* ${\*00\_sample\_metadata}:
  + lab\_sampleid: Sample name used in the lab
  + cell\_line\_id: Cell line ID
  + sample: Sample ID
  + population: Population abbreviation
  + color\_pop: Population color code
  + ooa: Africa / Out of Africa designation
  + color\_ooa: Africa / Out of Africa color code
  + sex\_code: Sex code; 1 = Male 2 = Female
  + color\_sex: Sex color code
  + hapmap\_DNA\_ID: Alternative cell line ID used for this cell line
  + hapmap: Boolean; whether sample was profiled in HapMap
  + GEUVADIS: Boolean; whether sample was profiled in GEUVADIS
  + MAGE: Boolean; whether sample was profiled in MAGE
  + 1000G: Boolean; whether sample was profiled in 1000G
  + ONT1000G: Boolean; whether sample was profiled in Oxford Nanopore 1000G
  + pangenome: Boolean; whether sample was used to build the Human Pangenome
  + GIAB: Boolean; whether sample was profiled in Genome in a Bottle
  + sex: Sex
  + two\_letter\_pop: Alternative 2-letter population abbreviation
  + reads\_fastq7: # reads after preprocessing with Q>7
  + reads\_fastq10: # reads after preprocessing with Q>10
  + map\_reads\_generalmap: # >Q7reads mapping to GRCh38 guided by GENCODE v47 gene annotation
  + map\_reads\_assemblymap: # >Q10reads mapping to GRCh38
  + trizol\_batchA: cell collection and lysis (TRIzol) batch A
  + trizol\_batchB: cell collection and lysis (TRIzol) batch B
  + rna\_extraction\_batchA: RNA extraction batch A
  + rna\_extraction\_batchB: RNA extraction batch B
  + captrap\_batch: CapTrap batch
  + quantification\_id: sample ID disambiguation when it has been sequenced twice
  + libprep\_batch: library preparation and sequencing batch
  + family\_member: wether samples is father, mother or child of a trio, or unrelated to others
  + family: trios family identity code
  + population\_full\_name: Full, unabbreviated name of population
  + population\_description: Description of population
  + extra\_info: Additional notes on cell line
  + coriell\_link: Link to Coriell catalog for this cell line
* ${\*02\_uma\_mt}; this table was used to devise the filters used to obtain the final set of PODER transcripts. Many columns come from SQANTI QC <https://github.com/ConesaLab/SQANTI3/wiki/Understanding-the-output-of-SQANTI3-QC#glossary-of-classification-file-columns-classificationtxt> and therefore many descriptions are taken from there:
  + isoform: Name of transcript
  + associated\_transcriptid.v: ENSEMBL transcript id version of the GENCODE transcript associated to “isoform” through SQANTI QC
  + associated\_geneid.v: ENSEMBL gene id version of the gene containing the GENCODE transcript associated to “isoform” through SQANTI QC
  + flair\_total\_counts: sum of all transcript counts across samples based on FLAIR quantification
  + flair\_mean\_counts: mean of transcript counts across samples based on FLAIR quantification
  + flair\_expressed\_samples: number of samples with non-zero counts based on FLAIR quantification
  + flair\_min\_counts: minimum of transcript counts across samples based on FLAIR quantification
  + flair\_max\_counts: maximum of transcript counts across samples based on FLAIR quantification
  + associated\_transcript\_biotype: GENCODE transcript biotype of GENCODE transcript associated to “isoform” through SQANTI QC
  + associated\_gene\_biotype: GENCODE gene biotype of the gene containing the GENCODE transcript associated to “isoform” through SQANTI QC
  + associated\_gene\_name: gene symbol of the gene containing the GENCODE transcript associated to “isoform” through SQANTI QC
  + associated\_transcripts\_per\_gene: number of different GENCODE transcripts from the same gene associated to any isoform
  + sj\_less\_recountsupported\_novel\_counts: minimum number of RECOUNT3 counts across samples and across novel splice junctions of a transcript
  + sj\_less\_recount\_counts\_novel\_info: novelty category of the splice junction with the least RECOUNT3 counts across samples and across novel splice junctions of a transcript
