Analysis of mRNA splicing using RNA-seq – case study:

Widespread splicing changes in human brain development and aging

Mazin *et al.* 2013

Molecular Systems Biology

Background

- mRNA spliced by the spliceosome – a large complex of structural RNA and proteins

- Alternative splicing allows individual genes to encode multiple proteins

- The majority of human genes are alternatively spliced (95% - Wang *et al. Science,* 2008, and others)

- Tissue specific splicing studied extensively (majority of alternative splicing is tissue specific), but not age related differences in splicing

- But alternative splicing associated with age related conditions (Alzheimer's, cancer etc.)

Objective

- Use RNA-seq to assess age-related changes in mRNA splicing (over entire human lifespan) in two brain regions:

    - pre-frontal cortex

    - cerebellum

Experiment

- Illumina sequencing of poly-A transcriptome
- 35 individuals aged from 2 days to 98 years

Two datasets

DS1: paired-end 75bp from 30 individuals; pre-frontal contex & cerebellum

DS2: strand-specific single-end 100bp 13 individuals; pre-frontal contex only

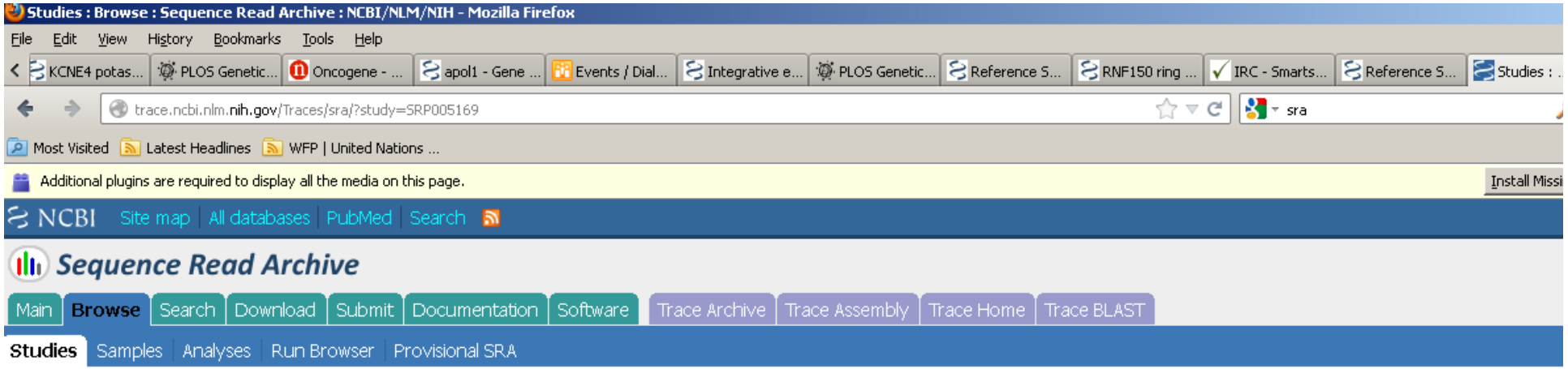
Mean reads per sample: 1.8×10^7 (single lane Genome Analyzer II)

In DS1 individuals were binned into 6 groups with similar ages.

NB: All data available from Short Read Archive (SRA) (SRP005169)

Exon array data also generated: Affymetrix Human Exon 1.0 ST arrays

GEO database (accession GSE26586



SRP005169 Widespread splicing changes in human brain development and aging

Study Type: Transcriptome Analysis

Submission: SRA028456 by on 2011-01-04 22:09:37

Abstract: Differences in transcript splicing are well documented among tissues and between sexes in humans and other organisms1-4. Further, splicing changes in selected genes have been reported in human and mouse brain development5-6. Here, using high-throughput RNA sequencing, we characterized the transcriptome-wide splicing changes occurring over the lifespan in the human brain. We found that in two brain regions, prefrontal cortex and cerebellum, as many as 37% of all expressed protein-coding multi-exon genes undergo significant splicing changes between birth and 98 years of age. Approximately 40% of these changes take place in aging. We confirmed our findings using exon arrays and PCRs and detected identified splicing changes at the protein level. We further associated distinct patterns of age-related splicing changes with expression of the corresponding splicing factors. Thus, splicing plays an important role in shaping the human brain transcriptome in both development and aging.

