molecular systems biology

Widespread splicing changes in human brain development and aging

Pavel Mazin, Jieyi Xiong, Xiling Liu1, Zheng Yan, Xiaoyu Zhang, Mingshuang Li, Liu He, Mehmet Somel1, Yuan Yuan, Yi-Ping Phoebe Chen, Na Li, Yuhui Hu, Ning Fu, Zhibin Ning, Rong Zeng, Hongyi Yang, Wei Chen, Mikhail Gelfand and Philipp Khaitovich

Corresponding author: Phillip Khaitovich, Max Planck Institute for Evolutionary Anthropology

Review timeline:

Submission date:	05 March 2012
Editorial Decision:	17 April 2012
Revision received:	16 July 2012
Editorial Decision:	17 August 2012
Revision received:	14 November 2012
Accepted:	16 December 2012

Editor: Andrew Hufton / Thomas Lemberger

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 17 April 2012

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from two of the three referees who agreed to evaluate your manuscript, and we have decided to render a decision now to avoid further delay. As you will see from the reports below, the referees find the topic of your study of potential interest. They raise, however, substantial concerns on your work, which, I am afraid to say, preclude its publication in its present form.

While the reviewers recognized that this work provided an potentially interesting analysis of alternative splicing in human brain development and aging, they raised a series of important concerns which overall were sufficient to reduce their confidence in these findings. One key concern was the effect that demographic differences in the pools might have on the data. Both reviewers felt that additional information was needed regarding ethnicity, sample origins, and cause of death. Moreover, the first reviewer clearly indicated that this issue needed to be investigated experimentally by validating some of the splicing differences in individual samples, and that this could potentially merit a more fundamental restructuring of the dataset to either eliminate potentially biased pools, or to sequence additional sample pools that better balance the dataset. The reviewers also felt that some of the claims made in this work required more direct statistical support and/or deeper analysis.

If you feel you can satisfactorily deal with these points and those listed by the referees, you may wish to 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. A revised manuscript will be once again subject to review and you probably understand that we can give you no guarantee

at this stage that the eventual outcome will be favorable.

PLEASE NOTE As part of the EMBO Publications transparent editorial process initiative (see http://www.nature.com/msb/journal/v6/n1/full/msb201072.html), Molecular Systems Biology now publishes online a Review Process File with each accepted manuscript. 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. Authors may opt out of the transparent process at any stage prior to publication (contact us at msb@embo.org). More information about this initiative is available in our Instructions to Authors.

Referee reports:

Reviewer #2 (Remarks to the Author):

Mazin et al have used a series of techniques (RNASeq, exon arrays and PCR) to address changes in exon inclusion/exclusion in development of the human prefrontal cortex and in aging. The main message is that there are very large scale changes in splicing during development (defined here as the ages 0-14) and more modest changes with aging (from ages 25 onwards). This is similar to the conclusion in a recent paper by Colantuoni et al except the novelty here is in use of different measures of splicing, particularly in the RNASeq approach, combined with some interesting informatic approaches. As such, the paper will be of interest to those working on gene expression, particularly in development. However, some of the aging related effects are less dramatic and might be emphasized a little less.
Some specific points that might be used to improve the manuscript:
1. One of the difficulties is the used of pooled samples in the RNASeq. The authors state that they used this to minimize variance, but also probably were limited due to the cost of RNASeq. However, this does not allow the authors to look at the source of variance - if one sample is an outlier the group mean will be shifted but there is no way to know this without breaking the groups out into individuals. Therefore, in some of the validation steps, the authors must break out the samples into individuals rather than groups and show results from the individuals. This should be performed for splicing assays and gene expression using RT-PCR techniques, although some of the exon array data may be helpful. For example, in 1D/E "two or three samples" were used. All samples should be used and the variance in the ratio on the PCR should be shown.
2. Related to the issues of pooling, this can lead to hidden variables. Looking at the samples in supplementary tables S1 and S2, it is notable that the newborn and young groups are racially mixed but the older group is only Caucasian; given that populations vary in genotype frequencies and there are strong genotype:expression correlations, it seems dangerous to include a mixed design like this where there isn't power to distinguish effects. Two possible solutions are either (a) to limit to a single ethnicity or (b) to include both ethnicities in the grouped design. Either way samples probably need to be excluded or the series expanded.
3. Again related to this; what are the NA samples in supplementary table S2? Also "Caucasian" and "African American" are very broad groupings; where exactly did these samples come from? Perhaps a more appropriate approach would be to genotype the samples and only include samples that are genetically close to each other.
4. Given the lack of an estimate of individual variance in the pools, the authors use a GLM approach with FDR correction to test for significance, and this seems reasonable although it might be challenged. However, this reviewer would be more convinced by genes that show differences in expression in both datasets in both regions (for the first experiment) ie the intersect of the datasets rather than the individual datasets. This is a particular problem in the cerebellum where there is no replication dataset. Please clarify which genes are considered significant in the subsequent analyses, perhaps using a venn diagram of 1B, and consider only using the intersect dataset (which might be 1484 segments in the cortex but maybe fewer if cortex plus cerebellum is used) or provide a strong justification for why not.
5. Why were three mismatches allowed in the RNAseq reads? Given the frequencies of SNPs in the human genome per 100 nt, this seems rather high. What happens in this key parameter is decreased to two or even one mismatch? Please show data.

