Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Which genes' splicing changes with age? A genome-wide analysis of 2.8 million junction tests in GTEx v8

John Hemming

john@hemming.email

Cirrostratus Exedra Ltd https://orcid.org/0000-0002-9520-8730

Research Article

Keywords: ageing, splicing

Posted Date: November 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11620386/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

Background.

Changes in pre-mRNA splicing are a recurrent feature of human aging, but which genes are most and least affected across tissues is not well established. Our earlier analysis of four kynurenine-pathway enzymes found age-associated shifts towards truncated isoforms, most strongly in AFMID. Here we extend the same approach to every gene in the GTEx v8 junction matrix.

Methods.

Junction read counts for 17,382 samples were restructured by tissue, and annotated junctions were grouped into intron clusters. In each of 49 tissues, junction usage was modelled on donor age alone with quasi-binomial regression, fitted simultaneously for all junctions; estimates matched a standard implementation (r = 1.000). False discovery rate (FDR) was controlled across 2.78 million tests. Models adjusted for sex, RNA integrity, ischaemic time and death circumstance were fitted as a secondary analysis.

Results.

Of 910,929 testable tissue × cluster events in 15,503 genes, 107,594 (11.8%) were associated with age at FDR < 0.05, and 10,786 genes (70%) had at least one such event. The share of age-associated events ranged from 1% to 33% between tissues, and was highest where age and ventilator-case status were most tightly linked. Among 9,320 genes testable in at least 20 tissues, splicing regulators were the most consistently age-linked: HNRNPH1, SF1, RBM39, BCLAF1 and LUC7L3 had age-associated splicing in 34–39 of 49 tissues, and regulation of mRNA splicing was the most enriched biological process (7.4-fold). At the other extreme, 584 genes, including COX4I1, DDOST and RPL13, showed no age association in any tissue despite precise estimates. Age shifted splicing away from principal isoforms: in 62% of significant events, the principal-isoform junction lost share. After adjustment, 97.7% of effects kept their direction, at a median 57% of their crude size; 3,662 genes retained an event, enriched for actin-cytoskeleton, muscle and extracellular-matrix functions. AFMID ranked at the 95th percentile genome-wide.

Conclusions.

Age-associated splicing is widespread but concentrated in a few hundred genes, led by the splicing machinery itself, and tends to move transcripts away from their principal isoforms. Much of its tissue-to-tissue variation in GTEx reflects how strongly age is linked to death circumstance, so crude and adjusted estimates should be read together.

1. Introduction

Alternative splicing lets one gene produce several proteins, and its regulation changes with age. In human blood, mRNA processing is among the few pathways associated with age, with shifts in the balance of alternative isoforms [1]. In senescent human endothelial cells, cardiomyocytes and astrocytes, many splicing factors are dysregulated, with downstream changes in isoforms of tissue-relevant genes [2, 3]. Many splicing regulators also control their own expression through unproductive splicing coupled to nonsense-mediated decay [4], so a change in their activity can feed back on their own transcripts.

In a companion analysis of four kynurenine-pathway enzymes in GTEx v8 [9], we found that splicing moves with age towards truncated isoforms, most strongly in AFMID, where skipping of exons 3–10 rose in almost every peripheral tissue. That study could not say whether AFMID is unusual or typical. Here we apply the same method to every gene with annotated junctions in the GTEx v8 junction matrix, to answer three questions:

1. Which genes have splicing most consistently linked to age across tissues, and which are most stable?

2. Is there a general direction to age-associated splicing, for example away from principal isoforms?

3. How much of the age signal is shared with post-mortem and RNA-quality variables that co-vary with age in GTEx [6]?

As in the companion study, the age-only association is the primary result, and confounder-adjusted estimates are reported second.

2. Methods

2.1 Data

The GTEx v8 STAR junction read-count matrix (357,746 junctions × 17,382 samples, 948 donors, 54 tissues) [5] was streamed once from the open-access bucket. Junctions with fewer than 500 reads across all samples were dropped, leaving 313,994. Counts were written to one file per tissue as 16-bit integers; the 2,115 junctions with any per-sample count above 65,535 were stored separately at full precision. Donor age (10-year bins, coded at the midpoint), sex and Hardy scale came from the donor table; RIN and ischaemic time from the sample table. Junctions were annotated against GENCODE v26 [5], the annotation used by GTEx v8. A junction was called **principal** if it is an intron of any transcript tagged APPRIS principal.

