molecular systems biology

Cell-to-cell variability of alternative RNA splicing

Zeev Waks, Allon Klein, Pamela Silver

Corresponding author: Pamela Silver, Harvard Medical School

Review timeline:

Submission date:	10 March 2011
Editorial Decision:	14 April 2011
Revision received:	26 April 2011
Accepted:	29 April 2011

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision 14 April 2011

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from the three referees whom we asked to evaluate your manuscript. As you will see from the reports below, the referees are generally supportive and find the topic of your study interesting. However, they make several constructive suggestions for modifications and also raise some concerns, which should be carefully addressed in a revision of the present work. The recommendations provided are very clear in this regard and refer to the need for additional clarifications and toning down some of the conclusions (eg on spatial location of splicing).

Note: in addition to our capacity to host datasets in our supplementary in formation section, we provide a functionality on our website, which allows readers to directly download the 'source data' associated with selected figure panels (see example at <http://tinyurl.com/365zpej>), for the purpose of alternative visualization, re-analysis or integration with other data. These files are separate from the traditional supplementary information files and are directly linked to specific figure panels.

In the case of this study, we would strongly encourage you to submit the individual files corresponding to the quantitative data showed in the various figure panels included in this study. We provide below some general guidelines with regard to the format of such data tables.

*** PLEASE NOTE *** As part of the EMBO Publications transparent editorial process initiative (see our Editorial at http://www.nature.com/msb/journal/v6/n1/full/msb201072.html), Molecular Systems Biology will publish online a Review Process File to accompany accepted manuscripts. When preparing your letter of response, please be aware that in the event of acceptance, your cover letter/point-by-point document will be included as part of this File, which will be available to the scientific community. More information about this initiative is available in our Instructions to

Authors. If you have any questions about this initiative, please contact the editorial office msb@embo.org.

If you feel you can satisfactorily deal with the points listed by the referees, please submit a revised version of your manuscript. Please attach a covering letter giving details of the way in which you have handled each of the points raised by the referees.

I thank you again for submitting this manuscript and look forward to reading your revised study.

Best wishes,

Editor

Molecular Systems Biology

REFEREE REPORTS

Reviewer #1 (Remarks to the Author):

"Cell to cell variability of alternative splicing" (from Waks, Klein & Silver) is a creative and clearly written piece of important science, that opens the door to studying an underappreciated and challenging set of questions about splicing. These questions have to do with the within-cell and between-cell type fluctuations in isoform composition for mRNA pools from alternatively spliced genes. These concerns have been ignored in the study of splicing thus far, which have employed average measurements of large cell populations. I think this is a very strong effort and have only a few quibbles; these are listed below. Despite my characterization of them as quibbles, this is a seminal paper in the field and thus precision may matter more here than in a typical paper.

Quibble 1. Abstract term "Non-genetic heterogeneity". Non-genetic is a kind of non-term that is not required at this high position in a paper which does not work very hard to rule out "genetic heterogeneity", which I guess the authors strictly mean: gene expression noise due to gene loss or amplification.

Quibble 2. Also in the abstract: "Analysis of the potential sources of isoform ratio heterogeneity indicates that misregulation of splicing factor activity is the primary origin of this increase." This statement is based on a knockdown of a splicing factor that dramatically reduces the uniformity of splicing isoform composition observed in untreated cells. But the statement sounds like a biological mechanism. In addition (and also in the text where this is discussed) is the assumption that good cell regulatory mechanisms are necessarily free of noise and that bad ones are noisy (Hela cells being "bad" relative to Rpe1 cells). Perhaps a statement that relies less on our expectations of good and bad cells would be that cells differ in the extent to which they control the noise in alternative splicing, and that one way in which they might do this is to impose stricter control on splicing factor expression. This is really a hypothesis that comes out of a gross disruption, rather than say, a correlation between cellular ASF expression and cellular FIA of the target. It is really important not to overstate these things in the abstract.

Quibble 3. I don't prefer random as a descriptor for alternative splicing, not because it isn't strictly technically accurate but because many readers who may not be familiar with splicing won't recognize that the authors means "a stochastic or probabilistic choice between strict alternatives."

Reviewer #2 (Remarks to the Author):

In this manuscript, the authors make use of single molecule FISH to detect alternatively spliced isoforms of 2 genes at the resolution of individual RNA molecules in single cells. Using this assay, they are able to measure the variability in relative isoform abundance from cell to cell, showing that in some cases, the variability is near the theoretical minimum, and that the variability is higher in other cases where there is potential misregulation of splicing. Overall, I found this paper to be of high quality and feel that it makes a substantial contribution to (and connection between) the fields of alternative splicing/RNA processing and stochastic gene expression.