6. The authors should try and be clearer about how many tests were performed and how many passed validation. In figure 1D/E, is this all PCR products attempted? The text seems to imply that only some PCR products were included; how many and of the exclusions why were they excluded? If some splicing events were not included in the graphs, was it because no splicing was detected and, if so, shouldn't these be shown as IR(PCR)=0? Also, were the PCR products sequenced?
7. Figure 1F is a little hard to follow. If read correctly, 19 genes were tested and of these, some show correlation coefficients between mRNA and protein near 0. How many showed a positive correlation more than chance for 19 regressions, and which genes were these?
8. With this sample series, it is hard to really distinguish aging from development as the power for the former is much less than the latter because of a smaller n and smaller effect size. Looking at figure 2A it is hard to be convinced that there really are age-related changes as the distribution is centered on zero. Please clarify in the text the number of significant age-related changes.
9. Please clarify figure 2B. Are these only significant features and, if so are they those that are significant in development, aging or both? Equally , are these replicated significant between the datasets and brain regions or only seen in one?
10. The amount of intron retention in 2C is quite high; have the authors validated these using RT-PCR? This would seem to be helpful to exclude reads that were mismapped to the transcripts. The authors are also encouraged to show some of the read aligments in a genome viewer.
11. Are the relationships indicated in Figures 1F-K significant? Out of the whole dataset, how many genes show this relationship, ie is there something special about the NMD genes or is this common? It is not discussed, but the same trends are seen in the splicing machinery in figure 3C-H, so this may be a relatively common response.
12. Are the changes in figure 3A significant for aging? Most of them look to be significant during development, but it is hard to see that the right hand portions of the graphs are changing. Please clarify, perhaps changing the text to reflect more about development and less about aging.
13. Figure 4 would benefit from validation with RT-PCR and additional examples other than protocadherin would be helpful.
14. Please also clarify which assays were replicated in the exon arrays, as this did not appear to be found in the manuscript.

Reviewer #3 (Remarks to the Author):

This is a potentially interesting manuscript on n important topic, the importance of alternative splicing on human brain development and maturation. It uses a number of molecular methods including microarrays and RNA Seq to mesure expression in two braion regions, but it is at times hard to follow. there are a number of methodological concerns not the least of which are the samples including who they are, where they are from (which banks as well as the neuroanatomicsal descriptions), inadequate descriptions of race, cause of death and overlap betyween the two groups of subjects (DS1 and DS2). For instance three of the suibjects in DS1 clearly overlap with DS2, but in the latter there ethnicity (actually race) is listed as N.A. in DS2. Although we are assuredthat all brain come from normals, one sublect has died from euthanasia and another from hanging. although it is suggested that there are samples from China there are no Asioans in the databases. In some places Caucasians are listed with a smasll "c" and others with a "C". In addition there is a reference to the Colantuoni study, but it cannot be found in the manuscript. In sdhort, this level of sloppiness makes one less certain about the science in the manuscript. At times there are stayements not backed up with any statistics. perhaps this carelessness is the result of two co-first authors or three corresponding authors. It can and should be fixed. It is important in a study such as this that one knows which bank the specimens are from and whether they havbe used similar dissections.

1st Revision - authors' response 16 July 2012

Response to Referees

Dear Dr Hufton,

Thank you very much for handling our manuscript. In the new version of the manuscript, we have addressed all concerns raised by the Reviewers.

Concerns raised by the Reviewer#2 regarding potential artifacts caused by sample pooling were already addressed in the previous version of our manuscript. Specifically, dataset 2 (DS2) included in our study was based on 13 individual prefrontal cortex samples, rather than pools. The pool-based dataset 1 (DS1) samples and individual-based DS2 samples spanned the same age interval: from birth to 98 years. Furthermore, individual samples included in DS2 covered some of the intermediate age points resulting in more reliable identification of continuous changes in exon inclusion/exclusion with age. All results presented in the original version of our manuscript, with the sole exception of the comparison between two brain regions, were based on a set of splicing changes which had been cross-verified between DS1 and DS2. A selected subset of these changes was further verified using RT-PCR and proteomics data. Furthermore, both the age-related splicing changes, and the comparison of age-related splicing changes between the two brain regions, were cross-verified by Affymetrix exon arrays. We have now conducted additional analysis showing that the DS2 individual samples did not contain any outliers, and results obtained by DS1 and DS2 are broadly consistent. These results exclude any detectable effect of potential outlier samples in the DS1 pools on the conclusions made in this study.

We further conducted additional analyses, based on DS2 splicing profiles, to show that the age-related splicing changes detected in our study are consistent between European and African American individuals. No Asian samples were used in this study; they were all obtained from US and European Brain Banks. To clarify this point we further improved sample information following the comments from Reviewers 2 and 3. We have also made detailed improvements to the manuscript to remove any types of small errors mentioned by the Reviewer#3.

Below please find detailed responses to these and other Reviewer's concerns, as well as a modified version of our manuscript with all changes marked in blue. We hope you will find our manuscript sufficiently improved and suitable for publication in Molecular Systems Biology.

Sincerely,

Philipp Khaitovich, Wei Chen and Mikhail Gelfand

Reviewer #2

Mazin et al have used a series of techniques (RNASeq, exon arrays and PCR) to address changes in exon inclusion/exclusion in development of the human prefrontal cortex and in aging. The main message is that there are very large scale changes in splicing during development (defined here as the ages 0-14) and more modest changes with aging (from ages 25 onwards). This is similar to the conclusion in a recent paper by Colantuoni et al except the novelty here is in use of different measures of splicing, particularly in the RNASeq approach, combined with some interesting informatic approaches. As such, the paper will be of interest to those working on gene expression, particularly in development.

However, some of the aging related effects are less dramatic and might be emphasized a little less. Some specific points that might be used to improve the manuscript:

1. One of the difficulties is the used of pooled samples in the RNASeq. The authors state that they used this to minimize variance, but also probably were limited due to the cost of RNASeq. However, this does not allow the authors to look at the source of variance - if one sample is an outlier the group mean will be shifted but there is no way to know this without breaking the groups out into individuals. Therefore, in some of the validation steps, the authors must break out the samples into individuals rather than groups and show results from the individuals. This should be performed for splicing assays and gene expression using RT-PCR techniques, although some of the exon array data may be helpful. For example, in 1D/E "two or three samples" were used. All samples should be used and the variance in the ratio on the PCR should be shown.