2.2 Clusters, filters and events

Within each gene, junctions sharing a donor or an acceptor were joined into intron clusters [7]; single-junction clusters were not tested. Per tissue and cluster, samples needed at least 20 cluster reads and

complete covariates, and at least 80 such samples were required. Junctions with mean usage below 1% or above 99% were skipped. An **event** is one tissue × cluster pair, represented by its junction with the smallest age p-value.

2.3 Models

Junction usage (junction reads / cluster reads) was modelled with a quasi-binomial generalised linear model: logit link, cluster-read weights and Pearson dispersion. Three models were fitted to every testable junction:

Model	Predictors	Role
Age only	age per decade	primary
Adjusted	age, sex, RIN, ischaemic time, Hardy scale (indicators)	secondary
Non-ventilator	age per decade, excluding Hardy 0 donors (≥ 30 samples)	sensitivity

To make 2.78 million fits feasible, iteratively reweighted least squares was implemented in NumPy for batches of 1,500 junctions at once. On lung and whole blood, the estimates matched those from statsmodels for all 35 overlapping kynurenine-pathway tests (correlation 1.000 for coefficients and standard errors; largest coefficient difference 0.002). Benjamini–Hochberg FDR was applied across all junction tests, separately for each model. The full run took about 30 minutes on two CPU cores.

2.4 Gene-level measures

For each gene we counted the tissues in which it was testable and the tissues with at least one event at FDR < 0.05. The main measure of **age linkage** is the fraction of testable tissues with an age-associated event, computed for the 9,320 genes testable in at least 20 tissues. Because statistical power rises with read depth, we also report a coverage-adjusted version: the gene's fraction minus the mean fraction of genes in the same decile of median cluster reads. **Age-stable** genes were those testable in at least 20 tissues with no age-associated event. They were ranked by the largest upper 95% confidence bound of $|\beta|$ across their events, so the top of the list contains genes whose effects are precisely estimated as near zero.

2.5 Enrichment

Gene Ontology biological-process annotations (GOA human, propagated through is_a and part_of relations) were tested for enrichment by one-sided hypergeometric tests against the 9,320-gene background. Terms with 15–500 background genes were tested, with BH-FDR. Three sets were tested: the top 5% by coverage-adjusted age linkage (467 genes), the age-stable genes (584) and genes with adjusted-model events in at least five tissues (132).

3. Results

3.1 Age-associated splicing is widespread

Across 49 tissues, 2,777,394 junction tests formed 910,929 events in 15,503 genes. In the age-only model, 107,594 events (11.8%) were associated with age at FDR < 0.05, and 10,786 genes (70%) had at least one. Among the 9,320 genes testable in at least 20 tissues, the median gene had an age-associated event in 17% of its tissues. Only 584 (6%) had none, and 491 (5%) had one in at least half.

3.2 Tissues differ ten-fold, in step with confounding

The share of age-associated events ranged from 1.1% in cervical spinal cord to 33% in whole blood (Fig. 1). The highest rates were in whole blood, visceral adipose, transverse colon, oesophageal mucosa, skeletal muscle and tibial artery, and the lowest in brain, pituitary, spleen, pancreas and lymphoblastoid cells. Across tissues, the rate rose with sample size (Spearman $\rho = 0.54$). It rose more strongly with the tissue's age–ventilator correlation ($\rho = 0.66$): where young donors were mostly ventilator cases, age effects were common. All 13 brain regions, where only 3–6% of donors were ventilator cases, had rates of 1–10%. After adjustment, every tissue fell to 4.7% or below; tibial artery (4.7%) and skeletal muscle (3.7%) retained the most.

3.3 The most age-linked genes

Age linkage increased with coverage ($\rho = 0.35$): the mean fraction of age-associated tissues rose from 10% in the lowest coverage decile to 27% in the highest. Ranking by the coverage-adjusted score changed little at the top.

Table 1
The 20 most age-linked genes (coverage-adjusted rank; genes testable in ≥ 20 tissues).

