Methods:

**Construction of initial tissue specific transcriptomes.**

We identified studies with healthy, unperturbed RNA-seq samples from 52 distinct subtissue regions of the body, downloaded and performed quality control on the pertinent sequencing data from the sequence read archive (SRA) using methods from Swamy et al. We aligned each sample to the Gencode V28 hg38 assembly using the genomic aligner STAR\* and sorted the resulting BAM files using samtools sort\*. For each sorted BAM file, we constructed a per sample base transcriptome using stringtie with the Gencode v28 comprehensive annotation as a guiding annotation and default parameters. Next, for each base transcriptome we used stringtie merge to remove transcripts with Transcript-per-Million(TPM) expression level < 1 and minimum isoform fraction < .01. The resulting filtered base transcriptomes were merged the subtissue level using gffcompare\* (default parameters) with the Gencode V28 comprehensive annotation as the reference, to create 52 distinct tissue specific transcriptomes.

**Refinement of tissue-specific transcriptomes by transcript construction frequency**

For each tissue-specific transcriptome, we used the transcript tracking file generated by gffcompare in the previous step to identify the number of samples from which a given transcript was constructed. We kept transcripts that were detected in samples from at least 3 different studies. For subtissues with fewer than 3 different studies, we kept transcripts constructed in at least 3 samples across all available studies. At this stage we also removed novel loci that overlapped any region within 5 kilobases of a known gene by using bedtools intersect. Additionally, we obtained a bed file of known genomic repeat regions from the UCSC table browser and removed any transcripts that overlapped repeat regions.   
**Refinement of tissue-specific transcriptomes by transcript quantification**

For each resulting filtered transcriptome, we extracted transcript sequences using the tool gffread\*, and used these sequences to build a tissue-specific quantification index using the index mode of the alignment free quantification tool Salmon\* using the following parameters: --type quasi --perfectHash --k 31. For each sample, we quantified transcript expression using the quant mode of salmon, using a samples respective tissue specific quantification index with the following runtime parameters: --gcBias --seqBias --numBootstraps 100 --validateMappings. For, for each tissue-specific transcriptome, we used salmon’s bootstrap quantification data to calculate the quantification variance for each transcript. We used this variance calculation to generate a distribution of transcript quantification variance for previously annotated transcripts, and then removed novel transcripts with a quantification variance greater than the 95th percentile of the reference transcript quantification variance. We additionally removed transcripts that had zero counts across all samples. This procedure produced a final set of tissue-specific transcriptomes. These 52 tissue-specific transcriptomes were merged into a single unified transcriptome using gffcompare using --strict-match mode, and DNTX as the transcript prefix.

**Identification of protein coding novel transcripts.**

We identified protein coding transcripts in our unified transcriptome using the TransDecoder\* suite. We extracted transcript sequences using the util script gtf\_genome\_to\_cdna\_fasta.pl and used TransDecoder to find a single best open reading frame from each transcript. Transcripts with no detectable open reading frame were considered noncoding. We used TransDecoder utility script cdna\_alignment\_orf\_to\_genome\_orf.pl to create a gff3 annotation for all protein coding transcripts.

**Calculation of PSI for novel exons.**

We define percent-spliced-in (PSI) as the following:

,

where inclusion reads are all reads spanning an exon’s splice junctions, and exclusion reads are all reads spanning the any exon-exon junction that excludes a given exon. For each sample, we calculated the inclusion and exclusion reads using the tool rMATS\*, using each tissue-specific transcriptome as the annotation for its respective tissues, and reference index built against hg38. We determined total inclusion and exclusion reads for a given exon by summing all inclusion and exclusion reads for all distinct alternative splice events associated with a given exon and used these to calculate PSI.

**Computing Resources**

All computation was performed on the National Institutes of Health cluster compute system Biowulf.

**Code availability**.

To improve reproducibility, we wrote all code used to generate both the data and figures for this paper as Snakemake\* pipelines. All code used for this pipeline is publicly available in the github repository <https://github.com/vinay-swamy/ocular_transcriptomes_pipeline>, and all code to produce these figures is available in the github repository <https://github.com/vinay-swamy/ocular_transcriptomes_paper>

**Figures and Tables**

All statistical analyses, figures and tables were generated using the R programming language. A full list of packages and versions can be found in supplementary file session\_info.txt