1 Variational EM for Model Learning

The variational approximation for the joint posterior distribution $Q(s^e, m^e, n^e)$ given the observed splice levels $x^e$ is given by:

$$Q(s^e, m^e, n^e) = \prod_{c=1}^{C} \left( Q(s_c^e)N(m_c^e; \eta_{s_c}^e, \sigma_{s_c}^e) \right) \prod_{t=1}^{T} Q(n_t^e) = \prod_{c=1}^{C} \left( q_c^e N(m_c^e; \eta_{s_c}^e, \sigma_{s_c}^e) \right) \prod_{t=1}^{T} q_{n_t}^e,$$

(1)

where $\eta_{s_c}^e, \sigma_{s_c}^e$ are variational parameters governing the posterior distribution over $m_c^e$, the probability $Q(s_c^e)$ is the posterior for signal assignment $s_c$, and $Q(n_t^e)$ gives the posterior for the measurement $x_t^e$ to come from the noise ($n_t = 1$) or signal ($n_t = 0$) model. Given this variational approximation, learning the model amounts to minimizing the free energy:

$$F = \sum_{e} \sum_{s^e, m^e, n^e} \int_{m^e} Q(s^e, m^e, n^e) \log \frac{Q(s^e, m^e, n^e)}{P(s^e, m^e, n^e, x^e)}.$$

(2)

This can be rearranged to the more intuitive form:

$$F = \sum_{e} D_{KL}(Q(s^e)\|P(s, M, N, x^e)) + \sum_{t=1}^{T} D_{KL}(Q(n_t^e)\|P(n_t)) + D_{KL}(Q(m^e \mid s^e, n^e)\|P(m, x \mid s, n)),$$

(3)

where $D_{KL}$ denotes the Kullback-Leibler divergence. Because of the Gaussian assumptions the integral is analytic and because $Q(n^e)$ and $Q(s^e)$ factorize across the $T$ conditions and the $C$ signals, the sums over $n$ and $s$ simplify so that their computation takes time that is linear in the number of conditions and the number of signals. Combining Eq. 2,3 from the main text with the form of the prior distributions and the above equations, we get after further mathematical manipulation:

$$F = \sum_{e} D_{KL}(Q(s_c)\|P(s_c)) + \sum_{t=1}^{T} D_{KL}(Q(n_t)\|P(n_t)) +$$

$$+ \sum_{t=1}^{T} \left[ \frac{1}{2} \log 2\pi + \frac{1}{2} q_{n_t=0}(\log \psi_t^2 + \psi_t^{-2} x_t^2) + \frac{1}{2} q_{n_t=1}(\log \phi_t^2 + \phi_t^{-2} x_t^2 - 2\lambda_t^2 x_t\eta_{s_c} + \phi^{-2} \mu_t^2) +$$

$$+ \sum_{k \neq k'} \sum_{s_c \neq 0} \frac{1}{2} q_{s_c} \left( -1 - \log \sigma_{s_c}^2 + \sigma_{s_c}^2 + \eta_{s_c}^2 - 2q_{n_t=0} \psi_t^2 x_t \lambda_t \eta_{s_c} + 2q_{n_t=0} \psi_t^{-2} x_t \lambda_t \eta_{s_c} + \right.$$

$$+ \left[ q_{n_t=0} \psi_t^{-2} \lambda_t \eta_{s_c} \sum_{s_c' \neq 0} \lambda_{t,k'} \sum_{s_c' \neq 0} Q(s_c') \eta_{s_c'} \right] \tau_{t,c} (\eta_{s_c}^2 + \sigma_{s_c}^2) \right].$$