Rank	Gene	Function	Tissues with age event / tested	Adjusted: tissues	Median cluster reads
1	HNRNPH1	splicing regulator	39 / 49	4	756
2	VWF	coagulation, endothelium	17 / 24	0	40
3	BCLAF1	splicing / transcription	37 / 49	4	81
4	UPP1	pyrimidine salvage	37 / 49	2	86
5	SF1	branch-point binding, splicing	38 / 49	6	303
6	KLC1	kinesin light chain	38 / 49	4	355
7	TPM1	tropomyosin	38 / 49	13	845
8	ABI3BP	extracellular matrix	28 / 38	6	79
9	MARK3	kinase	36 / 49	4	90
10	ANKRD11	chromatin regulator	34 / 49	0	56
11	RBM39	splicing regulator	37 / 49	8	480
12	MORF4L2	chromatin (NuA4)	37 / 49	1	413
13	NDRG2	stress response	36 / 48	8	1,056
14	SORBS1	adhesion / insulin signalling	32 / 45	12	89
15	KCNMA1	BK potassium channel	29 / 44	7	44
16	TACC1	centrosome	34 / 49	4	68
17	PTPA	PP2A activator	36 / 49	4	488
18	PPP2R5C	PP2A regulatory subunit	36 / 49	8	106
19	EIF4G1	translation initiation	36 / 49	1	164
20	DBNL	actin binding	36 / 49	0	182

Splicing and RNA-processing factors were prominent: HNRNPH1, SF1, RBM39, BCLAF1, LUC7L3, MATR3, RSRC2 and CCNL2, together with the snoRNP protein NOP58, all ranked in the top 1–2%. Gene Ontology enrichment for the top 5% confirmed this (Table 2).

Table 2
Top enriched biological processes (FDR < 0.05) for three gene sets.

Gene set	Biological process	Genes in set / term	Fold	FDR
Most age-linked (467)	regulation of mRNA splicing, via spliceosome	9 / 23	7.4	0.001
	regulation of RNA splicing	13 / 55	4.4	0.001
	muscle structure development	24 / 159	2.8	0.001
	regulation of alternative mRNA splicing	10 / 51	3.7	0.022
	heart development	22 / 182	2.3	0.022
	brain development	26 / 229	2.1	0.022
Robust to adjustment (132)	actin cytoskeleton organisation	16 / 318	3.2	0.006
	muscle system process	9 / 120	4.8	0.006
	extracellular matrix organisation	8 / 99	5.2	0.006
	synapse organisation	8 / 99	5.2	0.006
	regulation of focal adhesion assembly	4 / 23	11.2	0.010
	wound healing	8 / 109	4.7	0.009
Age-stable (584)	negative regulation of stress fibre assembly	7 / 19	6.2	0.049

3.4 Age moves splicing away from principal isoforms

The AFMID pattern generalised. Among the 190,733 junction tests significant for age, junctions of principal isoforms lost share in 61% and non-principal junctions gained share in 62%. At the event level, the principal junction fell with age in 62% of cases (binomial $p < 10^{-300}$). This held in brain (57%) and other tissues (63%), and among events significant after adjustment (56%).

3.5 The least age-linked genes

The 584 age-stable genes were mostly housekeeping genes with deep coverage (Table 3). For the most precisely estimated, any age effect was below about 0.05 log-odds per decade in every tissue. They showed no strong functional theme: only one process (negative regulation of stress fibre assembly) passed FDR < 0.05.

Table 3
The most precisely age-stable genes: no age-associated event in any tissue. Bound is the largest upper 95% confidence limit of |β| (log-odds per decade) across all of the gene's events.

Gene	Function	Tissues tested	Median cluster reads	Bound
DDOST	N-glycosylation (OST complex)	49	630	0.036
PNN	exon-junction complex	49	474	0.039
COX4I1	cytochrome c oxidase	49	2,266	0.045
KIF5B	kinesin heavy chain	49	276	0.056
ISCA2	iron–sulphur cluster assembly	49	107	0.056
CSK	tyrosine kinase	49	96	0.057
JOSD1	deubiquitinase	49	114	0.059
CSNK2A2	casein kinase 2	49	124	0.063
TMEM123	membrane protein	49	158	0.063
RPL13	ribosomal protein	49	6,512	0.085
LSM2	U6 snRNP	49	472	0.090

3.6 Effect of adjustment

Adjustment for sex, RIN, ischaemic time and Hardy scale reduced the number of significant events from 107,594 to 7,048 (6,630 significant in both models). The adjusted coefficient kept the sign of the age-only coefficient in 97.7% of significant events, at a median 57% of its size. Excluding ventilator cases, 88.1% kept their sign. In total, 3,662 genes retained at least one significant event after adjustment.

The genes most robust to adjustment were structural and contractile: TPM1 (13 tissues), SORBS1 (12), PALM, CALD1, RRBP1, TNC and TNNT3 (10 each). This set was enriched for actin cytoskeleton, muscle, extracellular-matrix and synapse functions (Table 2). Splicing regulators kept fewer adjusted events, for example HNRNPH1 in 4 tissues and RBM39 in 8, consistent with part of their crude signal being shared with RNA quality or death circumstance.