We thank the Reviewer for pointing out this issue. Indeed, the use of pooled samples could result in unforeseen artifacts due to the presence of outlier samples in the pool mixture. To address this issue, dataset 2 (DS2) was included in our study alongside dataset 1 (DS1). Unlike DS1 that was based on pooled samples, DS2 consisted of 13 individual samples that were sequenced separately, without pooling. Importantly, individual samples from DS2 spanned the same age range as DS1 and covered all the major ontogenetic stages present in the DS1: newborns, infants, young adults and old adults, as well as some of intermediate age points. This information is now presented more clearly in the main text (page 4) and in improved Tables S1 and S2.

Following the Reviewer's suggestion, we have now conducted outlier analysis based on splicing variation among individual samples constituting DS2 and pooled samples constituting DS1. Our results indicate the presence of a general effect of age on splicing variation among samples in both datasets, and the absence of outlier samples in DS2 (see Figure R1 below). This result is now included in the main text (page 13) and is shown in the new Supplementary Figure S14.

All splicing variation analysis presented in our study, with the sole exception of the comparison of splicing variation between the two brain regions, is based on results cross-

verified between DS1 and DS2. Given the following observations: (1) absence of samples showing outlier-like behavior in terms of splicing variation in DS2; (2) a high degree of agreement between age-related splicing changes observed in DS1 and DS2 (see Figure 1C for details); as well as (3) independent confirmation of splicing changes with age in a subset of individual samples using RT-PCR and protein measurements (see Figure 1C-F and Table S6 for details), we feel that our results, which are based on the combined analysis of pooled (DS1) and individual (DS2) samples conducted using different RNA-seq protocols, represent a true picture of splicing variation with age in the human brain.

2. Related to the issues of pooling, this can lead to hidden variables. Looking at the samples in supplementary tables S1 and S2, it is notable that the newborn and young groups are racially mixed but the older group is only Caucasian; given that populations vary in genotype frequencies and there are strong genotype:expression correlations, it seems dangerous to include a mixed design like this where there isn't power to distinguish effects. Two possible solutions are either (a) to limit to a single ethnicity or (b) to include both ethnicities in the grouped design. Either way samples probably need to be excluded or the series expanded.

We thank the Reviewer for pointing out this issue. Following the Reviewer's suggestion (a), we have now performed an analysis of splicing variation with age for each ethnicity separately. Here, we took advantage of DS2 measurements, which were conducted in 13 individual samples of different age and contained samples from both Caucasian (N=8) and African American (N=5) origins with overlapping age ranges. From these samples we selected five African American and five Caucasian samples spanning a similar age range (see Methods page 20 for details). Despite a limited number of samples and differences in age distributions, our analysis shows good agreement in direction and rate of age-related splicing changes between the two ethnicities (r=0.75 p<0.0001, see Figure R2 below). We have now added this result to the manuscript's main text (pages 13) and the supplementary information (new Supplementary Figure S15).

3. Again related to this; what are the NA samples in supplementary table S2? Also "Caucasian" and "African American" are very broad groupings; where exactly did these samples come from? Perhaps a more appropriate approach would be to genotype the samples and only include samples that are genetically close to each other.

We thank the Reviewer for pointing out this omission. All brain samples used in our study were obtained from the Brain and Tissue Bank for Developmental Disorders at the University of Maryland, USA and the Netherlands Brain Bank, Amsterdam, Netherlands. We have now completed the sample information table based on information provided by these Institutions (new Supplementary Tables S1 and S2). Given the excellent agreement

of splicing changes with age between the complete dataset that included both Caucasians and African Americans, and the Caucasian-only sample subset (see above), we feel that splicing variation presented in our study is not restricted to a particular ethnicity, but reflects a general ontogenetic trend characteristic of certain brain regions of the human brain.

4. Given the lack of an estimate of individual variance in the pools, the authors use a GLM approach with FDR correction to test for significance, and this seems reasonable although it might be challenged. However, this reviewer would be more convinced by genes that show differences in expression in both datasets in both regions (for the first experiment) ie the intersect of the datasets rather than the individual datasets. This is a particular problem in the cerebellum where there is no replication dataset. Please clarify which genes are considered significant in the subsequent analyses, perhaps using a venn diagram of 1B, and consider only using the intersect dataset (which might be 1484 segments in the cortex but maybe fewer if cortex plus cerebellum is used) or provide a strong justification for why not.

We thank the Reviewer for pointing out this issue. In fact, just as the Reviewer suspected, the set of 1484 segments used in most of our analyses represents an intersection dataset: for these segments splicing changes with age are (a) significant in DS1 both in the prefrontal cortex and in cerebellum, (b) significant in DS2, and (c) consistent between DS1 and DS2. We have now modified our manuscript (page 5) to clarify this.

5. Why were three mismatches allowed in the RNAseq reads? Given the frequencies of SNPs in the human genome per 100 nt, this seems rather high. What happens in this key parameter is decreased to two or even one mismatch? Please show data.

Mismatches in RNA-seq read alignments arise from several sources including (1) SNPs; (2) errors introduced by RNA polymerase during translation and by reverse transcriptase during library preparation; and (3) technical errors during sequencing. Since frequency of errors introduced by the latter two is much greater than the frequency of SNPs, allowing up to three mismatches per 100 nt reads is a standard practice that allows an increase in mapped read numbers without notable compromise with respect to mapping quality.

To test whether allowing mismatches affected our results, we have now restricted our analysis only to the reads that could be mapped to the human genome and splice junctions with no mismatches. In our data, 35% of reads could be mapped with no mismatches in DS1, and 72% in DS2. The low proportion of perfectly mapped reads in DS1 is caused by our requirement that both sequence reads in a pair must be mapped with no mismatches. We find that the age-related splicing changes based on the complete data (allowing mismatches) reported in our manuscript and those changes based on the “no

mismatch" subset of reads show excellent agreement (see Figure R3a and R3b below). We have now added this result to the main text of the manuscript (page 4) and the supplementary information (new Supplementary Figures S2, S3 and S4).

