Stem Cell Reports, Volume *8*

Supplemental Information

Mapping Whole-Transcriptome Splicing in Mouse Hematopoietic Stem Cells

Oron Goldstein, Karin Meyer, Yariv Greenshpan, Nir Bujanover, Mili Feigin, Hadas Ner-Gaon, Tal Shay, and Roi Gazit

Supplementary Figure legends

Supplementary Figure 1: Prediction of alternative splicing from microarray data; related to figure 1. (A) All the exons implicated in AS from the microarray analysis list (x-axis) plotted against their fold change values in comparison with the normalized exon expression within their specific gene. Marked in red are exons excluded from the cutoff range we determined (±1.2). Exons in the blue range have a higher probability of AS, and are included in the cutoff we determined. (B) A list of the genes we attempted to validate from the microarray predictions. (C) Representative amplification curves of RT-qPCR for genes from panel B. (D) An expression correlation of annotated splicing-factors between RNA-seq (sample A0) and ImmGen's microarray.

Supplementary Figure 2: Distribution of gene's isoforms across gene ontology (GO) annotation subcategories; related to figure 2. Histograms for isoform-per-gene distribution in: (A) the molecular function gene ontology annotation subcategory; (B) the biological processes gene ontology annotation subcategory; and (C) the cellular components gene ontology annotation subcategory. As a control, random groups of genes (D) are shown. Each plot contains the distribution of isoforms of all genes detected for comparison.

Supplementary Figure 3: ***HoxA9* and *Meis1* co-express two isoforms in mouse HSCs; related to figure 3.** (A) A schematic view of both isoforms and the positioning of the unique primers (red arrows) used in real time PCR. Known exons are represented by blue rectangles. The rectangles are wide for coding region and narrow for untranslated regions (UTRs). (B) Real time PCR amplification curves for the specific isoforms, each amplification curve is visualized using a separate color, variant and canonical are indicated within the figure. (C) A schematic view of the two major isoforms of *Meis1* and the loci of the unique primers. (D) A Sashimi plot of *Meis1* zooming in on the splice junction that skips exon 8. The data were taken from four samples (A0, B0, B1, and A3, in descending order, respectively). (E) RT-qPCR amplification curves of canonical and variant isoforms of *Meis1*. (F) Histogram of the expression of each isoform relative to *Hprt*. Data is shown of technical triplicate average ± SD from one representative experiment out of three.

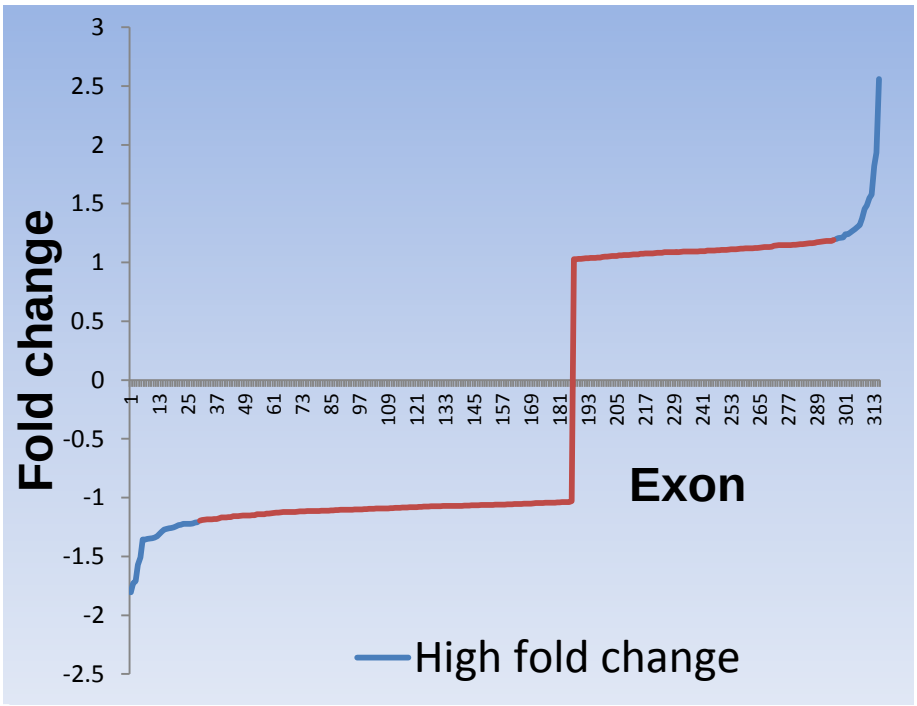
Supplementary Figure 4: Validation of alternatively spliced variants of *Prdm16* and *Nadk2* in HSC; related to figure 4. (A) A schematic view of the two major isoforms and the positioning of the unique primers used (red arrows). (B) Real time PCR amplification curves. Data shown is representative of one experiment out of four. (C) Raw RNA-Seq data showing two of *Nadk2's* isoforms (canonical: NM_001085410, variant: NM_001040395), raw reads, and junctions from the IGV browser. (D) A Sashimi plot zooming in on the splice junction that skips exon 8. The data were taken from four samples (A0, B0, B1, and A3, in descending order, respectively). (E) A schematic view of the canonical and the exon 8-skipping isoform, with the positioning of the unique primers used (red arrows). (F) Real time PCR amplification curves of the indicated isoforms. Data shown is representative of one experiment out of four. (G) A Sashimi plot of the entirety of *Nadk2* from a representative sample (A0), demonstrating both alternative splicing events. (H) Histogram of the expression of the canonical isoform and the variant that skips exon 8 relative to *Hprt*. Data is shown of technical duplicate average ± SD from one representative experiment out of four. . (I) A Sashimi plot zooming in on the splice junction that skips exon 5; data taken from four samples (A0, B0, B1, and A3, in descending order, respectively). (J) Histogram chart of the expression of the canonical isoform and the variant that skips exon 5. Data is shown of technical duplicate average ± SD from one representative experiment out of four.