To my knowledge, this is the first application of single molecule transcript quantification to the direct detection of alternative splicing, and that fact alone would make this a useful paper. I think the authors took a very interesting next step by looking at variability in alternative splicing, which is nice because the stochastic gene expression field has largely just focused on "vanilla gene expression" to this point. Moreover, their experiments suggest that cellular mechanisms govern alternative splicing variability. I also liked the model that they employed to place their data in context: it was not overly complex and provided insight that allowed one to make interesting interpretations.

I do have a few suggestions to improve the paper, though:

The authors make note in their abstract that they are able to extract spatial information about alternative splicing from their images. However, in the paper itself, the spatial information they report is mostly just observational, showing that the unspliced mRNA is visible away from the transcription site. However, they don't really use this information any further in the paper. A full spatial model of alternative splicing and how that may affect variability may be beyond the scope of this paper, but it would be nice if the authors could speculate on what the spatial distributions may mean with regards to variability in alternative splicing (or indeed what implications it has in general).

The authors ignore the binomial partitioning of biomolecules upon cell division. I understand that this increases the complexity of the modeling, but I think it would be useful to include at least a few simulations showing how much additional variability one might find if one included this effect. One might expect some contribution given the relatively long half-lives involved (comparable to the cell division times). In particular, are the division times of the two cell lines roughly the same? If binomial partitioning is a contributing factor, then a cell line with a faster division time would have higher variability just due to this effect.

Also, I find the arguments relating to the ASF/SF2 knockdown to be fairly unconvincing as presented. It is perhaps true that the removing this particular RNA binding protein might remove some autoregulation that normally decreases variability, but this is pure speculation and is a claim that requires direct proof. The IF images suggest that the RNAi is effective in every cell, but it is possible that variability in the small remaining amounts due to varying effectiveness of RNAi of ASF/SF2 is what is transmitting the variability. Without knowing, however, it's a bit hard to interpret the meaning of this particular experiment.

Minor point: the best primary references for mRNA diffusion measurements are Shav-Tal et al. Science 2005 and Vargas et al. PNAS 2006.

These points aside, however, this is a very interesting paper that brings some of the analysis of stochastic gene expression to the field of alternative splicing.

Reviewer #3 (Remarks to the Author):

The authors here analyze cell-to-cell variability of alternative splicing. They choose two transcripts that have known splice variants and can be detected using smFISH. DNA probes were designed to distinguish between the different isoforms. Using smFISH the authors quantified the amount of each isoform in individual cells. This technique was used to find the active transcription site - as the brightest point in the nucleus of both colors. The distance of the unspliced mRNA (co-localization of both colors) to the active transcription site indicates transcription was found not to be coupled to splicing. Next the variability of CAPRIN1 and MKNK2 isoforms was assessed in both Rpe1 and HeLa cells. The variability of CAPRIN1 isoform ratio was smaller than that of MKNK2. In addition the isoform ratio of both transcripts was greater in HeLa cells. Knock-down of SRSF1, a known regulator of MKNK2 splicing, increase the fluctuation of MKNK2 isoform ratio. A computational model is then generated to test some of these ideas.

This is overall a well-designed set of experiments that address an interesting topic. The study is innovative both at the technical and conceptual level. There are two main conclusions: The first regarding the localization of splicing relative to its transcription site is only poorly supported by the data and should be removed or strongly toned down (see below). The second regarding the stochastic nature of AS represents the major advance and is of interest.

Specific points:

There are several major problems with the conclusion regarding the location of splicing. These points should either be experimentally addressed or the conclusions removed or toned down:

- Figure 4D: The authors claim the identify the site of active transcription based on "high intensity in both channels" (p. 7); yet in figure 4D the intensity of the unspliced RNA is given as 0 at the transcription site? This seems a contradiction. Along the same lines, if the active transcription site is determined by nuclear spots with high intensity in both channels, how can the author distinguish if the mRNA is spliced or unspliced? Why is there ~ 40% mature mRNA at the transcription site and almost 0 unspliced mRNA?

-The use of ActD as a way to test the authors approach to identifying an active transcript site is problematic since it also affects the amount of the transcripts, and hence changes the entire picture. In order to better achieve this goal one would need to show co-localization of "the nuclear spots with high intensity in both channels" with PolII.