6. The authors should try and be clearer about how many tests were performed and how many passed validation. In figure 1D/E, is this all PCR products attempted? The text seems to imply that only some PCR products were included; how many and of the exclusions why were they excluded? If some splicing events were not included in the graphs, was it because no splicing was detected and, if so, shouldn't these be shown as IR(PCR)=0? Also, were the PCR products sequenced?

We thank the Reviewer for pointing out the lack of clarity in our description of this result. In fact, we originally tested 30 splicing events showing inclusion ratio changes with age using RT-PCR. Out of these 30 splicing events, 24 showed two bands of size expected for the predicted splicing products. These 24 splicing events were used in Figure 1D/E. For all 24 the inclusion ratio changes with age, determined using RNA-seq data, agreed with changes in the relative abundance of the PCR products calculated based on the bands intensity. For the remaining 6 splicing events tested, we did not observe two PCR bands of the expected size. These PCR products could not, therefore, be used to calculate changes in inclusion ratio with age. Some of these failed PCRs may indicate errors in our splicing variant predictions, while others might result from technical failures. To make this point clear, we have now included a more detailed description of the PCR verification experiment, including failed PCRs, into the manuscript's main text (pages 6, 25).

7. Figure 1F is a little hard to follow. If read correctly, 19 genes were tested and of these, some show correlation coefficients between mRNA and protein near 0. How many showed a positive correlation more than chance for 19 regressions, and which genes were these?

We thank the Reviewer for pointing out the lack of clarity in our description of this result. Figure 1F shows that the overall distribution of correlation coefficients between protein and RNA inclusion ratio changes with age is significantly more positive than expected by chance. This general statistical analysis was conducted due to the high technical variation of peptide measurements on the Thermo LTQ platform. Testing the Pearson correlation coefficient significance for inclusion ratio changes with age between protein and RNA data segment by segment, four out of 24 segments from 19 genes tested were marginally significant after correction for multiple testing (Pearson correlation test, followed by Benjamini & Hochberg correction, FDR<0.1). The four segments belong to genes *BIN1*, *DCTN2*, *MAPT* and *PTPRZ1*. The splicing changes in *MAPT* and *PTPRZ1* gene products

were previously implicated in neural function and dysfunctions (see Discussion page 11). We have now added this result to the main text of the manuscript (pages 33 [Figure 1 legend]).

8. With this sample series, it is hard to really distinguish aging from development as the power for the former is much less than the latter because of a smaller n and smaller effect size. Looking at figure 2A it is hard to be convinced that there really are age-related changes as the distribution is centered on zero. Please clarify in the text the number of significant age-related changes.

We thank the Reviewer for pointing out the lack of clarity in our description of this result. Indeed, splicing changes represented on the Figure 2A are based on gene segments showing significant changes in splicing over the entire lifespan, rather than aging. Following the Reviewer's suggestion, we have now determined the gene segments showing significant splicing changes in development and/or during aging by considering these two ontogenetic periods separately. Specifically, for the developmental period, we used 5 samples with ages between 2 days and 25 years and for the aging period – 5 samples with ages between 25 years and 100 years (see Methods page 20). Not all samples were used in the analysis to equalize the numbers of samples assigned to the developmental and aging intervals. Based on these samples, we conducted splicing variation analysis using binomial regression (see Methods page 20 for details). We find that out of 1484 gene segments showing significant splicing changes with age over the entire age span, 970 show significant splicing changes in development and 310 in aging. Out of them, 224 are significant in both periods. When we re-plot Figure 2A using gene only the segments showing significant splicing changes in at least of the two ontogenetic periods, the same result is observed: most of the developmental changes display inclusion ratio decrease with age, while most aging changes display inclusion ratio increase, although of a smaller magnitude (see Figure R4 below). We have now added this result to the main text of the manuscript (page 7) and the supplementary information (new Supplementary Figure S7).

9. Please clarify figure 2B. Are these only significant features and, if so are they those that are significant in development, aging or both? Equally, are these replicated significant between the datasets and brain regions or only seen in one?

We thank the Reviewer for pointing out the lack of clarity in our description of this result. Splicing changes represented in Figure 2B are significant based on splicing variation analysis of the entire age span: from two days to 98 years. These splicing changes are significant in both DS1 and DS2 and consistent between these two datasets. We have now clarified this point in the manuscripts' Methods section (page 21) and the figure

legend (page 34). Further, we have now re-plotted Figure 2B using only gene segments that show significant splicing changes in at least one of the two ontogenetic periods (see Figure R5 below). The same result could be observed: a large fraction of segments follow down-up ontogenetic patterns, with many of them corresponding to intron retention splicing events. We have now added this result to the main text of the manuscript (pages 7) and the supplementary information (new Figure S8).

10. The amount of intron retention in 2C is quite high; have the authors validated these using RT-PCR? This would seem to be helpful to exclude reads that were mismapped to the transcripts. The authors are also encouraged to show some of the read alignments in a genome viewer.

Among the 30 gene segments tested in the RT-PCR experiments, splicing changes in one gene (*UBP1*) corresponded to intron retention events and was successfully confirmed by RT-PCR. We have now added this information to Table S6. Additionally, following the Reviewer's suggestion, we tested whether intron retention events showing significant splicing changes with age, which were identified in our original analysis, are still present after restricting the analysis to sequence reads mapped with no mismatches. We find that the inclusion ratio calculated based on reads mapped without mismatches shows good correlation with the inclusion ratio based on all reads (see Figure R3A below) Visualizing some of the intron retention events in the genome viewer, as suggested by the Reviewer, showed that intron retention events are restricted to certain introns within a gene and are not influenced by exclusion of reads mapped with mismatches (see Figure R6 below for an example of such analysis). We have now added a description of this result to the main text of the manuscript (page 7) and the supplementary information (new Supplementary Figure S3 and S4).