Description: n/a

Show Entrez docsums for all experiments
Download reads for entire study in sra format (66.
(use Aspera plugin for fast download)

Experiments

Show RUNs for each experiment

Accession	Spots	Bases
Total: 41	1.0 G	113.6 G
SRX037403	16.1 M	2.4 G
SRX037404	15.9 M	2.4 G
SRX037588	17.8 M	2.7 G
SRX044099	21.3 M	2.1 G
SRX045372	21.3 M	2.1 G
SRX045373	20.8 M	2.1 G
SRX045374	23.7 M	2.4 G
SRX045375	23.4 M	2.3 G
SRX045376	22.7 M	2.3 G
SRX045377	23.9 M	2.4 G
SRX045378	17.8 M	1.8 G
SRX045379	19.9 M	2.0 G
SRX045380	23.2 M	2.3 G
SRX045381	16.0 M	1.6 G
SRX045382	20.9 M	2.1 G
SRX045383	21.0 M	2.1 G
SRX045384	20.3 M	2.0 G

Experiment (continued)

Exon array data also generated: Affymetrix Human Exon 1.0 ST arrays

Available from GEO database (accession GSE26586)

Analysis methods

- reads mapped to hg19 using Bowtie 1 (allowing maximum of 3 mismatches)

- reads mapped to a set of artificial junction sequences constructed from consecutive exons in all transcripts annotated by Ensembl

- also constructed artificial junctions from non-consecutive exons (for novel isoform discovery)

- used only uniquely mapped read-pairs

- defined ‘segments’ as regions between successive ‘segmentation sites’ (exon junctions or transcription start/stop)

-calculated ‘inclusion ratio’ for every junction:

$$\text{Inclusion ratio} = \frac{i/(\text{length (segment)} + 2\times\text{length (read)} - 1)}{i/(\text{length (segment)} + 2\times\text{length (read)} - 1) + e/(\text{length (read)} - 1)}$$

i = number of ‘inclusion reads’; *e* = number of ‘exclusion reads’

Analysis methods (continued)

- used generalized linear models (in R) to investigate the relationship between inclusion/exclusion read counts and age/brain region

models $(i_j, e_j) \sim \text{brain} - \text{region} + \text{age} + \text{age}^2 + \text{age} : \text{brain} - \text{region} + \text{age}^2 : \text{brain} - \text{region},$

$$(i_j, e_j) \sim \text{age} + \text{age}^2,$$

- treated the included and excluded reads like binomial success/failure counts

- methods used made available as an R package distributed by the authors

Analysis methods (Exon arrays)

- probes mapped to reference genome using Bowtie and only uniquely mapping probes retained

- background subtraction, normalization and summarization performed using affy package

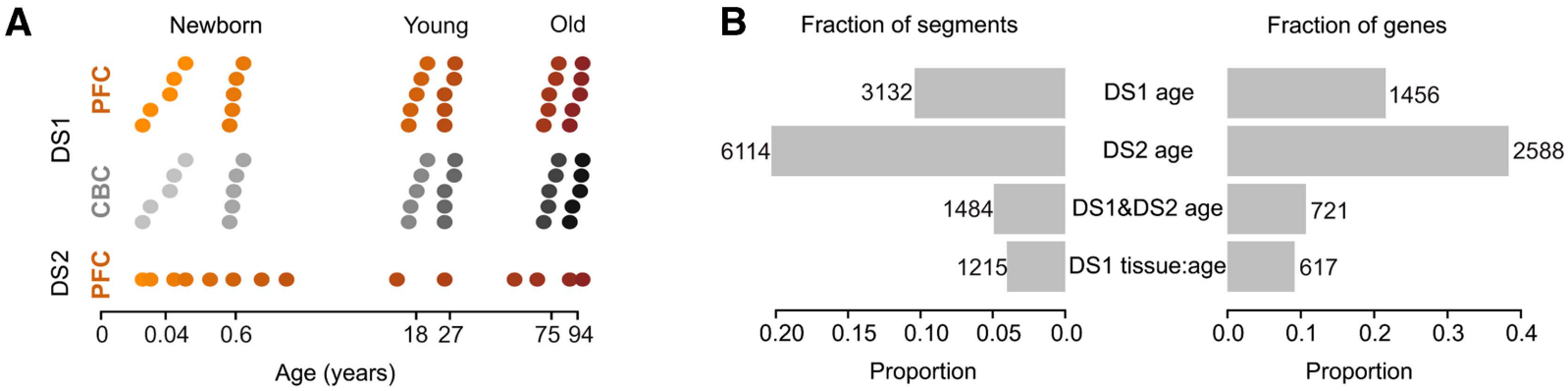
- excluded all probesets with intensity < 100