- CAPRIN1 and MKNK2 alternative isoforms both differ in their terminal exons. The authors use these two genes to extrapolate to all splicing events by making statements such as "unspliced mRNA is observed away from transcription sites". They have not tested whether other types of splicing events behave in a similar fashion. Generalizations based on these two genes should be toned down.

- Unspliced mRNA can only be detected in CAPRIN1 since the MKNK2 system is set up in such a way that isoform 1 can be distinguished from isoform 2, but the unspliced RNA can not be detected specifically. Hence, the analysis of distance of the unspliced MKNK2 mRNA from the active transcription site in not valid and should be removed.

Other points:

The authors state at various point in the manuscript that their observations suggest that compromising tight regulation of the splicing machinery may be relevant to disease. This is based on the observed differences between a "normal' and a "cancer" cell line. This seems a premature conclusion. Clearly more and better characterized cell lines would need to be tested to make this generalized statement. Please remove these statements.

Knocking down ASF/SF2 increased the FIA variability of MKNK2, but that does not indicate that "the regulatory splicing machinery is misregulated in HeLa cells" (mid page 13).

The authors conclude that "fluctuations in the regulatory splicing machinery" are responsible for the observed effects. This conclusion is mostly based on a SF2 knockdown experiment. If the effects are indeed due to fluctuations overexpression of SF2 should give the same result. This is a key experiment that should be done and would greatly strengthen the conclusion.

I am wondering whether the difference in the size of cell-to-cell variability is related to the difference in isoform ration between the two genes. CAPRIN1 expresses 95% of one isoform and shows low variability, MKNK2 isoforms are present at similar levels with larger variability. One explanation is that an active mechanism drives CAPRIN1 to a dominant expression pattern thus reducing the variability. This would have less to do with fluctuations of the splicing machinery but dedicated mechanisms. This model should be discussed.

Minor points

The sequence of figure panels in figures 2 and 3 does not follow the text and is very confusing. I suggest re-structuring the figures to make it correspond to the order the data is discussed in the text.

The nomenclature for SR proteins has recently been unified. See Krainer and Manley, G&D, 2010. The new nomenclature for ASF/SF2 is SRSF1. Please use throughout manuscript.

1st Revision - authors' response 26 April 2011

Please find enclosed a revised version of our manuscript, "Cell-to-cell variability of alternative RNA splicing".

We wish to thank you for considering our study for publication in Molecular Systems Biology. We are very pleased that the reviewers found our manuscript interesting, and would like to thank them for their comments, which help further improve the manuscript. Below you will find our specific, point by point, response to all of the reviewers' suggestions. As you will see, we agreed with the majority of the reviewers' suggestions and have implemented them accordingly. In addition, we submitted a thumbnail image, 'standfirst text', 3 bullet points summarizing the findings of our study, an extended synopsis, and a source data file that corresponds to quantitative data in figures 4, 5, and 6.

We hope our paper is now acceptable for publication and look forward to your response.

RESPONSES TO REVEIWERS

Response to Reviewer 1:

We wish to thank Reviewer 1 for describing our paper as "seminal".

Below we address the specific comments.

1. Abstract term "Non-genetic heterogeneity". Non-genetic is a kind of non-term that is not required at this high position in a paper which does not work very hard to rule out "genetic heterogeneity", which I guess the authors strictly mean: gene expression noise due to gene loss or amplification

RESPONSE:

We changed the first sentence of the abstract (page 2, lines 2-3) accordingly from "Non-genetic heterogeneity in the expression levels of mammalian cells is large and has phenotypic consequences" to: "Heterogeneity in the expression levels of mammalian genes is large even in clonal populations and has phenotypic consequences". We also replaced the keyword "non-genetic heterogeneity" with "cell-to-cell variability".

2. Also in the abstract: "Analysis of the potential sources of isoform ratio heterogeneity indicates that misregulation of splicing factor activity is the primary origin of this increase." This statement is based on a knockdown of a splicing factor that dramatically reduces the uniformity of splicing isoform composition observed in untreated cells. But the statement sounds like a biological mechanism. In addition (and also in the text where this is discussed) is the assumption that good cell regulatory mechanisms are necessarily free of noise and that bad ones are noisy (HeLa cells being "bad" relative to Rpe1 cells). Perhaps a statement that relies less on our expectations of good and bad cells would be that cells differ in the extent to which they control the noise in alternative splicing, and that one way in which they might do this is to impose stricter control on splicing factor expression. This is really a hypothesis that comes out of a gross disruption, rather than say, a correlation between cellular ASF expression and cellular FIA of the target. It is really important not to overstate these things in the abstract.

