BrainSpan alternative splicing

Li, Mingfeng

Alignment

- 607 QCed samples
- STAR mapping with additional parameters compatible with Cufflinks

  --outSAMstrandField intronMotif

  --outFilterIntronMotifs RemoveNoncanonical

Transcript quantification

- Cufflinks generate FPKM per transcript

  excluded chrM reads

  -G, using annotation, not assemble novel transcripts

  -multi-read-correct, do an initial estimation of multiple mapped reads

- RSEM generated FPKM per transcript

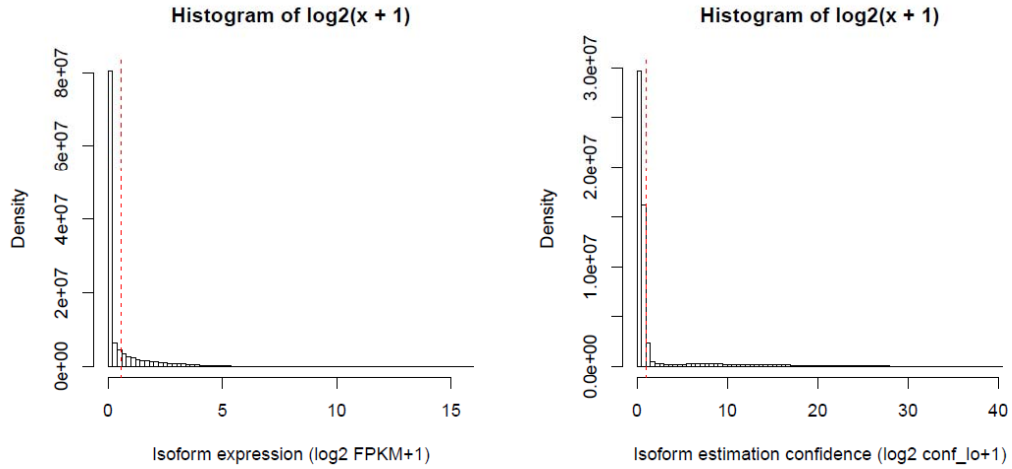
  excluded chrM reads

  with Gencode v21 annotations

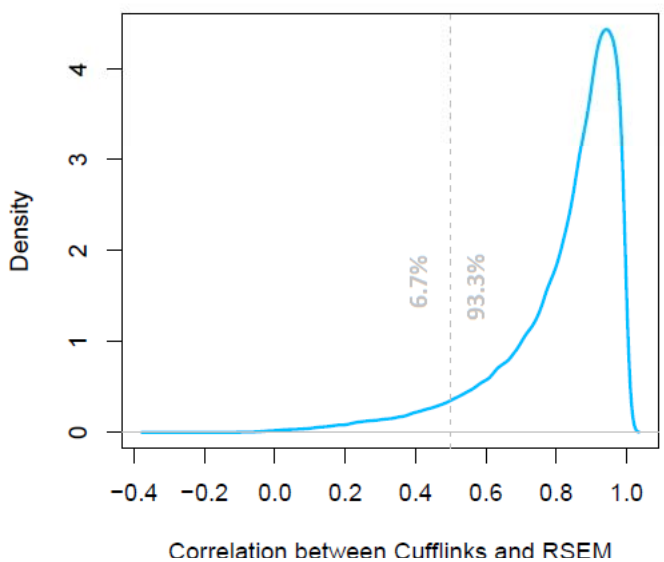
  with multiple mapped reads correction

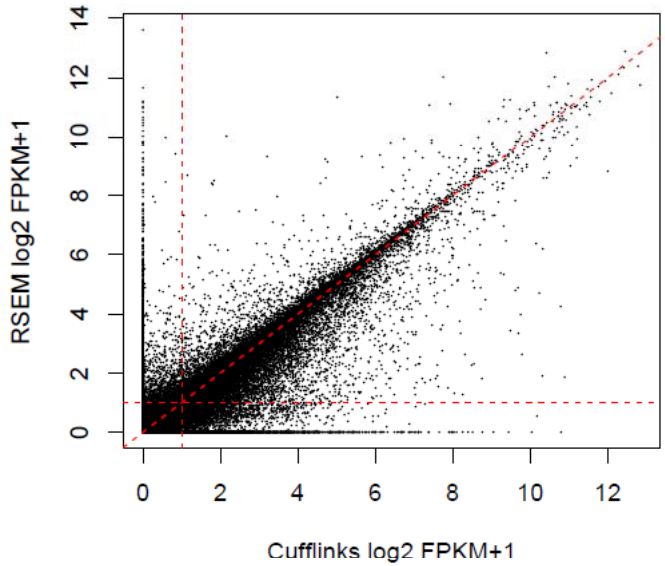
Filter low expression

- At least 5 samples FPKM >=1
- At least 5 samples CI < 1
- These 5 samples should come from more than one brains and more than one period