  + sj\_less\_recountsupported\_novel\_samples: minimum number of samples with at least 10 RECOUNT3 counts across novel splice junctions of a transcript
  + sj\_less\_recount\_samples\_novel\_info: novelty category of the splice junction with the least number of samples with at least 10 RECOUNT3 counts across novel splice junctions of a transcript
  + sj\_category\_full: concatenation of all novelty category of the splice junctions of a transcripts
  + sj\_recount\_counts: concatenation of RECOUNT3 counts across samples of all splice junctions of a transcript
  + sj\_less\_recountsupported\_counts: minimum number of RECOUNT3 counts across samples and across splice junctions of a transcript
  + sj\_less\_recount\_counts\_info: novelty category of the splice junction with the least RECOUNT3 counts across samples and across splice junctions of a transcript
  + sj\_less\_recountsupported\_samples: minimum number of samples with at least 10 RECOUNT3 counts across splice junctions of a transcript
  + sj\_less\_recount\_samples\_info: novelty category of the splice junction with the least number of samples with at least 10 RECOUNT3 counts across splice junctions of a transcript
  + length: Length of isoform in bp
  + exons: # exons
  + structural\_category: SQANTI structural category
  + ref\_length: Length of reference transcript
  + ref\_exons: Number of exons in reference transcript
  + subcategory: Additional splicing subcategorization
  + all\_canonical: All splice junctions use canonical splicing motifs
  + {AJI, CEU, HAC, ITU, LWK, MPC, PEL, YRI} (corresponds to population column in ${\*00\_sample\_metadata}): Boolean; transcript discovered in this population
  + {AJI1, AJI2, AJI3, AJI4, AJI5, AJI6, CEU1, CEU2, CEU3, CEU4, CEU5, HAC1, HAC2, HAC3, HAC4, HAC5, HAC6, ITU1, ITU2, ITU3, ITU4, ITU5, LWK1, LWK2, LWK3, LWK4, LWK5, MPC1, MPC2, MPC3, MPC4, PEL1, PEL2, PEL3, PEL4, PEL5, PEL6, YRI1, YRI2, YRI3, YRI5, YRI6, YRI7} (corresponds to sample column in ${\*00\_sample\_metadata}): Boolean; transcript discovered in this sample
  + population\_sharing: # populations this transcript was discovered in
  + sample\_sharing: # samples this transcript was discovered in
  + tool\_sharing: # different tools this transcript was discovered with
  + contig: Chromosome
  + strand: Strand
  + length: Length of isoform in bp
  + exons: # exons
  + structural\_category: SQANTI structural category
  + espresso: Boolean; transcript discovered with ESPRESSO
  + flair: Boolean; transcript discovered with FLAIR
  + isoquant: Boolean; transcript discovered with IsoQuant
  + lyric: Boolean; transcript discovered with LyRic
  + start: start transcript coordinate
  + end: end transcript coordinate
  + discovered\_transcripts\_per\_gene: number of discovered transcripts per gene
  + old\_associated\_geneid.v: non disambiguated ENSEMBL gene id version of the gene containing the GENCODE transcript associated to “isoform” through SQANTI QC
  + old\_associated\_gene\_biotype: non disambiguated ENSEMBL gene biotype of the gene containing the GENCODE transcript associated to “isoform” through SQANTI QC
  + existsFSMinGene: has any FSM “isoform” been discovered for this gene?
  + existsFSMinTranscript: has any FSM “isoform” been discovered with respect to this transcript?
  + strand: Strand
  + filter: Boolean; transcript passes final PODER filter or not
* ${\*04\_poder\_mt}; many columns come from SQANTI QC <https://github.com/ConesaLab/SQANTI3/wiki/Understanding-the-output-of-SQANTI3-QC#glossary-of-classification-file-columns-classificationtxt> and therefore many descriptions are taken from there:
  + isoform: Name of transcript
  + geneid.v: Name of GENCODE gene or buildLoci gene (for novel genes only)
  + associated\_gene\_biotype: GENCODE biotype of gene
  + proteinv47\_Start: Genomic location of transcript start
  + proteinv47\_Stop: Genomic location of transcript end
  + proteinv47\_CDS\_Source: Source of predicted CDS (CPAT or ORFanage)
  + proteinv47\_CDS\_Start: Genomic location of predicted CDS start
  + proteinv47\_CDS\_Stop: Genomic location of predicted CDS stop