RESPONSE:

We agree with this comment and have changed the relevant sentences accordingly to tone it down, specifically not to mention misregulation.

A. Abstract (page 2, lines 8 to 10):

a. Original sentence: "Analysis of the potential sources of isoform ratio heterogeneity indicates that misregulation of splicing factor activity is the primary origin of this increase."

b. New Sentence: "Analysis of the potential sources of isoform ratio heterogeneity indicates that a difference in the control over splicing factor activity is one origin of this increase."

B. Final paragraph in introduction (page 4, lines 11-13):

a. Original sentence: "Further analysis and knockdown of a key splicing factor showed this increased variability to be a function of misregulation of the regulatory splicing machinery."

b. New sentence: "Further statistical analysis of potential variability sources and knockdown of a key splicing factor suggested that the increased variability in HeLa cells is a function of reduced control over the regulatory splicing machinery."

C. In the results section pertaining to sources of variability (pages 9, lines 14-17) , the following paragraph was changed:

a. Original paragraph: "In comparison to this theoretical minimum, we found that FIA variability in HeLa cells was much larger than in Rpe1 cells (4.2 fold for CAPRIN1 and 3.1 fold for MKNK2), indicating that variability in isoform ratios is misregulated in this cancer cell line (Figure 5E). Remarkably, FIA variability in the non-transformed Rpe1 cells was close to the minimum possible (Figure 5E), with only a ~9% increase above the minimum for CAPRIN1 and ~20% for MKNK2."

b. New paragraph: "In comparison to this theoretical minimum, we found that FIA variability in HeLa cells was much larger than in Rpe1 cells (4.2 fold for CAPRIN1 and 3.1 fold for MKNK2) (Figure 5E). Remarkably, FIA variability in the Rpe1 cells was close to the minimum possible (Figure 5E), with only a ~9% increase above the minimum for CAPRIN1 and ~20% for MKNK2.

D. First paragraph in the discussion (page 13, line 21):

a. Original sentence: "We also found evidence that the regulatory splicing machinery is misregulated in HeLa cells."

b. New sentence: "We also found that isoform ratio variability is increased in HeLa cells."

3. I don't prefer random as a descriptor for alternative splicing, not because it isn't strictly technically accurate but because many readers who may not be familiar with splicing won't recognize that the authors means "a stochastic or probabilistic choice between strict alternatives."

RESPONSE:

We have changed the text accordingly to remove the word random in the following positions:

A. Figure 6A: The term "Random outcome of splicing (binomial)" was changed to "Stochastic outcome of splicing (binomial)".

B. Abstract (page 2, line 7): The term random was changed to stochastic in the following sentence: "We show that isoform variability in non-transformed, diploid cells is remarkably close to the minimum possible given the stochastic nature of individual splicing events, while variability in HeLa cells is considerably higher."

C. In the following sentence, the term randomness was changed to stochasticity (page 9, line 8): "As mRNAs are typically present in low numbers, the most obvious source of isoform ratio variability is the stochasticity of individual alternative splicing events."

D. In the discussion, the term randomness was changed to probabilistic chance in the following sentence (page 13, lines 18-21): "In this study, we acquired single molecule images of the alternative splicing of two endogenous genes and found that isoform ratio variability was close to the minimum possible variability resulting from the probabilistic chance of individual alternative splicing events."

Response to Reviewer 2:

1. The authors make note in their abstract that they are able to extract spatial information about alternative splicing from their images. However, in the paper itself, the spatial information they report is mostly just observational, showing that the unspliced mRNA is visible away from the transcription site. However, they don't really use this information any further in the paper. A full spatial model of alternative splicing and how that may affect variability may be beyond the scope of this paper, but it would be nice if the authors could speculate on what the spatial distributions may mean with regards to variability in alternative splicing (or indeed what implications it has in general).

RESPONSE:

We agree that this is an interesting point to discuss and have accordingly added the following paragraph to the discussion (page 15, lines 14-20).

"The observation that the alternative splicing of the terminal exon of CAPRIN1 is not necessarily co-transcriptional may have implications regarding isoform ratio variability. If the observed unspliced mRNAs are not misprocessed transcripts but rather functional pre-mRNAs, it is possible that the spatial localization of alternative splicing affects cell-to-cell variability in isoform ratios as the concentration of regulatory splicing factors may not be uniform within the nucleus. Conceptually, this source of variability would fall under the category of fluctuations in the regulatory splicing machinery."