3.7 The kynurenine-pathway genes in context

AFMID had an age-associated event in 26 of 49 tissues (53%), placing it at the 96th percentile of all genes testable in at least 20 tissues (95th after coverage adjustment). This is two tissues fewer than in the companion study, which applied FDR across four genes rather than the whole genome. KYAT1 (5 of 47 tissues, 11%) was near the genome-wide median after coverage adjustment (49th percentile). KYNU (3 of 13) and IDO2 (0 of 2) were testable in too few tissues to rank.

GTEx v8 junction counts · 910,929 events in 49 tissues · genome-wide FDR

Brain regions sit at the left, with weak age–ventilator coupling and low rates. EBV-transformed lymphocytes are the exception at the right, probably because cultured cell lines are little affected by the donor's death circumstance.

GTEx v8 junction counts · 9,320 genes testable in ≥ 20 tissues

The distribution is skewed: half of genes have age-associated splicing in fewer than one in six of their tissues, while a tail of about 500 genes is affected in half or more.

4. Discussion

4.1 What the age-only analysis shows

With age as the only predictor, age-associated splicing in GTEx is common but uneven. About one event in nine is associated with age, and most genes have at least one, yet the burden is concentrated. Half of genes are affected in fewer than one in six of their tissues, while about 500 are affected in half or more. The direction is not random: across tissues, the principal isoform's junction tends to lose share and alternative junctions to gain it. That is the same pattern seen in AFMID, and it fits a gradual loss of isoform specificity with age. It also agrees with multi-tissue GTEx work showing that aging alters splicing alongside expression and DNA methylation [8].

4.2 The splicing machinery is itself the most age-linked

The most consistently age-linked genes were splicing regulators, and "regulation of mRNA splicing" was the most enriched process among them. Many splicing factors regulate their own levels through unproductive splicing, in which a poison exon or retained intron sends the transcript to nonsense-mediated decay [4]. Age-associated changes in these genes' own junctions could therefore reflect, or cause, changes in splicing-factor abundance. That would link our findings to reports that splicing-factor expression is dysregulated in aging blood and in senescent cells [1–3]. It also offers a mechanism by which a small number of upstream changes could shift splicing across thousands of downstream genes.

4.3 Confounding is structural in GTEx

The tissue-level results show why crude and adjusted estimates must be read together. In GTEx, 85% of donors in their 20s are ventilator cases against 28% in their 70s, so age and death circumstance are strongly linked in most peripheral tissues. The share of age-associated events was highest exactly where that link was strongest, and lowest in brain, where it is absent. This does not prove that the peripheral signal is artefact. Directions were preserved in 97.7% of events after adjustment, and the brain, which is nearly free of this confounding, still showed age effects in 1–10% of events. But it does mean that the size of crude effects in blood, adipose, colon and lung is probably overstated.

The adjusted model is conservative for the same reason: because the covariates carry information about age, it absorbs some genuine age signal. The adjusted results are best read as a lower bound. On that reading, the genes whose age association survives adjustment in many tissues, led by contractile and cytoskeletal genes such as TPM1, SORBS1, CALD1 and TNNT3, are the most robust findings of this study.

4.4 The least age-linked genes

The age-stable genes are informative as controls. They include core housekeeping genes of respiration (COX4I1), glycosylation (DDOST), translation (RPL13) and splicing itself (LSM2, PNN), with deep coverage and precise estimates near zero. They show that age-associated splicing is not an artefact affecting every expressed gene equally. Their tight bounds also make them candidate reference events for future aging studies of splicing.

4.5 Limitations

• **Coverage dependence.** Power rises with read depth, so breadth of age linkage partly reflects expression level. The coverage-adjusted ranking reduces but does not remove this.

• **Events are not independent.** Junctions within a gene, and genes within a tissue, share samples and donors; gene-level counts are descriptive and not formal tests.

• **Annotated junctions only.** Cryptic splicing, the class most relevant to a loss-of-fidelity model, is not in the public matrix.

• **Binned age, cross-sectional design.** Ten-year bins and a single time point per donor limit resolution and cannot separate aging from cohort effects.

• **Tissue composition.** Bulk RNA cannot distinguish within-cell isoform changes from age-related changes in cell-type mix.

• **Single resource.** Results need replication in cohorts of living donors with continuous age.