11. Are the relationships indicated in Figures 2F-K significant? Out of the whole dataset, how many genes show this relationship, ie is there something special about the NMD genes or is this common? It is not discussed, but the same trends are seen in the splicing machinery in figure 3C-H, so this may be a relatively common response.

We thank the Reviewer for pointing out the lack of clarity in our results description. As it was mentioned in the original version of our manuscript, 2247 out of 6690 tested genes showed a "down-up" expression pattern (page 22). The fact that five out of the six age-related NMD genes showed a "down-up" expression pattern was not expected to occur by chance (Fisher's exact test, $p<0.02$). This observation, combined with enrichment of intron retention events in the "down-up" pattern (Fisher's exact test, $p<0.0001$), is coherent with the notion that part of the "down-up" segments could be associated with NMD (pages 7-8). In other of our results, splicing factors that were enriched in cluster 1,

representing the "down-up" splicing pattern, also showed significant overrepresentation of the corresponding "down-up" expression patterns (Fisher's exact test, $p<0.001$) (page 9). We have now modified our manuscript to clarify the description of these results.

12. Are the changes in figure 3A significant for aging? Most of them look to be significant during development, but it is hard to see that the right hand portions of the graphs are changing. Please clarify, perhaps changing the text to reflect more about development and less about aging.

We thank the Reviewer for pointing out the lack of clarity in our description of this figure. As described above, following the Reviewer's suggestion we have now determined the significant splicing changes in development and/or in aging separately (see Methods page 20 for details). Out of 1484 gene segments showing significant splicing changes with age over the entire age span, 1422 had detectable inclusion ratio values in samples of all ages. Out of these 1422 segments, shown in Figure 3A, 940 are significant in development and 306 in aging. 223 of them are significant in both ontogenetic periods. We have now modified the main text to include this information (page 7), added numbers labeling the significant gene segments for each or the two ontogenetic intervals to each of the cluster panes on Figure 3A and in the modified legend for figure 3A on page 35.

13. Figure 4 would benefit from validation with RT-PCR and additional examples other than protocadherin would be helpful.

We thank the Reviewer for pointing out this issue. As mentioned in the previous version of the manuscript (page 10), splicing changes with age observed for the protocadherin gamma transcript in the human brain show good agreement with splicing changes previously observed in the mouse brain. Following the Reviewer's suggestion, we have now added two additional examples showing different patterns of splicing changes with age in the two human brain regions, prefrontal cortex and cerebellum. One of the examples shows splicing of the *APP* (amyloid beta (A4) precursor protein) gene, which is involved in Alzheimer's disease (see Figure R7a below), while another shows splicing of the *BIN1* gene, which is involved in synaptic vesicle endocytosis (see Figure R7b below). This result has been added to the manuscript's main text (pages 10) and the supplementary information (new Supplementary Figure S11, S12 and S13). Differences in splicing changes with age between the prefrontal cortex and cerebellum determined using RNA-seq could be generally verified using Affymetrix exon array data (see response to point 14 below).

14. Please also clarify which assays were replicated in the exon arrays, as this did not appear to be found in the manuscript.

We thank the Reviewer for pointing out the lack of clarity in our description. We used an exon array to verify the validity of age-related splicing validation measurements in PFC and CBC conducted in DS1 samples using RNA-seq (page 6, Figure S3). Now, we have also used exon array data to assay the validity of the age-related splicing pattern differences between PFC and CBC that were found using RNA-seq data (Wilcoxon test, $p<0.0001$, see Figure R8 below). We have added this result to the main text of the manuscript (pages 10) and the supplementary information (new Supplementary Figure S11).

Reviewer #3

This is a potentially interesting manuscript on n important topic, the importance of alternative splicing on human brain development and maturation. It uses a number of molecular methods including microarrays and RNA Seq to measure expression in two brain regions, but it is at times hard to follow.

There are a number of methodological concerns not the least of which are the samples including who they are, where they are from (which banks as well as the neuroanatomicsal descriptions), inadequate descriptions of race, cause of death and overlap between the two groups of subjects (DS1 and DS2). For instance three of the subjects in DS1 clearly overlap with DS2, but in the latter there ethnicity (actually race) is listed as N.A. in DS2. Although we are assured that all brain come from normal, one subject has died from euthanasia and another from hanging. Although it is suggested that there are samples from China there are no Asians in the databases. In some places Caucasians are listed with a small "c" and others with a "C".

We thank the Reviewer for pointing out these problems in the sample description. All brain samples used in our study were obtained from NICHD Brain and Tissue Bank for Developmental Disorders at the University of Maryland, USA (NICHD) and the Netherlands Brain Bank, Amsterdam, Netherlands (NBB). We have now updated the sample information table based on information provided by these Institutions and fixed sample description problems that appeared in the original version of our manuscript. These changes are reflected in the new Table S1 and S2 containing sample information for both DS1 and DS2.

In addition there is a reference to the Colantuoni study, but it cannot be found in the manuscript.

We have corrected this error (page 8).

In short, this level of sloppiness makes one less certain about the science in the manuscript. At times there are statements not backed up with any statistics. Perhaps this carelessness is the result of two co-first authors or three corresponding authors. It can and should be fixed. It is important in a study such as this that one knows which bank the specimens are from and whether they have used similar dissections.

We thank the Reviewer for pointing out these problems. We have carefully corrected our manuscript to ensure that there are no other errors and unsupported statements.

2nd Editorial Decision 17 August 2012

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from the two referees who agreed to evaluate your revised manuscript. As you will see from the reports below, the referees felt that the revisions made to this work had improved this manuscript. They have, however, substantial remaining concerns, which, I am afraid to say, must preclude its publication in its present form.

Reviewer #2 felt that certain aspects of this analysis still remaining weakly supported and subject to potentially confounding variables. In particular, s/he noted that the sources of the tissue samples could potentially bias the age-related analysis and results.