2. The authors ignore the binomial partitioning of biomolecules upon cell division. I understand that this increases the complexity of the modeling, but I think it would be useful to include at least a few simulations showing how much additional variability one might find if one included this effect. One might expect some contribution given the relatively long half-lives involved (comparable to the cell division times). In particular, are the division times of the two cell lines roughly the same? If binomial partitioning is a contributing factor, then a cell line with a faster division time would have higher variability just due to this effect.

RESPONSE:

We thank Reviewer 2 for this request; indeed it is interesting to consider the effect of binomial partitioning on the isoform ratios. We did consider this point during the study, but noticed that the effects of binomial partitioning upon cell division were relatively small and thus did not mention them in the first submission for simplicity. However, given that the reviewer's interest may reflect a wider interest in this question, it should not be ignored.

Referring to Huh and Paulsson (2011), the effect of binomial partitioning is analogous to introducing an additional channel of mRNA degradation with an exponentially-distributed lifetime; both cases introduce a Poisson noise into the mRNA number, giving a variance in mRNA abundance that is proportional to the mean abundance. Significantly, the effect of binomial partitioning is captured by the Poisson term of the isoform variance (which implicitly accounts for cell division owing to the lower mean isoform abundance), and by revisions to the time-averaging function A[x] (see Supp Theory) that reflect the change in the mRNA lifetime distribution. Similarly, a change in cell cycle length will shorten the effective mRNA lifetimes (as already discussed in Supp. Theory), leading to the same revisions. We therefore should not expect significant changes following inclusion of binomial partitioning. (This would not be true if the mRNA lifetimes were far longer than the cell cycle time, in which case the Poisson term would be generated primarily by binomial partitioning).

While we feel that a complete discussion of this point would complicate the current paper, the reviewer's suggestion to conduct simulations of the effect of partitioning is reasonable. We have now conducted the relevant simulations and describe the result in the revised manuscript (page 12, lines 9-11). As we expected from the above discussion, the contribution of binomial partitioning of isoforms to FIA variability was found to be small. We used parameter values that correspond to MKNK2 and CAPRIN1 isoform statistics in Rpe1 cells, and found a drop of <10% from the original FIA variability in both cases (8% and 9%, respectively).

Following the reviewer's request, we also revisited our simulations assuming different cell cycle times while maintaining the same average isoform abundance. The division times had only a minor effect. Differences between a 16hr and 24hr division time resulted in a FIA standard deviation change of <4% (3% and 2% for MKNK2 and CAPRIN1). Given that the difference in cell cycle times between the two cell lines is commensurate with this range, we conclude that there is no significant contribution from cell cycle effects.

For completeness, it is worth noting that cell division does introduce additional terms to the variance and covariance, but these are not the result of binomial partitioning. Instead, they are the result of changing mRNA levels over the cell cycle. However, the effect on the isoform ratio, as seen in FIA variability, can be shown to be small. Demonstration of these cell cycle effects requires further (and

far from elegant) theoretical analysis, and, as the effect is small we have decided to exclude it from the current manuscript.

The changes that we made to the manuscript include the addition of a methods section that describes the simulations (page 22, lines 3-11), and a sentence stating that the effect of binomial partitioning upon cell division is small (<10%) for both genes. The added sentence in the results section is (page 12, lines 9-11):

"Simulations of the model that include the full effects of cell division and binomial mRNA partitioning resulted in only a minor increase in FIA variability for both genes (<10% increase) (Materials and methods)."

The cell division times are mentioned in the original and revised draft in the methods (page 22, lines 17-20):

"Figure 6D presents results using doubling times of Tc=24hr and 18hr for HeLa and Rpe1 cells, as previously reported (Uetake and Sluder, 2004; Zocchi et al, 1998). Simulations using cell cycle durations ranging from 16-24hrs had minimal effect on the CV, at most 5%."

3. Also, I find the arguments relating to the ASF/SF2 knockdown to be fairly unconvincing as presented. It is perhaps true that the removing this particular RNA binding protein might remove some autoregulation that normally decreases variability, but this is pure speculation and is a claim that requires direct proof. The IF images suggest that the RNAi is effective in every cell, but it is possible that variability in the small remaining amounts due to varying effectiveness of RNAi of ASF/SF2 is what is transmitting the variability. Without knowing, however, it's a bit hard to interpret the meaning of this particular experiment.