4.6 Next steps

The most direct tests would be replication in living-donor cohorts, where death circumstance is absent; annotation-free junction calling on GTEx BAM files to capture cryptic splicing; and testing whether age-associated changes in splicing-factor transcripts predict the downstream events in the same samples.

5. Conclusions

Across 15,503 genes and 49 tissues, splicing changes with age in roughly one event in nine. The changes concentrate in a few hundred genes, led by the splicing regulators themselves, and tend to move transcripts away from principal isoforms. A small group of housekeeping genes is strikingly stable. In GTEx, much of the tissue-to-tissue variation in crude age effects reflects how closely age is

tied to death circumstance; adjustment shrinks effects by about 40% but almost never reverses them. AFMID, highlighted in our pathway study, is among the 5% most age-linked genes genome-wide.

Declarations

Acknowledgements

The Genotype-Tissue Expression Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health and by NCI, NHGRI, NHLBI, NIDA, NIMH and NINDS. Data analysis, figure preparation and drafting were assisted by Claude (Anthropic), an AI system; the author is responsible for the content.

Data availability

All inputs are public: GTEx v8 junction counts and annotations from the GTEx Portal open-access bucket, GENCODE v26 from the GENCODE FTP site, Gene Ontology annotations from GOA and the ontology from the GO repository. The gene-level ranking for all 15,503 genes, tissue rates, enrichment tables and the three analysis scripts are available from the author.

References

1. Harries LW, Hernandez D, Henley W et al (2011) Human aging is characterized by focused changes in gene expression and deregulation of alternative splicing. Aging Cell 10(5):868–878. 10.1111/j.1474-9726.2011.00726.x
2. Latorre E, Pilling LC, Lee BP et al (2018) The VEGFA156b isoform is dysregulated in senescent endothelial cells and may be associated with prevalent and incident coronary heart disease. Clin Sci (Lond) 132(3):313–325. 10.1042/CS20171556
3. Lye JJ, Latorre E, Lee BP et al (2019) Astrocyte senescence may drive alterations in GFAPα, CDKN2A p14ARF, and TAU3 transcript expression and contribute to cognitive decline. *GeroScience* ;41(5):561–573. 10.1007/s11357-019-00100-3
4. Lareau LF, Inada M, Green RE, Wengrod JC, Brenner SE (2007) Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. Nature 446(7138):926–929. 10.1038/nature05676
5. The GTEx Consortium (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 369(6509):1318–1330. 10.1126/science.aaz1776
6. Ferreira PG, Muñoz-Aguirre M, Reverter F et al (2018) The effects of death and post-mortem cold ischemia on human tissue transcriptomes. Nat Commun 9(1):490. 10.1038/s41467-017-02772-x
7. Li YI, Knowles DA, Humphrey J et al (2018) Annotation-free quantification of RNA splicing using LeafCutter. Nat Genet 50(1):151–158. 10.1038/s41588-017-0004-9

8. Ramirez JM, Ribeiro R, Soldatkina O et al (2025) The molecular impact of cigarette smoking resembles aging across tissues. Genome Med 17(1):66. 10.1186/s13073-025-01485-x
9. Hemming J (2026) Age-associated alternative splicing of four kynurenine-pathway enzymes (AFMID, IDO2, KYNU, KYAT1) across human tissues. Unpublished manuscript, Draft

