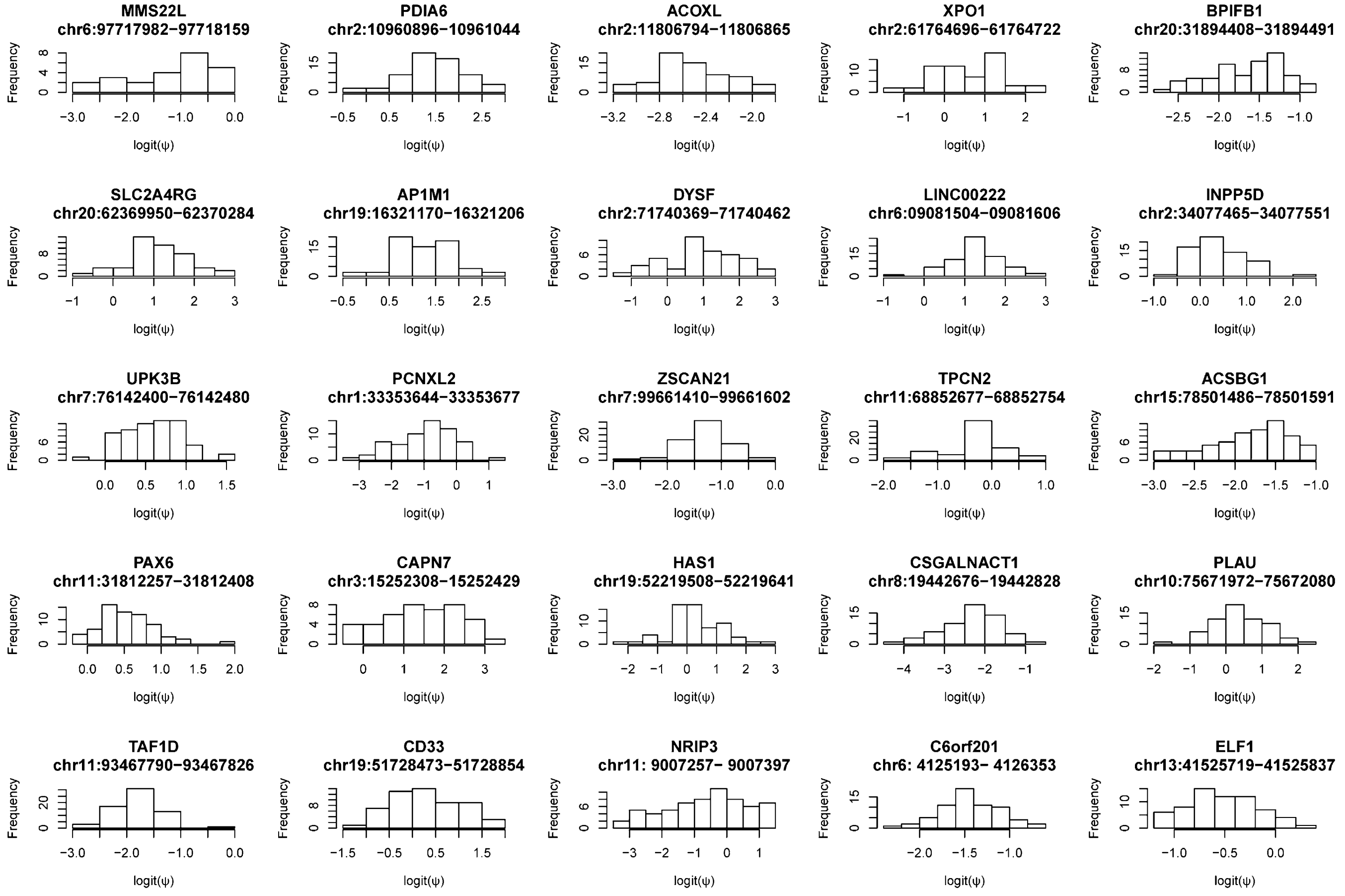

Fig. S6. Histograms of the logit exon inclusion levels of 25 randomly selected alternative exons in TCGA ccRCC normal samples.