  + proteinv47\_pid: Protein ID from GENCODE of closest protein match from BLASTP
  + proteinv47\_blastp\_identity: BLASTP identity score for predicted amino acid sequence against annotated GENCODE protein translations
  + proteinv47\_blastp\_bitscore: Bitscore from BLASTP of predicted amino acid sequence against annotated GENCODE protein translations
  + proteinv47\_orf\_length\_nt: Length of predicted ORF in bp
  + proteinv47\_protein\_length\_cd: Length of predicted protein sequence in amino acid residues
  + proteinv47\_protein\_is\_nmd: Boolean; whether predicted ORF is predicted to be subject to nonsense mediated decays
  + proteinv47\_protein\_has\_stop\_codon: Boolean; whether predicted ORF has a stop codon
  + proteinv47\_protein\_has\_start\_codon: Boolean; whether predicted ORF has a start codon
  + proteinv47\_protein\_sequence: Predicted amino acid sequence
  + trunc\_annot\_aa: For novel transcripts of known protein coding genes, whether the predicted amino acid sequence represents a truncation of an annotated amino acid sequence from the same gene in GENCODE v47
  + elong\_annot\_aa: For novel transcripts of known protein coding genes, whether the predicted amino acid sequence represents an elongation of an annotated amino acid sequence from the same gene in GENCODE v47
  + annot\_aa: For novel transcripts of known protein coding genes, whether the predicted amino acid sequence is an exact match for an annotated amino acid sequence from the same gene in GENCODE v47
  + aa\_novelty: For novel transcripts of known protein coding genes, novelty as indicated in the main figure (NMD>Known>Known truncation>Known elongation>Novel)
  + espresso: Boolean; transcript discovered with ESPRESSO
  + flair: Boolean; transcript discovered with FLAIR
  + isoquant: Boolean; transcript discovered with IsoQuant
  + lyric: Boolean; transcript discovered with LyRic
  + {AJI, CEU, HAC, ITU, LWK, MPC, PEL, YRI} (corresponds to population column in ${\*00\_sample\_metadata}): Boolean; transcript discovered in this population
  + {AJI1, AJI2, AJI3, AJI4, AJI5, AJI6, CEU1, CEU2, CEU3, CEU4, CEU5, HAC1, HAC2, HAC3, HAC4, HAC5, HAC6, ITU1, ITU2, ITU3, ITU4, ITU5, LWK1, LWK2, LWK3, LWK4, LWK5, MPC1, MPC2, MPC3, MPC4, PEL1, PEL2, PEL3, PEL4, PEL5, PEL6, YRI1, YRI2, YRI3, YRI5, YRI6, YRI7} (corresponds to sample column in ${\*00\_sample\_metadata}): Boolean; transcript discovered in this sample
  + population\_sharing: # populations this transcript was discovered in
  + sample\_sharing: # samples this transcript was discovered in
  + tool\_sharing: # different tools this transcript was discovered with
  + chrom: Chromosome
  + strand: Strand
  + length: Length of isoform in bp
  + exons: # exons
  + structural\_category: SQANTI structural category
  + associated\_gene: Reference gene
  + associated\_transcript: Reference transcript
  + ref\_length: Length of reference transcript
  + ref\_exons: Number of exons in reference transcript
  + diff\_to\_TSS: Distance of 5’ end of transcript to reference TSS
  + diff\_to\_TTS: Distance of 3’ end of transcript to reference TTS
  + diff\_to\_gene\_TSS: Distance of 5’ end of transcript to closest TSS annotated to the same gene
  + diff\_to\_gene\_TTS: Distance of 3’ end of transcript to closest TTS annotated to the same gene
  + subcategory: Additional splicing subcategorization
  + RTS\_stage: Demonstrates evidence of at least one possible RT artifact
  + all\_canonical: All splice junctions use canonical splicing motifs
  + bite: Boolean; whether transcript contains at least one "bite" positive SJ
  + FSM\_class: This feature classifies the transcript according to the expression of other isoforms in the gene to which the transcript belongs. Transcripts belonging to genes that only express one isoform are classified as A. Transcripts belonging to genes that express more than one isoform but none is a FSM are classified as B. Transcripts belonging to genes which express more than one isoform and other isoforms and at least one is a FSM are classified as C.