RESPONSE:

While knockdown appears efficient across all cells, it is true that there may be small differences in knockdown efficiency between cells that can also lead to an increase in isoform ratio variability. To address this possibility, we have added the following sentence into the text (page 11, lines 9-11): "Nonetheless, although knockdown was apparent across all cells, it is possible that variability in the residual abundance of SRSF1 (ASF/SF2) contributed to the observed effect as well."

4. The best primary references for mRNA diffusion measurements are Shav-Tal et al. Science 2005 and Vargas et al. PNAS 2006.

RESPONSE:

The suggested references were added to the text in the sentence that discusses mRNA diffusion (page 8, lines 4-8). We believe the reviewer was referring to the Shav-Tal Y et al. Science paper from 2004 and the Vargas DY et al. PNAS paper from 2005.

Response to Reviewer 3:

1A. - Figure 4D: The authors claim they identify the site of active transcription based on "high intensity in both channels" (p. 7); yet in figure 4D the intensity of the unspliced RNA is given as 0 at the transcription site? This seems a contradiction.

RESPONSE:

We apologize for the lack of clarity and have corrected it in the manuscript. Figure 4D is not an intensity plot but rather a histogram of the distance of mRNAs from sites of transcription. The x-axis is the distance from the transcription site. The y-axis is the frequency, or percentage of mRNAs that

are present within a certain distance away from the transcription site. The following clarifications have been made to the manuscript:

A. Figure caption (page 29, line 19): The word "plot" in the caption was changed to histogram to further clarify the nature of the panel.

B. Figure panel, y-axis: The y-axis was changed from "mRNA fraction of total (%)" to "Percentage of mRNA".

1B. Along the same lines, if the active transcription site is determined by nuclear spots with high intensity in both channels, how can the author distinguish if the mRNA is spliced or unspliced?

RESPONSE:

Again we apologize for the confusion. Our text states that spliced and unspliced mRNA were detected away from the transcription sites, where there is no ambiguity about spot identity. The analysis in Fig. 4D excludes the transcription sites themselves (which are located at "x=0"). We have made the following change in the text to clarify this point.

A. We changed the following sentence in the results section pertaining localization (page 7, line 21) from: "Spatial analysis of the spots revealed that unspliced mRNA is enriched near transcription sites, up to ~6-8um away" to "Spatial analysis of the spots revealed that CAPRIN1 mRNA that has not been alternatively spliced is enriched near transcription sites, up to ~6-8um away, when compared to mature mRNA isoforms." This emphasizes that we are looking at enrichment not in the transcription site, but rather when comparing to mature mRNA isoforms.

B. We changed the following sentence in the spatial localization section of the results from (page 8, lines 1-3): "In addition to pre-mRNA, these incompletely processed transcripts may represent misprocessed mRNA since co-transcriptional splicing is thought to be more efficient." to "In addition to pre-mRNA, non-alternatively spliced transcripts may represent misprocessed mRNA since co-transcriptional splicing is thought to be more efficient. Thus, the imaged transcription sites may contain mature mRNAs in addition to unspliced transcripts."

Additionally, we have made the following changes in the text to further clarify how transcription sites were detected.

A. Although the method for detecting transcription sites is mentioned both in the Materials and Methods and in the caption of figure 3, a reference to the method was lacking in the text. In the revised manuscript, we have added this reference at the appropriate location within the text which describes how transcription sites were detected (page 7, line 17).

B. We also changed the relevant sentence (page 7, line 15-18):

a. Original sentence: "The former were detected as co-localized nuclear spots (which also show isoform 2 for MKNK2), while the latter were identified as nuclear spots with high intensity in both channels (Figure 3)."

b. New sentence: "The former were detected as co-localized nuclear spots (which for MKNK2 also show isoform 2), while the latter were identified as co-localized nuclear spots whose intensity in both channels was substantially higher than that of average mRNAs (Figure 3 and Materials and methods)."

1C. Why is there ~ 40% mature mRNA at the transcription site and almost 0 unspliced mRNA?

RESPONSE:

We are very grateful to Reviewer 3 for identifying this problem. The black line and red line were accidently mislabeled in the original submission. This problem has now been corrected. Now it correctly shows that there is actually ~40% unspliced mRNA and almost 0% mature mRNA very close to the transcription site.