Figures

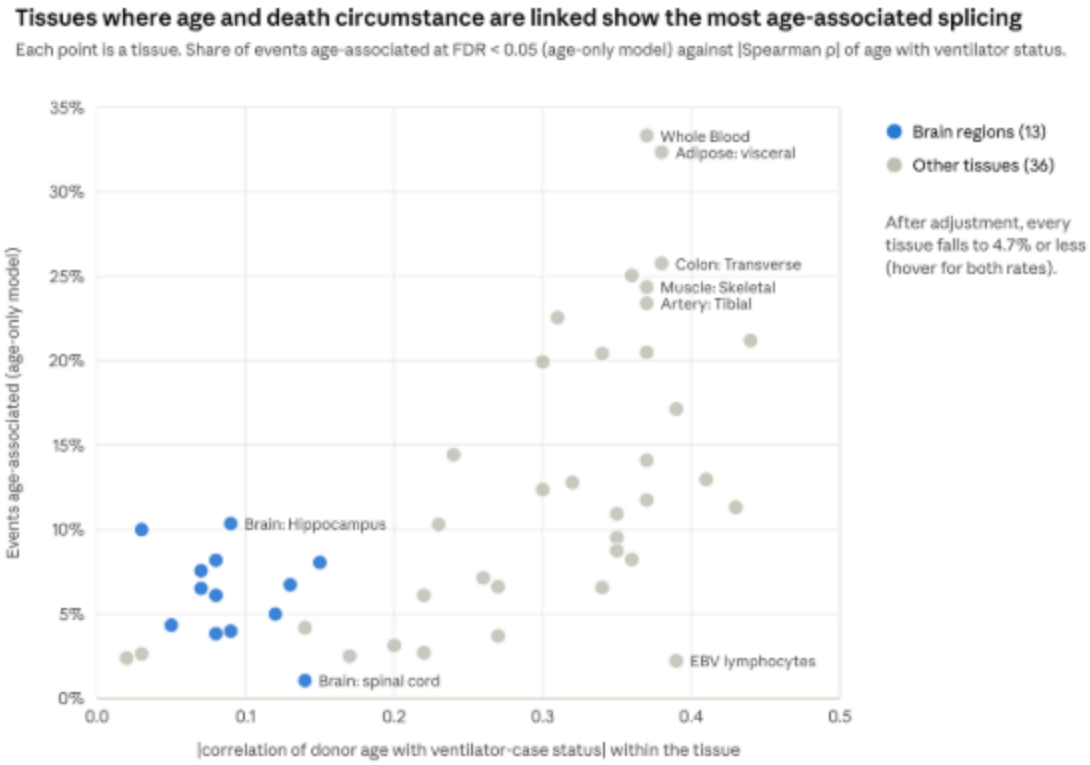


Figure 1

Share of age-associated events per tissue against how strongly age is tied to ventilator-case status in that tissue (Spearman ρ = 0.66 across tissues).

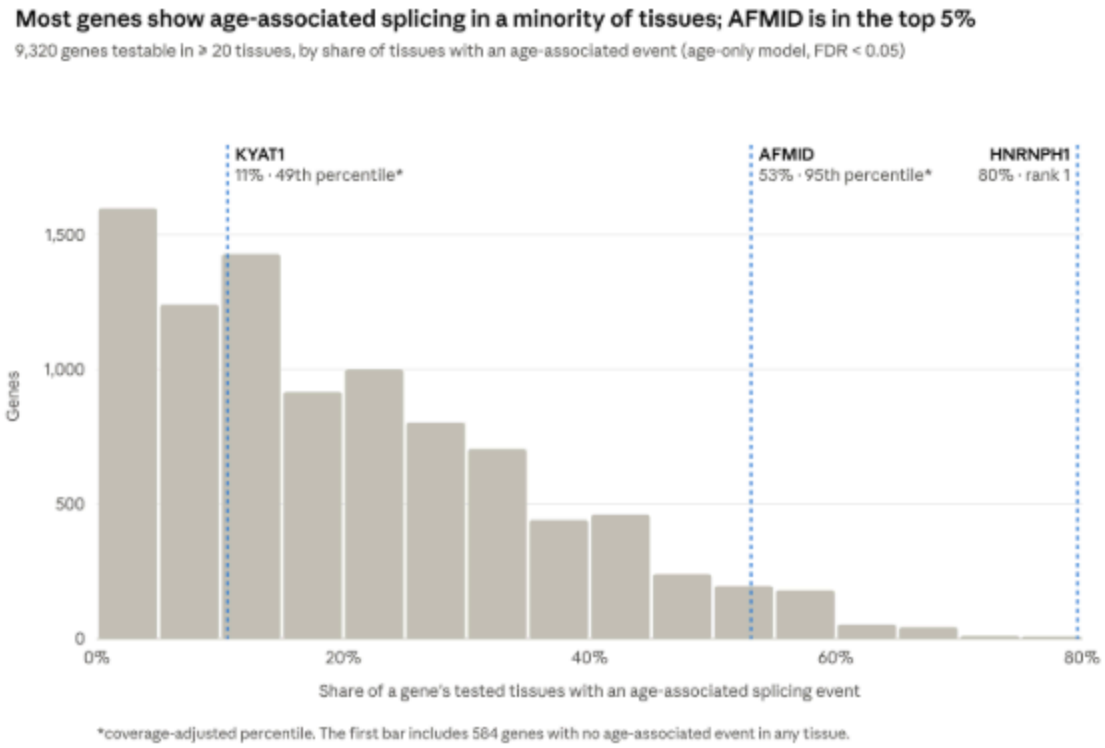


Figure 2

Distribution of gene-level age linkage, with three genes marked.