  + perc\_A\_downstream\_TTS: % of genomic "A"s in the downstream 20 bp window. If this number if high (say > 0.8), the 3' end site of this isoform is probably not reliable
  + seq\_A\_downstream\_TTS: Sequence of the downstream 20 bp window
  + pop\_spec\_t: Boolean; whether transcript is population-specific discovered (>=2 samples of same population but not discovered in any other population)
  + tau: Tau value of transcript. Transcripts are required to pass minimal reproducibility and expression thresholds described in the methods
* ${\*06\_poder\_t\_counts}:
  + rows: one row for each isoform
  + columns: one column for each sample (using identifier of “cell\_line\_id”\_1)
  + values: counts per sample / isoform pair
* ${\*07\_poder\_g\_counts}:
  + rows: one row for each gene
  + columns: one column for each sample (using the "sample" identifier)
  + values: counts per sample / gene pair
* ${\*08\_mage\_t\_counts\_poder}, ${\*09\_mage\_t\_counts\_gencode}, ${\*10\_mage\_t\_counts\_enh}:
  + rows: one row for each isoform
  + columns: one column for each MAGE sample
  + values: counts per sample / isoform pair
* ${\*11\_astu}:
  + sample: Sample ID
  + population: Population abbreviation
  + map\_reads\_assemblymap: # >Q10reads mapping to GRCh38
  + gene\_variant: Gene + variant pair being tested (gene\_chromosome\_position\_ref\_alt)
  + statistic: chi-squared statistic
  + p.value: P-value from testing this gene + variant pair
  + FDR: Corrected p-value from testing this gene + variant pair
  + geneid.v: Gene ID
  + gene\_biotype: GENCODE gene biotype of the gene containing the GENCODE transcript
  + variant: Variant ID (chromosome\_position\_ref\_alt)
  + transcriptid.v: Transcript ID
  + refCount: Reference allele count for this transcript
  + altCount: Alternative allele count for this transcript
  + Count: Total number of counts for this transcript
  + transcript\_variant: Transcript + variant pair (transcript\_chromosome\_position\_ref\_alt)
  + significant: Whether / how gene + variant pair passes significance threshold
  + tested\_genes: Number of testable genes
  + significant\_genes: Number of significant genes
  + trx\_tencounts: Boolean; whether this transcript has at least 10 counts in this sample
  + gene\_numtrx: Number of transcripts belonging to this gene
  + gene\_twentycount: Boolean; whether this transcript has at least 20 counts in this sample
  + gene\_heterozygous: Boolean; whether this gene has a heterozygous SNP
  + gene\_testable: Boolean; whether this gene was testable for ASTU
  + annot: Either PODER, GENCODE, or Enhanced (Enhanced GENCODE) for where the gene + variant pair was tested
* ${\*12\_ase}:
  + annot: Either PODER, GENCODE, or Enhanced (Enhanced GENCODE) for where the gene + variant pair was tested
  + contig: Chromosome
  + position: Variant position
  + variant: Variant ID (chromosome\_position\_ref\_alt)
  + refAllele: Reference allele
  + altAllele: Alternative allele
  + refCount: Reference allele count for this gene
  + altCount: Alternative allele count for this gene
  + totalCount: Total number of counts for this gene
  + GENOTYPE: Genotype of individual (always heterozygous)
  + geneid.v: Gene ID
  + p.value: P-value from testing this gene + variant pair
  + FDR: Corrected p-value from testing this gene + variant pair
  + tested\_genes: Number of testable genes
  + significant\_genes: Number of significant genes
  + cell\_line\_id: Cell line ID
  + sample: Sample ID
  + population: Population abbreviation
* ${\*13\_mage\_tau}:
  + isoform: isoform ID
  + tau: Tau value of population-specificity computed in the MAGE RNA-seq dataset.
* ${\*14\_gwas\_enrichments}
  + ID: ID of GWAS trait
  + Description: More in-depth description of GWAS trait
  + GeneRatio: Number of ASTU genes over number of total GWAS genes in this trait
  + BgRatio, RichFactor, FoldEnrichment, zScore, pvalue, p.adjust, qvalue: ClusterProfiler: EnrichR output
  + geneID: Gene ID
  + Count: Number of genes found in this trait
  + annot: Either PODER, GENCODE, or Enhanced (Enhanced GENCODE) for where the gene + variant pair was tested
* ${\*16\_inter\_catalog\_overlap}
  + Strand: Strand of transcript
  + Coordinates: Intron chain coordinates of transcript (hyphen-separated)
  + Chromosome: Chromosome of transcript
  + CHESS3: Boolean; whether this transcript is in CHESS3
  + ENCODE4: Boolean; whether this transcript is in the ENCODE4 LR-RNA-seq transcripts
  + GENCODE v47: Boolean; whether this transcript is in GENCODE v47
  + GTEx: Boolean; whether this transcript is in GTEx LR-RNA-seq transcripts
  + PODER: Boolean; whether this transcript is in PODER
  + RefSeq v110: Boolean; whether this transcript is in RefSeq v110
* ${\*17\_personalized\_hg38}:
  + ic\_id: Unique identifier for intron chain (chromosome\_strand\_hyphen-separated coordinates)
  + structural\_category: SQANTI structural category
  + cell\_line\_id: Cell line ID
  + hap1: Boolean; whether this transcript is detected using personalized-GRCh38 1
  + hap2: Boolean; whether this transcript is detected using personalized-GRCh38 2
  + hg38: Boolean; whether this transcript is detected using standard GRCh38