Analysis methods (continued)

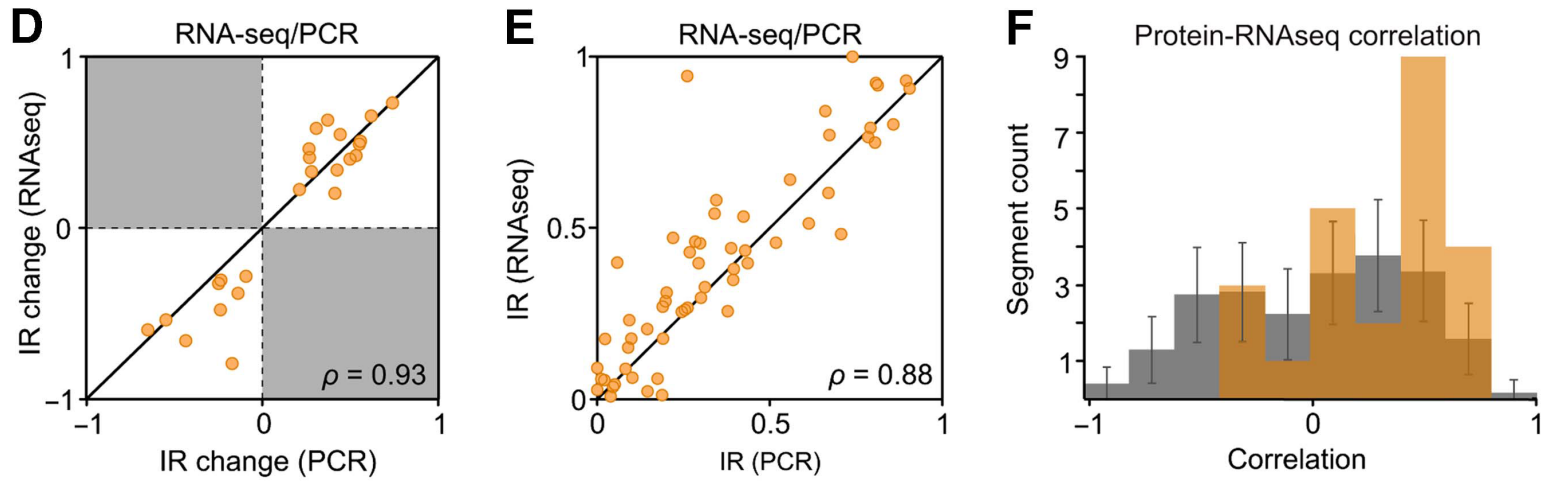
- validated differentially included segments using RT-PCR

- also obtained mass spec data from public domain to compare peptide changes over time to inferred mRNA differential inclusion