Correlation between Cufflinks and RSEM

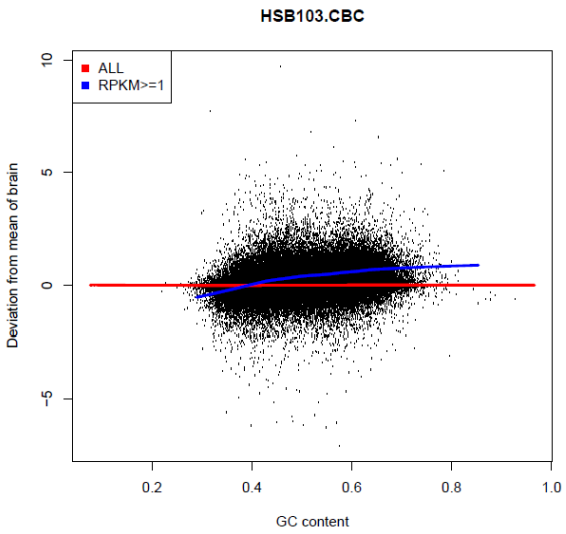




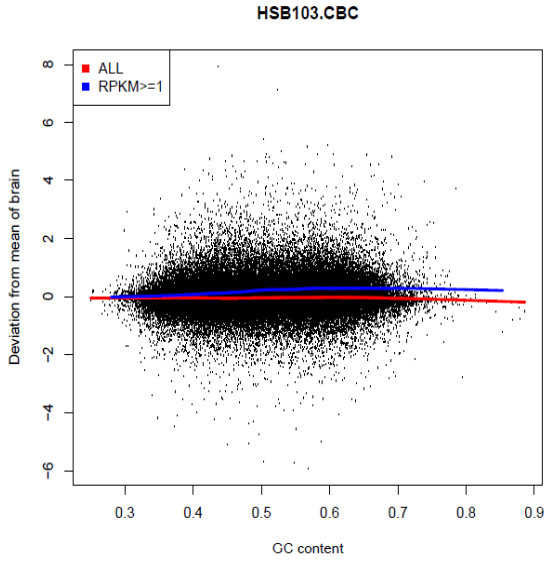
Exclude transcripts with correlation less than 0.5 estimated by two software

Normalization

- CQN correct GC and transcript length
- Combat correct batch



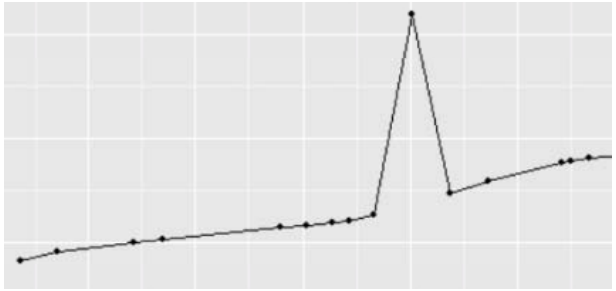
Before normalization



After normalization

More filters

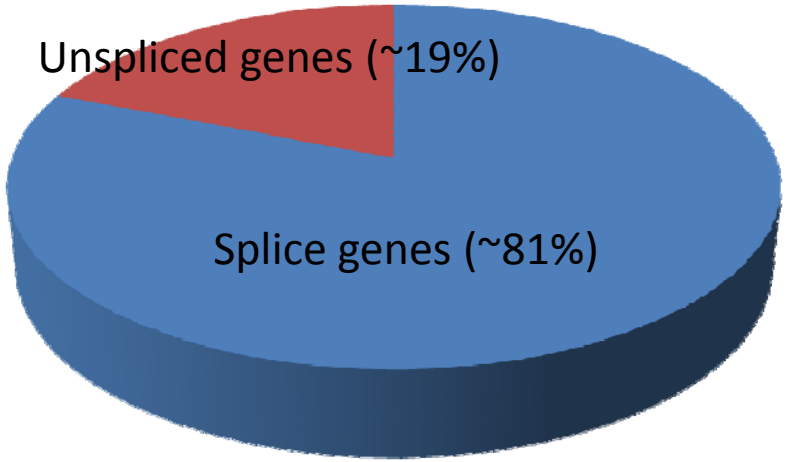
- Using pseudo replicates, more than one sample in each time period, fixing single sample sudden change