Molecular Systems Biology now encourages reviewers to comment on each other's reports, and during this process Reviewer #3 wrote:

"I believe the other reviewer has raised a very important point re the different sources of tissue. There are, however, statistical methods such as surrogate variable analysis that was used in the Colantuoni et al study that could determine to what degrees age and/or tissue source contribute to the variance in the data. This should be done in a revised manuscript."

Overall, both reviewers felt that this work could provide a valuable additional to the literature if these important concerns are convincingly addressed with a combination of new statistical analyses and additional discussion of important caveats. Reviewer #2 felt that a reduction of the age-related claims and analyses may also be needed, especially if additional analysis cannot exclude the possible confounding factors.

As such, we would like to take the exceptional step of offering you a second and final chance to prepare a revised manuscript.

Referee reports

Reviewer #2 (Remarks to the Author):

The revised manuscript by Mazin et al has clarified some of the major problems with the prior version, which were largely due to an opaqueness in the description, particularly around sample collection and characteristics. Having re-evaluated the manuscript, this reviewer is of the opinion that the dataset is likely powered to address some of the problems in the study (eg splicing in human brain development) but by conflating these with both other biological (aging) and some technical (sample sourcing) variables, conclusions about some aspects are limited. The validation by additional techniques and using independent sample series is also limited and the authors should really spend more time on this aspect as it would allow for more robust interpretations of the significant 'hits' from the study. The recommendation is that the manuscript is focussed on consistent, validated changes with development of the human prefrontal cortex, which is where the best data is found. That would be a substantive addition to the literature and of interest to the field. It is acknowledged that this means removing some work that was probably hard to generate and analyze, but a better more robust paper would be much better received and is therefore advisable.

Specific points;
1. Previously it was very difficult to follow the key aspects of the sample series, which is gratefully now corrected with tables S1 and S2. However, this reveals a concern that had previously been hidden, namely that the samples from age 0-23 from series 1 (apart from one sample at 73) are from one brain bank whereas those from 70-98 are from another. SImilarly, sample series 2 are split between brain banks at 0-53 vs 66-98. Therefore, age in the range of 70-98 is confounded by brain bank. Array studies have indicated that sample source is as much a source of variation as age in human brain; therefore there is no correction that can adequately discriminate these two variables. If one considers this in depth, this is a particular problem for the 'down-up' segments, where the 'up' is especially confounded; this analysis is therefore considered especially weak. One solution is to limit

the samples to those obtained from a consistent source; if this is not acceptable then the analysis should be limited to the developmental series and not to aging. The text should be clarified to that end in either case.. Again, previous studies (eg Colantuoni et al,. 2011) show that development has much stronger effects on development than age >20 years so the former are more likely to be robust than the latter.

2. Validation was questioned previously and the authors have responded including the letter to the editor. However, the strength of evidence for validation is quite variable and the authors are strongly encouraged to limit their analysis to fully validated developmental changes (after excluding the aging only and 'down-up' segments. This is probably best discussed in terms of the validation data:

- Verification between brain regions. This is more difficult than the authors discuss, as each is strictly a separate hypothesis and the cerebellum was included only once and then only in the pooled approach. However, now the data has been clarified to be only the intersect of the different datasets, this is acceptable as the cerebellum is effectively a filter for false positives from the frontal cortex. This needs only a short sentence or two for the reader to clarify this point.

- RT-PCR verification. This is quite disappointing - only some changes were validated and the authors don't seem to have made an effort to understand why some of their PCR primers failed. Based on their own data, 24/30 events were validated, which is ~80% rate is not that impressive for an RNASeq dataset. Much more of a concern is that only two or three samples appear to have been sued based on table S7 and it looks like samples were chosen for each gene. This is not enough, and was raised in the prior review- all available samples should have been used to rigorously test the hypotheses, treating each gene as a separate event. It should also be noted that; many of the chosen samples for validated either span a huge age-range (eg DCLK1) and do not distinguish development from aging; that none of the 'down-up' segments are validated; and no independent cohort was used. All these together limit interpretation for what is a straightforward technique.

- Proteomics. The authors admit that these measures have high variance, so their ability to formally test the hypothesis is weak. Furthermore, given that RNA and protein are only partially correlated, this is a weak validation. It is recommended that this data is included but not considered validation.

-Affy exon arrays. Again, the data is not spectacular and is limited to one supplementary figure (S11). If one looks hard at this, the plot shows an excess of positive correlations, but there is still a long tail in the density function of the significant correlations (blue line) between -1 (ie negative correlation, I assume the x axis is mislabeled otherwise it makes no sense) and 0.5. Where is the significance threshold of adjusted $p<0.05$? One would expect it to be ~0.3 or better, in which case a large number of segments fail to be significantly replicated by both techniques.
The authors are therefore suggested to present the validation data in a clearer, more transparent way and consider only techniques that can be considered legtimate tests of the hypothesis.

The other points originally raised are considered adequately dealt with.

Reviewer #3 (Remarks to the Author):

The authors have responded appropriately to suggestions for improving the manuscript. It now looks fine to me.

2nd Revision - authors' response 14 November 2012

Response to Referees

Dear Editor,

As both Reviewers point out, differences in the brain bank origin between the old and the young human brain samples could be a confounding variable complicating the interpretation of the splicing changes found in aging. To address this and other concerns, we have conducted additional analyses and, more importantly, measured and analyzed RNA splicing in an additional sample set composed of 15 macaque prefrontal cortex samples with an age distribution between 0 and 28 years. In contrast to humans, all macaque samples were obtained from the same primate facility where animals were raised in the same environment and were sacrificed in the same way for purposes other than participation in this study. Thus, we believe this additional sample set can be used to address possible confounding effects present in the human data, including the brain bank source, as well as other possible demographic variables.

The comparison between the human and the macaque datasets showed significant and obvious conservation of general age-related splicing patterns observed over the entire postnatal lifespan, as well as conservation of splicing changes found in aging. This result is notable given the existing evolutionary divergence between the human and the macaque species, as well as the fact that the oldest macaque present in our sample set is substantially younger than the oldest human: the age of the oldest macaque sample corresponds to 84 years of human age if corrected for lifespan duration, while the oldest human sample is 98 years old.