2. The use of ActD as a way to test the authors approach to identifying an active transcript site is problematic since it also affects the amount of the transcripts, and hence changes the entire picture. In order to better achieve this goal one would need to show co-localization of "the nuclear spots with high intensity in both channels" with PolII.

RESPONSE:

It is true that actinomycin D results in an overall decrease of mRNA transcripts. However, the mRNAs in this study are relatively stable and actinomycin D is a fast acting drug that halts transcription. Therefore, after 1hr of actinomycin D treatment there was a very minimal reduction in mRNA abundance, at most a ~10% reduction (Figure S1). This should have a negligible effect on how transcription sites were detected (Figure 3 & Figure 3 caption), as there are still many mRNAs present in the cells. As a reminder, the requirements for identification of an active transcription site were defined as: (1) a nuclear spot, (2) signal in both channels, and (3) substantially higher intensity in each channel compared to the average of all mRNA spots in that channel (at least 1.5 times greater than the average, or the maximum intensity on our 12 bit CCD detector-4096 arbitrary units).

Taken together, the fact that mRNA abundance is only slightly decreased after 1hr, yet transcription site abundance went down by >90% indicates that our method for detection does not pick up false positives. And, it is our experience that co-localization by smFISH coupled to RNA polII immunofluorescence would be technically problematic as smFISH coupled to IF causes a reduction in FISH signal, making FISH intensity quantifications less accurate.

3. - CAPRIN1 and MKNK2 alternative isoforms both differ in their terminal exons. The authors use these two genes to extrapolate to all splicing events by making statements such as "unspliced mRNA is observed away from transcription sites". They have not tested whether other types of splicing events behave in a similar fashion. Generalizations based on these two genes should be toned down.

RESPONSE:

We apologize for the lack of clarity. We did not intend to generalize to all splicing events in the original submission. Thus, we changed the term "unspliced mRNA" to non-alternatively spliced mRNA, and we also specifically mention that we are looking at CAPRIN1 and/or MKNK2 on multiple occasions in the text in order not to generalize to other genes. Changes were made in the abstract, in the results section pertaining to localization, and in the caption of figure 4. Examples of changes are:

A. The last sentence of introduction (page 4, lines 14-15) was changed from "Finally, making use of the spatial resolution of our data, we showed that not all nascent transcripts are alternatively spliced co-transcriptionally." to "Finally, making use of the spatial resolution of our data, we showed that not all nascent CAPRIN1 transcripts are alternatively spliced co-transcriptionally."

B. Sentence in results (page 7, lines 13-15): We changed the following sentence from "Our imaging strategy allowed us to assess this distance by identifying unspliced mRNA and active transcription sites (Figure 4A-C)" to "Our imaging strategy allowed us to assess this distance selectively for CAPRIN1 by identifying non-alternatively spliced mRNA and active transcription sites (Figure 4)."

C. Sentence in results (page 7, line 21): We changed the following sentence from "Spatial analysis of the spots revealed that unspliced mRNA is enriched near transcription sites, up to ~6-8um away." to "Spatial analysis of the spots revealed that CAPRIN1 mRNA that has not been alternatively spliced is enriched near transcription sites, up to ~6-8μm away (Figure 4D)".

D. We changed the title of the relevant section in the discussion (page 7, line 7) from "Unspliced mRNA is observed away from transcription sites" to "CAPRIN1 mRNA alternative splicing in relationship to transcription sites". This emphasizes that we are not looking at all splicing events, but rather we are looking only at the alternative splicing event of our genes.

E. First paragraph in discussion, last sentence (page 14, lines 3-4): We changed the sentence from: "Finally, we used the visual nature of our assay to show that not all nascent mRNA transcripts are alternatively spliced co-transcriptionally." to "Finally, we used the visual nature of our assay to show that not all nascent CAPRIN1 mRNA transcripts are alternatively spliced co-transcriptionally."

F. Figure 4 title and caption (page 29, line 11): The original title of figure 4 was: "Unspliced CAPRIN1 and MKNK2 mRNA is present away from transcription sites." The new title again emphasizes that we are only looking at the alternative splicing event, and not all splicing events: "Non-alternatively spliced CAPRIN1 mRNA is present away from transcription sites". Further the caption of title 4 specifically discusses CAPRIN1 and MKNK2 alternative splicing in specific, and not general terms.

4. - Unspliced mRNA can only be detected in CAPRIN1 since the MKNK2 system is set up in such a way that isoform 1 can be distinguished from isoform 2, but the unspliced RNA can not be detected specifically. Hence, the analysis of distance of the unspliced MKNK2 mRNA from the active transcription site in not valid and should be removed.