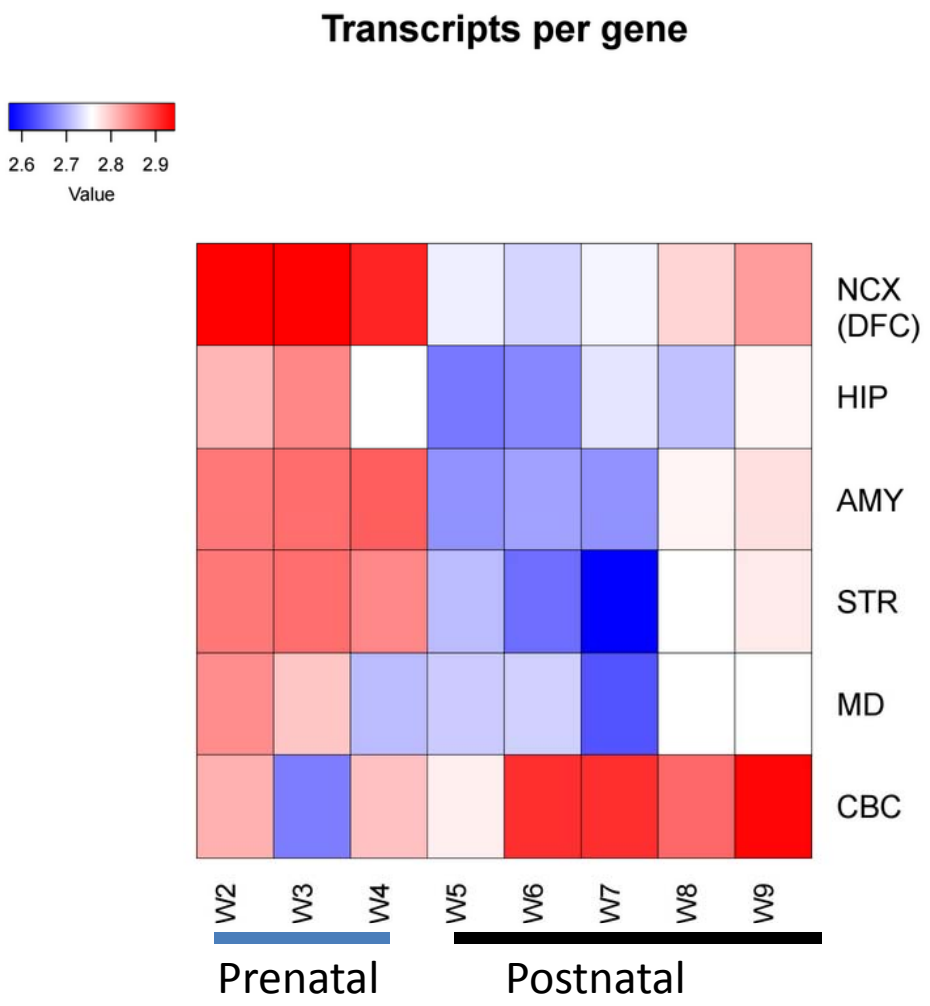
- Using WGCNA, exclude transcripts clustered in grey modules, those suffer from finding co-expressed transcripts and mostly uncertainly inferred transcripts

Summary data

Transcript	Gene (total)	Gene1 (>1 trans)	Gene2 (==1 trans)
61,787	15,128	12,244 (~81%)	2,884(~19%)



Transcripts abundance



Ref: The human transcriptome across tissues and individuals, Science, 8 MAY 2015,V348