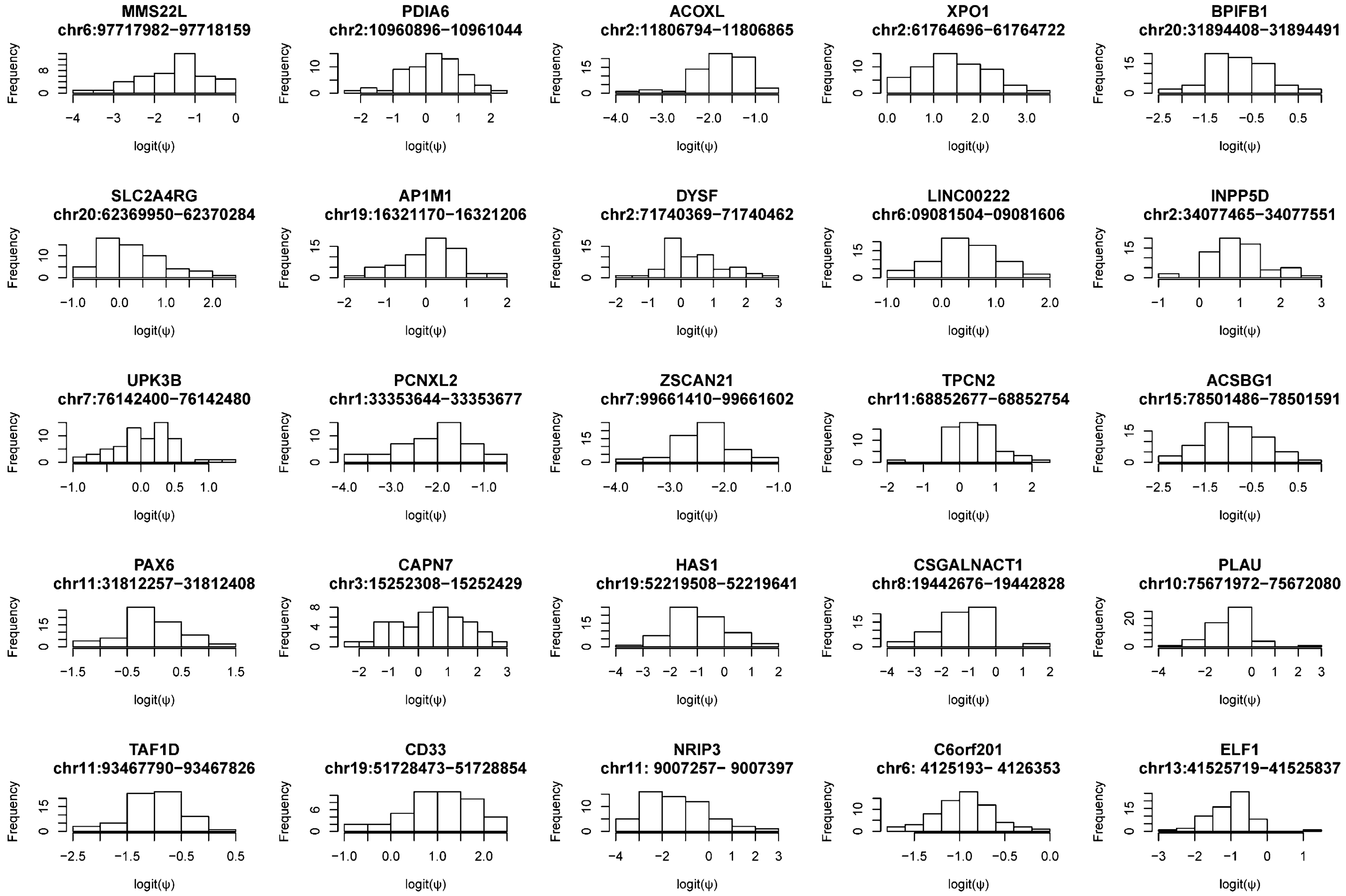

Fig. S7. Histograms of the logit exon inclusion levels of 25 randomly selected alternative exons in TCGA ccRCC tumor samples.

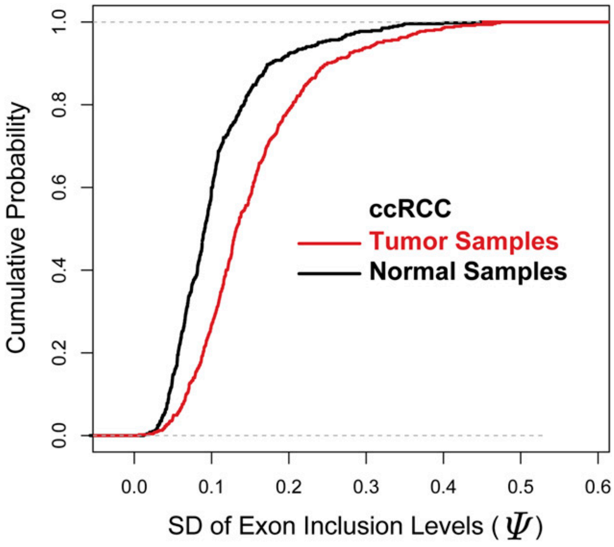

Fig. S8. The cumulative distribution of the SDs of exon inclusion levels in TCGA ccRCC normal or tumor samples. The estimate of SD was obtained by rMATS.

Table S1. Summary and mapping statistics of RNA-Seq data on the PC3E and GS689 cell lines

Samples	Total no. RNA-Seq reads, million*	No. uniquely mapped reads, million (%)	Reads mapped to genomic regions, million (%)	Reads mapped to splice junctions, million (%)
PC3E-1	126	101 (80)	62 (49)	39 (31)
PC3E-2	129	104 (81)	65 (50)	39 (31)
PC3E-3	126	101 (80)	63 (50)	38 (30)
GS689-1	132	104 (79)	68 (52)	36 (27)
GS689-2	114	90 (79)	58 (51)	32 (28)
GS689-3	119	93 (78)	60 (50)	33 (28)

*RNA-Seq reads are 2 × 101-bp paired-end reads.

Table S2. rMATS analysis of PC3E and GS689 cell lines

Alternative splicing events	Total no. alternative splicing events	No. significant alternative splicing events*
Skipped exon	91,856	467
Alternative 5′ splice site	3,664	24
Alternative 3′ splice site	5,138	21
Mutually exclusive exon	24,168	159
Retained intron	3,315	50

*Significant events are based on FDR ≤ 1% and $|\Delta\psi| > 5\%$.

Other Supporting Information Files

Dataset S1 (XLS)
Dataset S2 (XLS)
Dataset S3 (XLS)
Dataset S4 (XLS)
Dataset S5 (XLS)