Confirmation of the human age-related splicing patterns by the measurements taken in the macaque brains provides a very strong indication that the human age-related splicing patterns described in our manuscript cannot be caused by technical or biological confounding variables. As human splicing patterns can be reproduced by an entirely new dataset in a different species, we believe that this observation also alleviates the Reviewer's concerns regarding confirmation of the RNA-seq results by arrays, PCR and proteomics measurements.

Below please find a point-by-point response to the Reviewer's remarks, as well as a new version of the manuscript's main text with all changes marked in blue. We hope that you will find the new version of our manuscript suitable for publication in Molecular Systems Biology.

Many thanks,

Philipp Khaitovich and Mikhail Gelfand

1. Previously it was very difficult to follow the key aspects of the sample series, which is gratefully now corrected with tables S1 and S2. However, this reveals a concern that had previously been hidden, namely that the samples from age 0-23 from series 1 (apart from one sample at 73) are from one brain bank whereas those from 70-98 are from another. Similarly, sample series 2 are split between brain banks at 0-53 vs 66-98. Therefore, age in the range of 70-98 is confounded by brain bank. Array studies have indicated that sample source is as much a source of variation as age in human brain; therefore there is no correction that can adequately discriminate these two variables. If one considers this in depth, this is a particular problem for the 'down-up' segments, where the 'up' is especially confounded; this analysis is therefore considered especially weak. One solution is to limit the samples to those obtained from a consistent source; if this is not acceptable then the analysis should be limited to the developmental series and not to aging. The text should be clarified to that end in either case. Again, previous studies (e.g. Colantuoni et al., 2011) show that development has much stronger effects on development than age >20 years so the former are more likely to be robust than the latter.

We thank the Reviewer for pointing out this issue. To test the influence of the brain bank source on age-related splicing patterns and, specifically, on splicing changes found in aging we used two approaches. Firstly, we re-analyzed splicing changes with age within the existing human dataset, using samples from each brain bank separately. Secondly, we measured and analyzed RNA splicing in an additional sample set composed of 15 macaque prefrontal cortex samples with an age distribution between 0 and 28 years, using RNA-seq methodology.

With respect to the re-analysis of existing human data:

First, we repeated the analysis of developmental and aging splicing changes using only the 10 samples from DS2, which were obtained from the Brain and Tissue Bank for Developmental Disorders (NICHD) at the University of Maryland, USA. These samples are the youngest among the 13 samples of DS2, with the oldest NICHD derived individual being 53 years old. Still, 1,024 (71%) out of 1,451 age-related segments identified using the full dataset showed consistent splicing change patterns between the NICHD-only set and the full dataset (one-sided Fisher's exact test, $p<0.0001$). Importantly, consistent with our observations made using the full dataset, the "down-up" pattern remained a dominant pattern in the NICHD-only set (one-sided Fisher's exact test, $p<0.0001$) and more segments showed inclusion ratio increase, rather than decrease, during aging (one-sided binominal test, $p<0.01$).

Secondly, we tested the consistency of splicing changes found in aging between the two brain banks. Specifically, we tested the consistency of slopes from the linear regression models based on (a) the two oldest samples from the NICHD set plus the mean value of the other three samples from Netherlands Brain Bank (NBB), and (b) the three NBB samples only. In case splicing changes are caused by differences in tissue source, these two slopes should be independent (**Figure R1, A**), while if splicing changes in aging agree between the two brain banks, the slopes should correlate (**Figure R1, B**). By checking 310 aging-related splicing segments identified using the full DS2 sample set, we found significant positive correlation between the slopes of the two linear models (Spearman correlation, $\rho=0.72$, $p<0.0001$) (**Figure R1, C**). Accordingly, 256 (82.6%) out of 310 segments showed consistent slope directions, which is significantly more than expected by chance (one-sided Fisher's exact test, $p<0.0001$).

With respect to analysis of age-related splicing changes in the macaque dataset:

We measured RNA splicing patterns in prefrontal cortex samples from 15 rhesus macaques with ages distributed between one day and 28 years using RNA-seq (Supplementary Table S12). Importantly, these rhesus macaque samples were obtained from the same primate facility, where animals were raised in the same environment and were sacrificed in the same way for purposes other than participation in this study. Furthermore, all samples were frozen and preserved in the same way prior to the RNA-seq experiment.

Among the 1,484 segments significant in the two human datasets, 496 could be mapped unambiguously to the rhesus macaque genome and had sufficient sequence read coverage (see Methods page 27 for details). Of these 496 segments, 290 segments also showed significant age-related splicing changes in the macaque dataset. For the 496, as well as for 290 segments, we observed significantly greater positive correlations between the trajectories of splicing changes with age in the human and the macaque prefrontal cortex (permutation test, mean $r>0.4$, $p<0.0001$). Further, considering development and aging separately, splicing changes of 321 segments significant in development, as well as 131 segments significant in aging, identified among the 496 segments, were consistent between the human and the macaque datasets for each of these ontogenetic intervals (Fisher's exact test, $p<0.0001$ for development and $p<0.005$ for aging).

We further tested the consistency of the four main splicing patterns observed in the human prefrontal cortex, including the "down-up" splicing pattern, between the human and the macaque datasets, by analyzing each of the four patterns separately. All four main splicing patterns showed significantly greater positive correlations (mean r range 0.2-0.5) between the human and the macaque trajectories of splicing changes with age. This was true for the 496 segments identified in humans that also mapped to macaques (mean r between 0.2 and 0.5; permutation test, $p<0.05$), as well as the 290 segments identified in the human or the macaque time series (mean r between 0.4 and 0.6; permutation test, $p<0.005$), and also the 181 segments identified in humans or macaques that showed consistent slinging changes in human DS1 and DS2 (mean r between 0.4 and 0.6; permutation test, $p<0.005$) (**Figure R2** or Supplementary Figure S9).