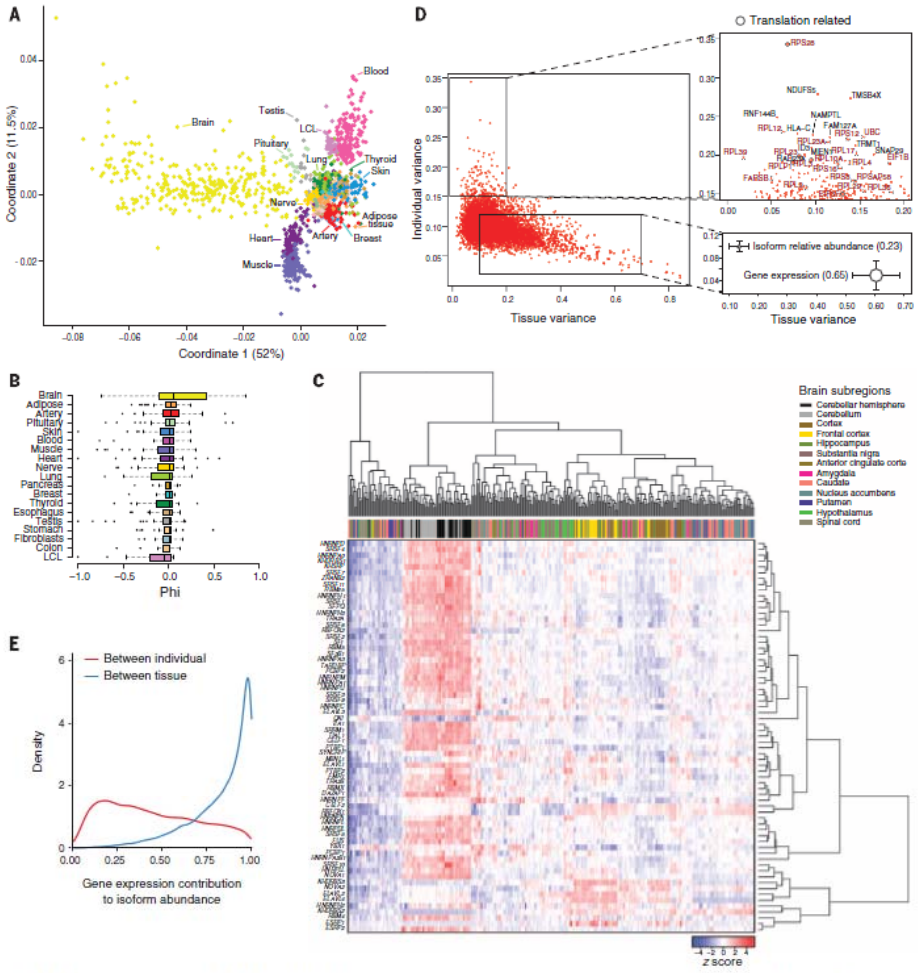


Fig. 3. Splicing across tissue and individuals. (**A**) Multidimensional scaling of all samples on the basis of exon inclusion levels (Percent spliced in, PSI). (**B**) Microexon inclusion across tissues. Values of tissue exon inclusion close to 1 (−1) indicate that the microexon is included (excluded), in nearly all samples from the tissue, and excluded (included) in nearly all samples from the rest of the tissues. Tissues are sorted according to tissue exon inclusion (phi) median value. (**C**) Clustering of brain samples on the basis of the normalized expression levels of 67 RNA binding proteins involved in splicing. The order of samples and genes is obtained by biclustering the expression matrix. (**D**) Left: Contribution of tissue and individual to splicing variation in PCGs. Bottom right: Mean ± SD of individual and tissue contributions to splicing and to gene expression variation. Circle size is proportional to the sum of tissue and individual variation and segment length corresponds to 0.5 SD. Top right: Genes with high splicing variation across individuals. (**E**) Contribution of gene expression to the between-individual and between-tissue variation in isoform abundance

rain (*20*). Furthermore, on of RNA-binding pro- tial and preferential hed in the brain (figs. 5). We found very few l or excluded in a given e S16), 40% of which in the brain. We also <15 bp) are overwhelm- compared to other tissues (Wilcoxon test, $P < 1 \times 10^{-7}$, Fig. 3B). This pattern is not obvious in short exons longer than 15 bp ($P = 0.3$, fig. S21). This observed brain-specific splicing pattern may result from differential splicing in the cerebellum, because expression clustering of the brain regions reveals a general up-regulation of RNA-binding proteins specifically in the cerebellum (Fig. 3C). This is also the brain region exhibiting the largest proportion of novel splicing events (fig. S22).

67 RBP involved in splicing