(4)

where, with a slight abuse of notation, we removed the explicit dependency on $e$ and the sum over it for compactness. During learning, the free energy defined above is minimized iteratively using a variational
EM procedure[7, 5]. In short, at every maximization (M) step, the derivative of Eq. (4) with respect to the model’s parameters is computed, while the expectation (E) step involves a short series of point estimates for the expected assignment or sufficient statistic \(Q(n_t), Q(s_c), \eta_{t,c}, \sigma_{t,c}^{}\) with the model’s parameters held fixed. As usual with EM based algorithm, this iteration continues until convergence to a local minimum is reached. In the experiments described, we used \(10^{-5}\) fold change as the stopping criteria. Finally, given the learned model we can infer the desired signal assignments to each exon by performing inference and computing \(\{q_{c,e}^*\} \forall c, e\). Using Eq. (4) we solve for \(\frac{\partial F}{\partial q_{c,e}} = 0\) to get:

\[
\log q_{c,e} = \log P(s_c) - 1 + \frac{N}{2} \left(1 + \log \sigma_{s_c}^2 - \sigma_{s_c}^2 - \eta_{s_c}^2 + 2\eta_{s_c}^2 \nu_c - \nu_c^2\right) + \\
+ \sum_{t=1}^T q_{n_t=0} v_t^{-2} \lambda_{t,c} \left[ x_t \eta_{s_c} - \eta_{s_c} \left( \sum_{c' \neq c} \lambda_{t,c'} \sum_{s_{c'} \neq 0} q_{n_{t,c'}} \eta_{s_{c'}} \right) \right] - \frac{1}{2} \lambda_{t,c} (\eta_{s_c}^2 + \sigma_{s_c}^2) \tag{5}
\]

## 2 Using Gene Expression to Update Signal Model Posterior

High-throughput AS measurements typically include estimation for the overall gene expression for the genes corresponding to each exon monitored in the experiments. For the data set of [2], two sets of expression measurements were available: log abundance estimates for the genes corresponding to each exon monitored in the experiment, denoted \(\{v_t^*\} \in \mathcal{R}\), and a set of measurements from negative probes \(\{v^*\} \in \mathcal{R}\), for genes presumably not expressed. The distribution over the two sets is shown in Fig. 1A. A common approach to utilize such information is to set a threshold over expression values, to make sure at least a predefined percentage of the negative probes are excluded. The result of applying this approach is shown in Fig. 1B. The red line correspondence to a threshold that removes 95% of the negative probes, while the blue dashed line indicates that using that threshold would retain about 75% of the original measurements and remove about a quarter. Fig. 1C shows the result of applying this approach to a noisier data set, derived from a higher density micro array and various human tissue (Blencow lab, unpublished data). In this case, using such a threshold would result in removing over 70% of the original data.

Instead of using a hard threshold to include or reject data in subsequent analysis, our probabilistic framework allows us to incorporate the gene expression measurement \(v_t^*\) into the model in order to update the posterior belief that an AS measurements \(x_t^*\) comes from the signal or the noise model. For the expression measurements we assume a generative mixture distribution model where the observed expression level can be generated from two distinct distributions: The first is the distribution over noise measurements, which is the distribution that the measurements for the negative probes \(\{v^*\}\) were sampled from. The second distribution is an unknown distribution over expression values corresponding to the signal model. According to our model, the observed expression values \(\{v_t^*\}\) are a set sampled from a mixture of these two distributions, with some unknown mixing proportions. The probability of an observed expression value under this mixture model can be written as:

\[
P(v_t^*) = P(n=1)P(v_t^* | \Theta_{n=1}) + P(n=0)P(v_t^* | \Theta_{n=0}),
\]

Where \(P(n=1) = 1 - P(n=0)\) give the mixing proportions, while \(\Theta_{n=0}\) and \(\Theta_{n=1}\) are the parameters that govern the signal and noise expression level distributions. According to this model, given the observed data \(\{v_t^*\}\) and \(\{v^*\}\), our task is to find the model parameters \(P(n=1), \Theta_{n=0}, \text{ and } \Theta_{n=1}\) that maximize the likelihood of the data. We do this by first estimating \(\Theta_{n=0}\) from \(\{v^*\}\), then we fix these parameters and run EM for mixture model learning to fit \(P(n=1), \text{ and } \Theta_{n=1}\). As usual with EM based learning, a good initialization point is important. To achieve this, we used Matlab’s built in package for parametrized distributions fitting to derive the initial parameters. After the model’s parameters have been learned we compute for each AS measurement \(x_t^*\) the posterior it came from the signal model based on the observed expression value:

\[
P(n=0 | v_t^*) = \frac{P(v_t^* | \Theta_{n=0})}{P(v_t^*)}
\]