Taken together, these results strongly indicate that human age-related splicing patterns described in our manuscript cannot be caused by technical or biological confounding variables in the human dataset. Instead, we find significant conservation of all the major splicing patterns reported in our study between the human and rhesus macaque prefrontal cortices, including the splicing changes found in aging. We have now modified our manuscript to include these analyses (pages 8 and 13, Supplementary Figures S9).

2. Validation was questioned previously and the authors have responded including the letter to the editor. However, the strength of evidence for validation is quite variable and the authors are strongly encouraged to limit their analysis to fully validated developmental changes (after excluding the aging only and 'down-up' segments. This is probably best discussed in terms of the validation data:

- Verification between brain regions. This is more difficult than the authors discuss, as each is strictly a separate hypothesis and the cerebellum was included only once and then only in the pooled approach. However, now the data has been clarified to be only the intersect of the different datasets, this is acceptable as the cerebellum is effectively a filter for false positives from the frontal cortex. This needs only a short sentence or two for the reader to clarify this point.

We thank the Reviewer's for pointing out this omission. We have now modified the main text to include this information (page 15).

- RT-PCR verification. This is quite disappointing - only some changes were validated and the authors don't seem to have made an effort to understand why some of their PCR primers failed. Based on their own data, 24/30 events were validated, which is ~80% rate is not that impressive for an RNA-Seq dataset. Much more of a concern is that only two or three samples appear to have been sued based on table S7 and it looks like samples were chosen for each gene. This is not enough, and was raised in the prior review- all available samples should have been used to rigorously test the hypotheses, treating each gene as a separate event. It should also be noted that; many of the chosen samples for validated either span a huge age-range (eg DCLK1) and do not distinguish development from aging; that none of the 'down-up' segments are validated; and no independent cohort was used. All these together limit interpretation for what is a straightforward technique.

We completely agree with the Reviewer that much more vigorous validation of our RNA-seq results by PCR would be needed if PCR were the sole means of verification. To do so would require a huge number of PCR experiments. Therefore, we used other data types to validate our results. As indicated in our previous response, splicing patterns were cross-validated with two independent RNA-seq datasets, and by comparing splicing changes found in two different brain regions, and this was supported by Exon Arrays and by proteomic data. Certainly, each of these techniques alone does not confirm all the identified age-related splicing changes in their entirety. We note, however, that throughout the manuscript we do not draw conclusions based on specific splicing changes, some of which could still represent false-positives, but rather based on general patterns of splicing changes with age.

To confirm the validity of these general patterns, we have now included an additional analysis of the rhesus macaque time series data described above. The results showed that all four major splicing patterns identified in the human time series, as well as splicing changes taking place in development and in aging when considered separately, can be confirmed in the macaque dataset. Thus, although we completely agree with the Reviewer's point that our PCR results alone are not sufficient to support the full scope of the conclusions drawn by the study, we believe that all validation results considered together along with the addition of a completely independent dataset from another species, do confirm the validity of our conclusions.

- Proteomics. The authors admit that these measures have high variance, so their ability to formally test the hypothesis is weak. Furthermore, given that RNA and protein are only partially correlated, this is a weak validation. It is recommended that this data is included but not considered validation.

We have now modified the main text according to the Reviewer's suggestions (page 12).

- Affy exon arrays. Again, the data is not spectacular and is limited to one supplementary figure (S11). If one looks hard at this, the plot shows an excess of positive correlations, but there is still a long tail in the density function of the significant correlations (blue line) between -1 (ie negative correlation, I assume the x axis is mislabeled otherwise it makes no sense) and 0.5. Where is the significance threshold of adjusted p<0.05? One would expect it to be ~0.3 or better, in which case a large number of segments fail to be significantly replicated by both techniques.

We thank the Reviewer for pointing out the error in labeling Figure S12 (former Figure S11) axis. It has now been fixed. Further, we agree with the Reviewer's point that some segments do not show correlated splicing changes with age between the RNA-seq and Exon Array datasets. We think that this could be caused by the low signal to noise ratio in the Exon Array data. Now, we have used only the segments that show at least moderate (ANOVA, $p<0.1$) age-related splicing changes based on the Exon Array data. Taking the intersection of segments showing significant age-related splicing changes from RNA-seq and on the Exon Array platforms, we now find a much improved positive correlation between the datasets for each of the two brain regions when considered separately (Figure S6), as well as for the splicing pattern comparison between the brain regions (Figure S11) (see **Figures R3 and R4** below).

The authors are therefore suggested to present the validation data in a clearer, more transparent way and consider only techniques that can be considered legitimate tests of the hypothesis.

We have now modified our manuscript according to the Reviewer's suggestions, including additional datasets and analyses, as well as replacing and adding Supplementary Figures and Tables (pages 8 and 13, Supplementary Figures S9 and Tables S7). For a detailed view of manuscript modifications, please see the main text below with all changes marked in blue.

Acceptance letter 16 December 2012

Thank you again for sending us your revised manuscript. We are now satisfied with the modifications made and I am pleased to inform you that your paper has been accepted for publication.

Thank you very much for submitting your work to Molecular Systems Biology.

Reviewer #2 (Remarks to the Author):

The paper is now supplemented with a dataset from macaques, where a subset of the splicing patterns seen in development and aging in the human brain could be mapped to RNA-Seq data. Some overlap reasonably well - the linear patterns of up and down changes across lifespan have a mean r of ~0.6, suggesting these are well preserved across species. But others are more marginal, and it is of note that the mean r value for the 'up-down' and 'down-up' segments are lower (~0.37). I don't see why it would have been difficult to perform a few RT-PCR reactions to validate some candidates. Certainly it would have been quicker, cheaper, and easier to do that than use a whole different species, but that is the authors' decision. It remains the firm view of this reviewer that the developmental changes are robust, but that the signal for aging, independent of development, is weaker and poorly developed. However, I have no specific changes requested at this time.