RESPONSE:

While it is true that for MKNK2 we can only distinguish between two groups, (A) isoform 1 and (B) isoform 2 together with unspliced mRNA, we notice a strong enrichment of the second group close to transcription sites. This is something that would not be expected, as it is unlikely that the two isoforms would have different spatial localizations within the nucleus. Thus, this suggests that unspliced mRNA is present away from transcription sites.

Nonetheless, as Reviewer 3 suggests, this is not definitive proof. Therefore, in response to Reviewer 3's request, we have removed the quantitative data panel in figure 4 regarding MKNK2 localization and instead describe our observations only qualitatively without drawing conclusions. We only state that we also observed an enrichment of co-localized nuclear spots proximal to transcription sites for MKNK2, and don't interpret that result. Second, we have revised the manuscript to emphasize that our quantification is specific for CAPRIN1.

5. The authors state at various point in the manuscript that their observations suggest that compromising tight regulation of the splicing machinery may be relevant to disease. This is based on the observed differences between a "normal' and a "cancer" cell line. This seems a premature conclusion. Clearly more and better characterized cell lines would need to be tested to make this generalized statement. Please remove these statements.

RESPONSE:

The relevant statements were removed from the text. The two statements that were removed and their corresponding locations are:

A. Abstract (page 2, line 14) "and that this control may be compromised in disease states."

B. Last sentence of discussion (page 16, line 20): ", and may lead to disease states."

6. Knocking down ASF/SF2 increased the FIA variability of MKNK2, but that does not indicate that "the regulatory splicing machinery is misregulated in HeLa cells" (mid page 13).

RESPONSE:

The sentence was toned down accordingly (page 13, line 21). The original sentence was: "We also found evidence that the regulatory splicing machinery is misregulated in HeLa cells." The revised sentence is: "We also found that isoform ratio variability is increased in HeLa cells."

7. The authors conclude that "fluctuations in the regulatory splicing machinery" are responsible for the observed effects. This conclusion is mostly based on a SF2 knockdown experiment. If the effects are indeed due to fluctuations overexpression of SF2 should give the same result. This is a key experiment that should be done and would greatly strengthen the conclusion.

RESPONSE:

We thank the reviewer for the suggestion and agree that is a worthwhile experiment. Unfortunately, we have already attempted to overexpress ASF/SF2 (SFRS1) in the course of this study and found that it is too toxic.

8. I am wondering whether the difference in the size of cell-to-cell variability is related to the difference in isoform ratio between the two genes. CAPRIN1 expresses 95% of one isoform and shows low variability, MKNK2 isoforms are present at similar levels with larger variability. One explanation is that an active mechanism drives CAPRIN1 to a dominant expression pattern thus reducing the variability. This would have less to do with fluctuations of the splicing machinery but dedicated mechanisms. This model should be discussed.

RESPONSE:

We agree that it is certainly plausible possibility that there may an active mechanism that drives the alternative splicing of CAPRIN1 primarily towards isoform 1, i.e. a dominant expression pattern. If this mechanism exists, it would essentially fall under the category of alternative splicing regulation, but perhaps more upstream than the core splicing machinery. Nonetheless, it constitutes a form of regulation of alternative splicing and therefore fluctuations in its control would be considered fluctuations in the regulation of the alternative splicing. Further, such a mechanism would not be able to decrease variability below the binomial isoform partitioning as it would simply set a different average (p) for a binomial distribution. We feel that fluctuations in the regulation of alternative splicing have already been discussed in detail in the manuscript, and that adding this point would add unnecessary complication as it falls within the discussed category.

9. The sequence of figure panels in figures 2 and 3 does not follow the text and is very confusing. I suggest re-structuring the figures to make it correspond to the order the data is discussed in the text.

RESPONSE:

We thank Reviewer 3 for identifying this issue and have corrected it by moving the following sentence, "No bright spots were observedÖ (Figure 2E and 2F)" to an earlier paragraph such that if follows the reference to Figure 2D (page 6, lines 20-21). In the revised version of the manuscript, all of the panels of figure 2 are referenced prior to the first mention of figure 3.

10. The nomenclature for SR proteins has recently been unified. See Krainer and Manley, G&D, 2010. The new nomenclature for ASF/SF2 is SRSF1. Please use throughout manuscript.

RESPONSE:

The text was changed accordingly.

