**Supplementary Figures**

Figure S1: Relative abundance of the subset of transcripts in each position of the ranking for all datasets, including different expression thresholds. Only genes with more than one transcript annotated are represented here.

Figure S2: Relative abundance of the subset of transcripts in each position of the ranking for the BM dataset, separating genes by the number of annotated transcripts.

Figure S3: Relative abundance of the subset of transcripts in each position of the ranking for all datasets, including different expression thresholds, when grouping transcripts by TSS. Only TSS with more than one transcript annotated are included here.

Figure S4: Ratio of expression levels between the major transcript and the second most abundant one for all datasets, relative to the number of annotated transcripts and including different expression thresholds.

Figure S5: Proportion of dominant vs non-dominant major transcripts detected in the BM dataset for different gene expression thresholds and taking into account different dominance degrees.

Figure S6: Relative abundance of the subset of transcripts in each position of the ranking for the simulated datasets, including different sequencing depths. Major transcript abundance is underestimated by the analysis pipeline used here, thus suggesting that it is not biased towards the identification of a single transcript per gene.

Figure S7: Length distribution for major transcripts, relative to the longest transcript annotated for each gene.

Figure S8: Example of a non-canonical major transcript: CD47 (CD47 molecule, ENSG00000196776). Read coverage for the gene (a) and screenshot from the Zmap manual annotation interface. UTR exons and splice variants with no annotated CDS are shown in red, coding exons are shown in green and the CDS portion of models annotated as NMD are shown in purple. Horizontal pink bars represent manually annotated polyadenylation signals and adjacent black bars manually annotation polyadenylation sites. Clusters containing >8000 CAGE tags defining transcription start suites are shown as small blue boxes, CpG islands are shown as yellow boxes broken by horizontal red bars representing TSS predictions from EPONINE [59]. Pale blue boxes represent BLAST alignments of Swissprot proteins. The short horizontal green bars represent polyadenylation sites identified by polyAseq [60].
Figure S9: Summary of Cufflinks *de novo* quantification results for the tissue dataset. (a) Relative abundance of the subset of transcripts in each position of the ranking. (b) Percentage of the studied mRNA pool explained by each category of transcripts. (c) Expression distribution for major and minor transcripts. (d) Relative abundance of the subset of transcripts in each position of the ranking, separating genes by the number of annotated transcripts. (e) Ratio of expression between the major transcript and the second most abundant one, relative to the number of annotated transcripts and including different expression thresholds.

Figure S10: Comparison of expression patterns for major and minor transcripts. (a) Percentage of major and minor transcripts expressed in a given number of tissues (left). Proportion of major vs minor transcripts in each number of tissues (right). Major transcripts tend to be broadly expressed (Spearman correlation: 0.90, p-value<2.2e^{-16}). (b) Recurrence degree of major and minor transcripts. Transcript breadth of expression (ie number of tissues in which a transcript was detected) relative to gene breadth of expression (ie number of tissues in which the corresponding gene was detected) was used as a recurrence measure. Compared to minor transcripts, major ones tend to be expressed in a recurrent fashion.

Figure S11: Expression profile for a subset of the identified switch events across tissues. Strong switch events with changes at the protein sequence level are represented here.

Figure S12: Read coverage for the *MBP* gene across the primary tissues analysed. Read coverage across the 16 tissues in the BM dataset (a) and screenshot from the Zmap manual annotation interface (b). UTR exons and splice variants with no annotated CDS are shown in red, coding exons are shown in green and the CDS portion of models annotated as NMD are shown in purple. Clusters containing >8,000 CAGE tags defining transcription start suites are shown as small blue boxes, CpG islands are shown as yellow boxes broken by horizontal red bars representing TSS predictions from EPONINE [59]. The short horizontal green bars represent polyadenylation sites identified by polyAseq [60].

Figure S13: Average number of genes involved in switch events in cell lines.

Figure S14: Focus on major noncoding transcripts. Protein coding genes for which the most abundant transcript is noncoding are expressed at higher levels in the nucleus, while the trend gets inverted in the cytosol (left). The difference in expression between the first and second most abundant transcripts becomes reduced when the dominant transcript is noncoding and in the nucleus in general (right).
Figure S15: Focus on major retained introns.

(a) Normalised intron expression in different cellular compartments. FPKMs were calculated for all the introns and normalised by gene expression levels (see Methods). Intron expression is higher in the nucleus than in the cytosol (Wilcoxon test p-value < 2.2e^{-16}).

(b) Location of the dominant retained introns within the context of protein coding transcripts. Genes for which the most abundant transcript is a retained intron were initially considered in the analysis. Further filtering of this subset of genes included those for which the second most abundant transcript is protein coding and includes the dominant retained intron. Similar criteria were applied to analyse other less abundant retained introns. The location of the retained introns was then represented by measuring the distance from their centre to the transcriptional start of the selected coding transcript, as illustrated in the panel below the figure. Retained introns are preferentially located towards the transcriptional end of protein coding transcripts, and this trend becomes more apparent in the cytosol.

(c) Number of genes with a major retained intron in nucleus and cytosol.

(d) Expression levels in each cellular compartment for genes with a major retained intron both in nucleus and cytosol (n=81). Expression of this set of genes is higher in the nucleus (Wilcoxon test p-value < 2.2e^{-16}).

Figure S16: Overlap in major transcript predictions within and across replicates for the ENCODE dataset. The detected overlap is higher when comparing replicates than across different cell lines (Wilcoxon test p-value = 2.514e^{-06}).
SFIG 5

- two-fold predominance
- five-fold predominance

FPKM threshold

Major predominant / major nonpredominant

two-fold predominance

five-fold predominance
major transcript length relative to the longest annotated density
SFIG 8

(a)

(b)
SFIG 11

transcript B

transcript A

expression profile = log10 expA

expB

ovary

adrenal

breast

adipose

testis

colon

brain

liver

prostate

lung

heart

lymph_node

white_blood_cells

kidney

skeletal_muscle
average number of genes per sample involved in switch events

2-fold switch events

5-fold switch events

ENCODE (cell lines) ENCODE (replicates)

SFIG 13
SFIG 14

Gene expression (FPKM) ratio second/first transcript

- Major coding
- Major noncoding

BM PE
ENCODE
Cell
ENCODE
Cytosol
ENCODE
Nucleus
(a) (b) (c) (d)

SFIG 15

- Retained intron as major isoform
- Retained intron not as major isoform

% transcript

FPKM

nucleus cytosol cell

425 81 42

protein coding transcript

% transcript

nucleus cytosol

FPKM
% overlap
within replicates
across replicates
density
0 20 40 60 80 100
0.00 0.02 0.04 0.06 0.08 0.10

% overlap

within replicates
across replicates

density

0.00 0.02 0.04 0.06 0.08 0.10
0.00 0.02 0.04 0.06 0.08 0.10
0.00 0.02 0.04 0.06 0.08 0.10
0.00 0.02 0.04 0.06 0.08 0.10