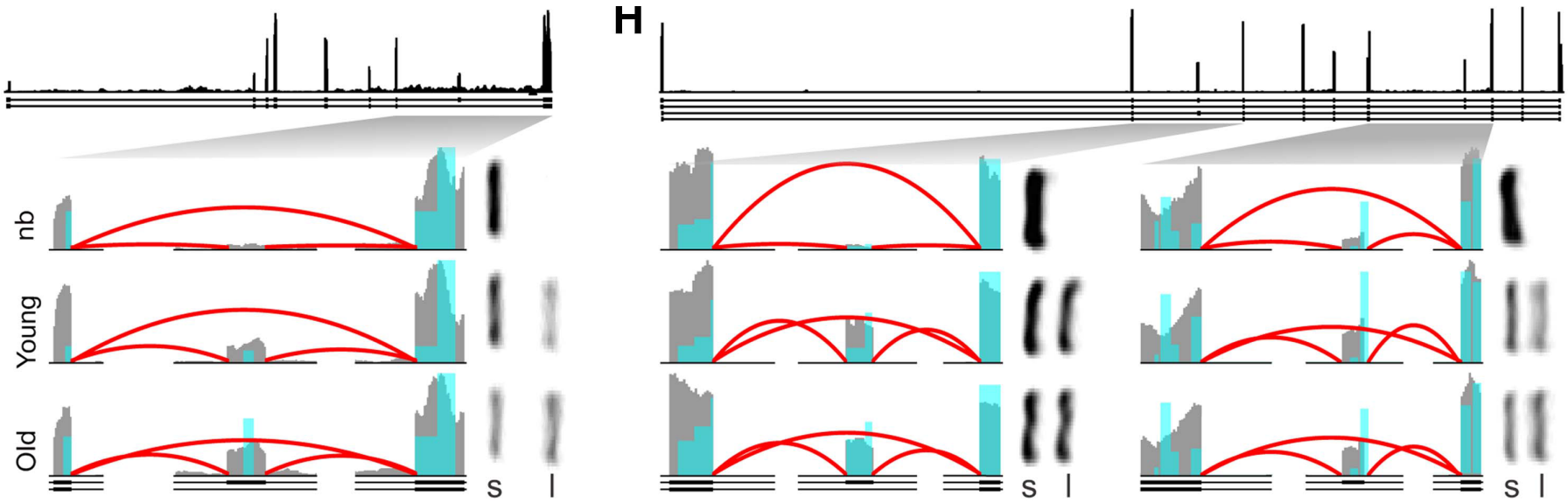
Results



Validation by RT-PCR & consistency with public mass spec data



Examples: *PALM* and *MART*



Grey: RNA; Blue: Protein

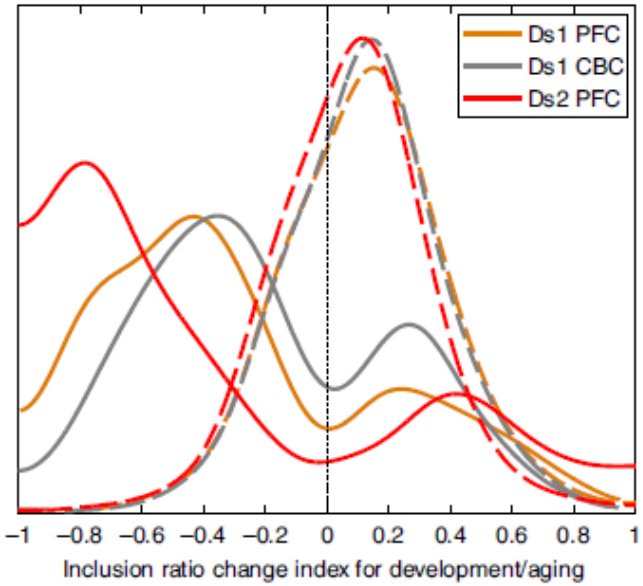
Heights of red arcs represent junction coverage

Results continued

The majority of splicing changes occur in ‘development’ (< 14 years; 970 splice changes) and minority in ‘aging’ (> 25 years; 310 changes)