This posterior as a function of the expression value is plotted in Fig. 1B,C as a dashed black line. The value \(P(n=0 | v_t^*)\), is plugged into the AS model presented in the main text, to update the belief each AS measurement comes from the signal rather than the noise model. Overall, for the data set of [2], the noise model marginal probability we inferred was \(P(n=0) \sim 11\%\), while for the second data set shown in Fig. 1C, the noise model marginal was more than doubled \((P(n=0) \sim 23\%)\).
3 The Feature Information Index

In order to have a more quantitative measure of how identified AS signals correspond to known regulatory features, we did extensive literature survey for cis elements associated with splice factors that are known to have tissue-specific regulatory effect. These include motifs for CUG-rich, Mbnl and the CUGBP binding tracts [4], PTB/nPTB binding tracts [1, 8, 9], Fox [6], Quaking [3] and Nova [10]. We note that since current available regulatory information involves almost exclusively CNS and muscle tissues, our analysis concentrated only on AS signals reflecting splicing changes in those tissues. For each of the cis element in our set, we scored its enrichment, using the hyper-geometric p-value (-log scaled), in a group of exons. These motifs were searched in the alternative exon, the constitutive exons flanking it, and the intronic regions flanking those exons. The feature information index (FII) for a splicing signal in a given data set (e.g., splicing changes in CNS), was computed by summing over all the features enrichment scores for this group. When different motifs were available for the same SF, we only included the higher scoring one.

References

1. M. Ashiya and P.J. Grabowski. A neuron-specific splicing switch mediated by an array of pre-mRNA repressor sites: evidence of a regulatory role for the polypyrimidine tract binding protein and a brain-specific PTB counterpart. RNA, 3:996–1015, September 1997.
2. M. Fagnani, Y. Barash, J. Ip, C. Misquitta, Q. Pan, A. L. Saltzman, O. Shai, L. Lee, A. Rozenhek, N. Mohammad, S. Willaine-Morawek, T. Babak, W. Zhang, T. R. Hughes, D. van der Kooy, B. J. Frey, and B. J. Blencowe. Functional coordination of alternative splicing in the mammalian central nervous system. Genome Biology, 8:R108+, June 2007.
3. A. Galarmeau and S. Richard. Target RNA motif and target mRNAs of the Quaking STAR protein. Nature Structural & Molecular Biology, 12(8):691–698, July 2005.
4. Thai H. Ho, N. Charlet-B, M. G. Poulos, G. Singh, M. S Swanson, and T. A. Cooper. Muscleblind proteins regulate alternative splicing. EMBO, 23:3103–3112, 2004.
5. M. I. Jordan, Z. Ghahramani, T. Jaakkola, and L. K. Saul. An Introduction to Variational Methods for Graphical Models. Machine Learning, 37(2):183–233, 1999.
6. S. Minovitsky, S. L. Gee, S. Schokrpur, I. Dubchak, and J. G. Conboy. The splicing regulatory element, UGCAUG, is phylogenetically and spatially conserved in introns that flank tissue-specific alternative exons. *NAR*, 33:714–724, 2005.

7. R. Neal and G. Hinton. A View of the EM Algorithm that Justifies Incremental, Sparse, and other Variants. In M. I. Jordan, editor, *Learning in Graphical Models*. Kluwer, 1998.

8. F. C. Oberstrass, S. D. Auweter, M. Erat, Y. Hargous, A. Henning, P. Wenter, L. Reymond, B. Amir-Ahmady, S. Pitsch, D. L. Black, and F. H.-T. Allain. Structure of PTB Bound to RNA: Specific Binding and Implications for Splicing Regulation. *Science*, 309(5743):2054–2057, 2005.

9. I. Perez, C.H. Lin, J.G. McAfee, and J.G. Patton. Mutation of PTB binding sites causes misregulation of alternative 3’ splice site selection in vivo. *RNA*, 3:764–768, 1997.

10. J. Ule, G. Stefani, A. Mele, M. Ruggiu, X. Wang, B. Taneri, T. Gaasterland, B. J. Blencowe, and R. B. Darnell. An RNA map predicting Nova-dependent splicing regulation. *Nature*, October 2006.