Development: 75% of differentially included segments show *decreased* inclusion

Aging: 77% show *increased* inclusion

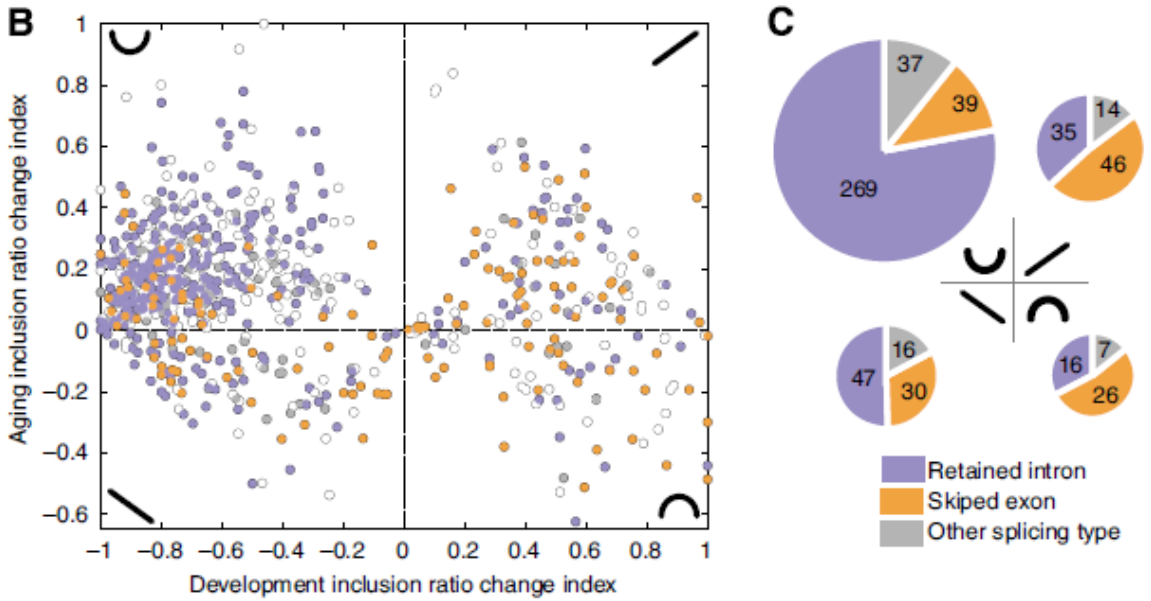


Dashed: aging

Solid: development

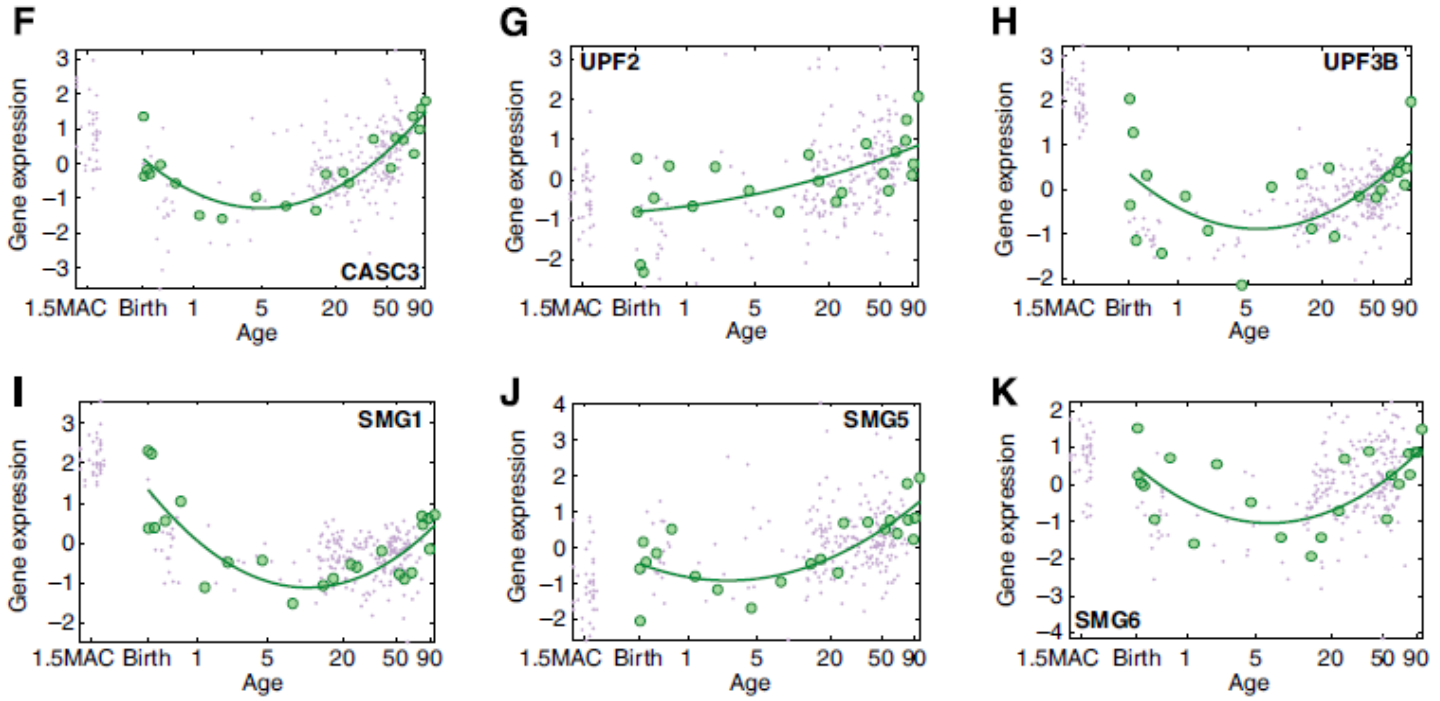
Results continued

Differences in the types of splicing events, depending on pattern of splice change



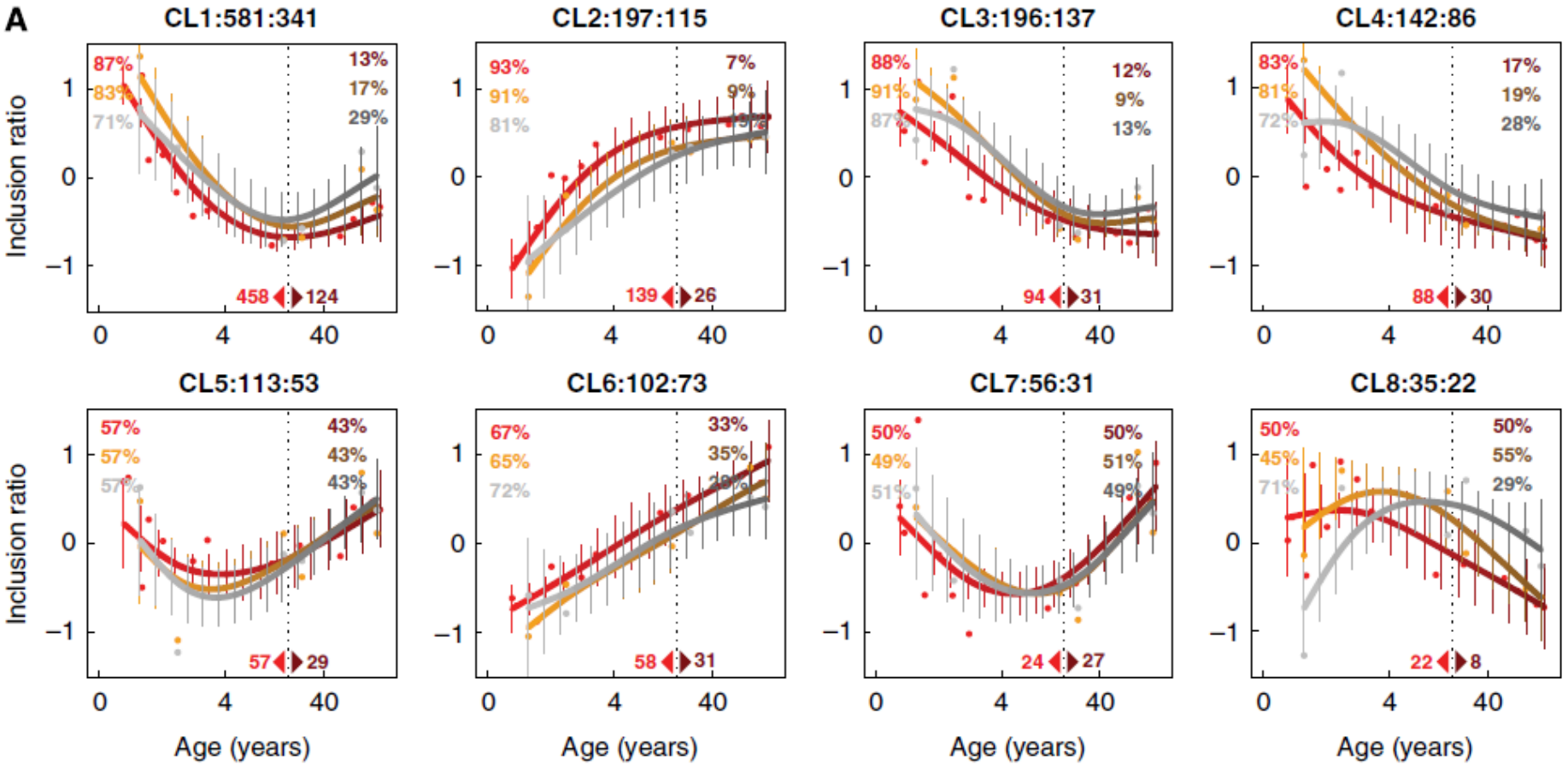
Results continued

Splicing changes displaying ‘down-up’ pattern (down in development; up in aging) enriched for targets of nonsense-mediated decay (NMD). Expression of genes involved in NMD mostly shows similar patterns.



Results continued

Results of hierarchical clustering to identify groups of isoforms showing similar behaviour with age.



Results continued

Results of hierarchical clustering to identify groups of isoforms showing similar behaviour with age.

In each of the clusters:

- performed gene set analysis
- searched for enrichment of sequence motifs predicted to regulate splicing
- determined whether the corresponding splicing factor was over-expressed