Supplementary Figure 5: Validation of the unannotated alternatively spliced variant of *CDKn1c* in mouse HSC; related to figure 5. (A) Sashimi plot of *CDKn1c* taken from all 10 samples (A0, B0, A1, B1, A2, B2, A3, B3, A4, and B4, in descending order, respectively). (B) Schematic view of both isoforms and the positioning of the unique primers (red arrows). (C) RT-qPCR amplification curves of indicated isoforms.

Supplementary Figure 6: Two isoforms of *Pbx1*s are expressed to a similar extent in HSCs; related to figure 6. (A) Raw RNA-Seq data showing two of the *Pbx1* isoforms that are transcribed in mouse HSCs (from the Ensembl database), along with splice junctions from the IGV browser. (B) A Sashimi plot zooming in on the splice junction that skips exon 7, taken from a representative sample (A0). (C) A schematic view of the two major isoforms and the positioning of the unique primers used in real time PCR. (D) RT-qPCR amplification curves. Data shown is representative of one experiment out of four. (E) Histogram of the expression of each isoform relative to *Hprt*. Data is shown of technical duplicate average ± SD from one representative experiment out of four.

Supplementary Figure 1; related to figure 1

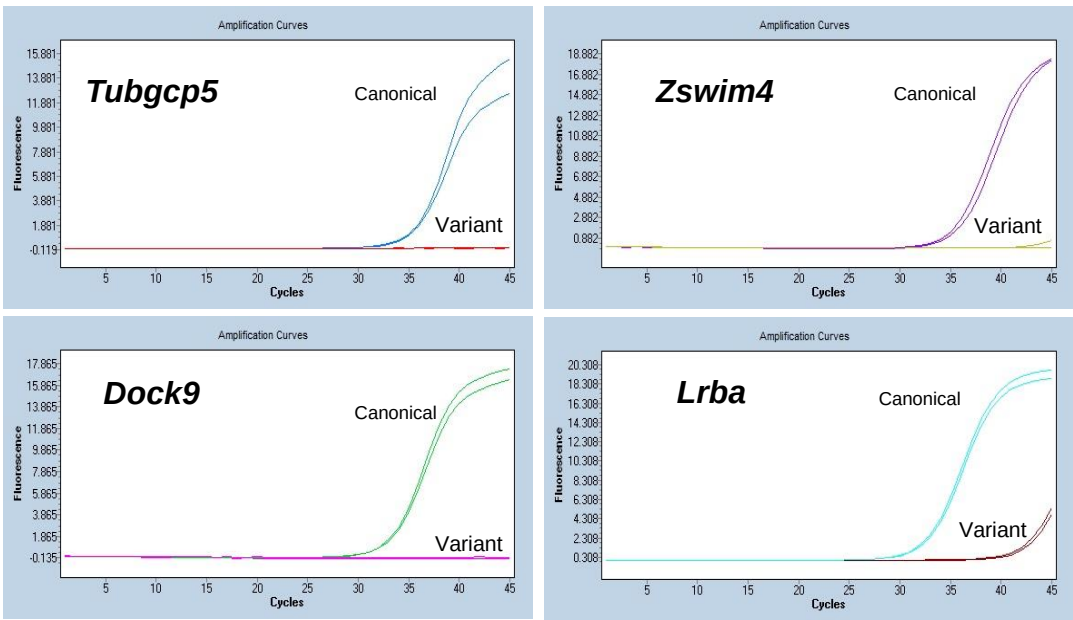
A



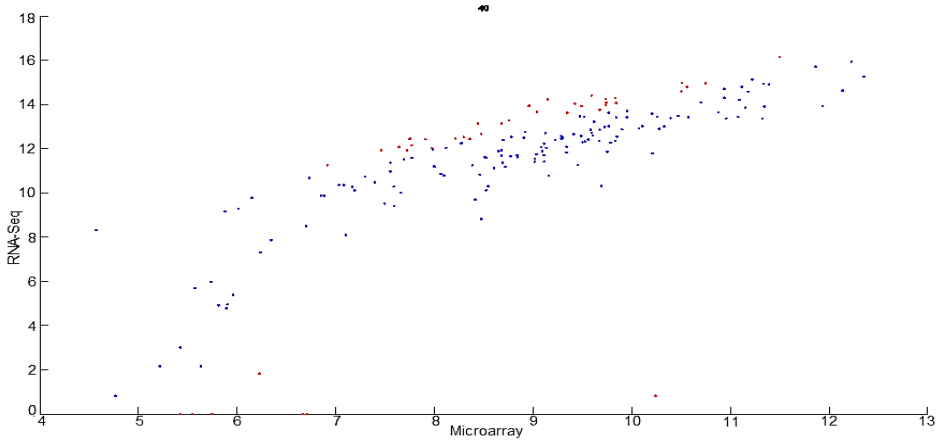
B

Gene symbols
ANKrd27
Zswim4
Lrba
Dock9
Smpd4
Tubcgp5
Wdfy2

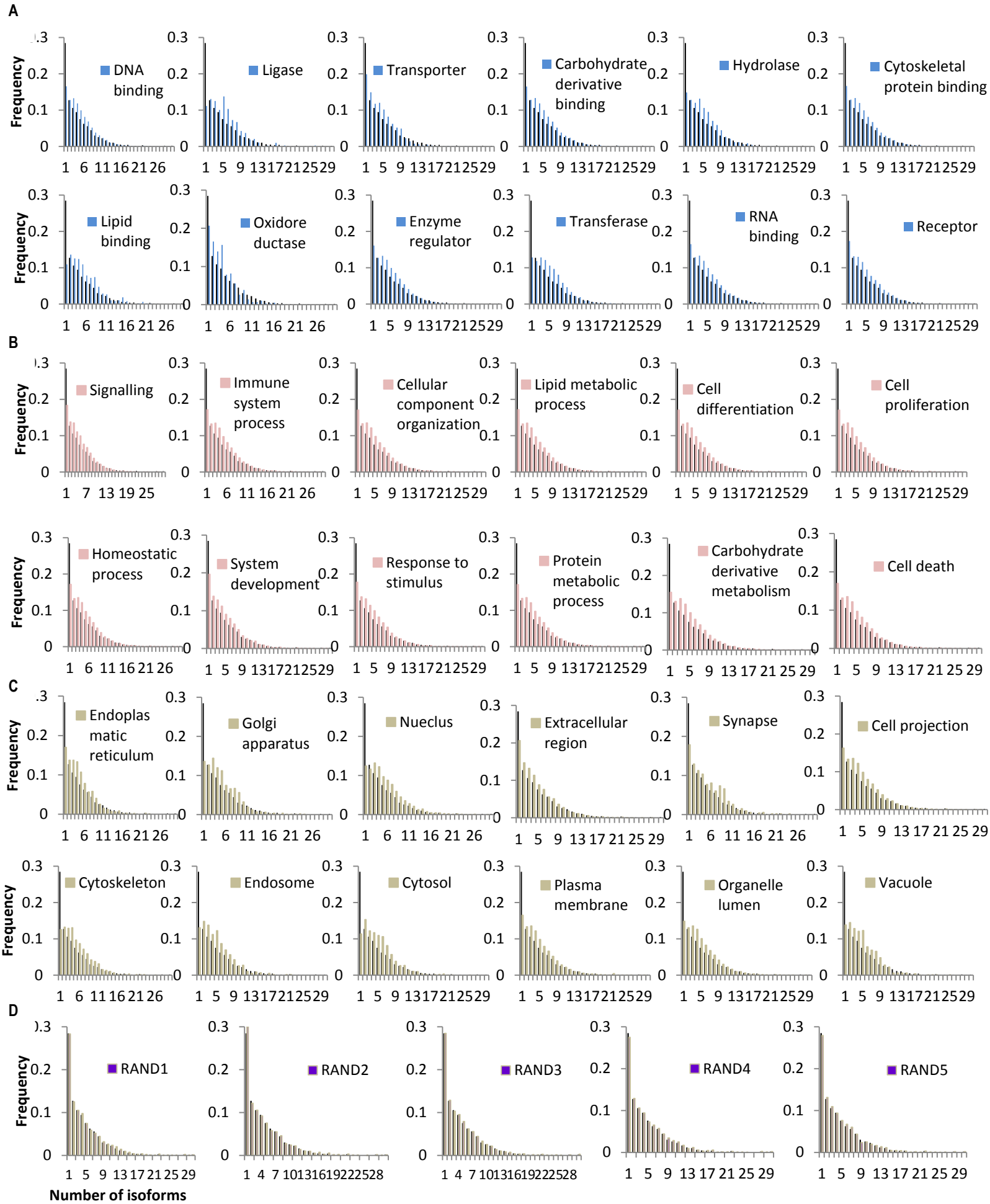
C



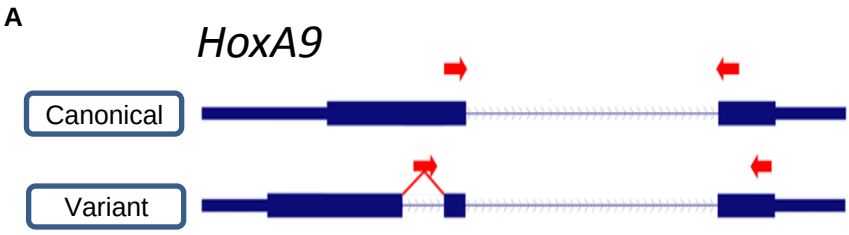
D

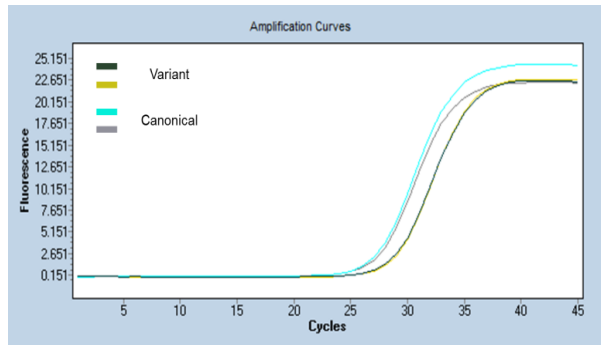


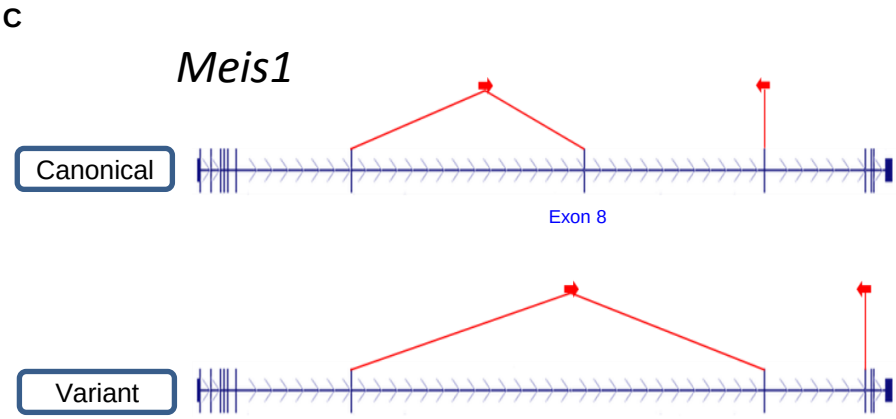
Supplementary Figure 2; related to figure 2

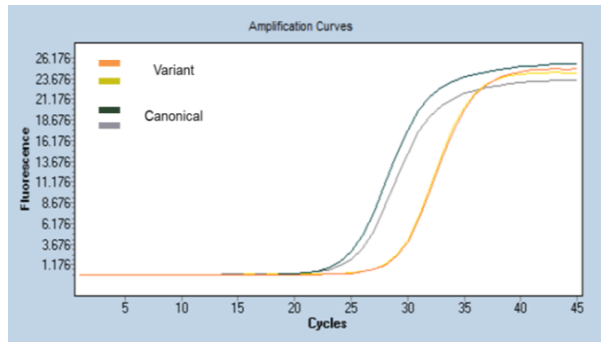


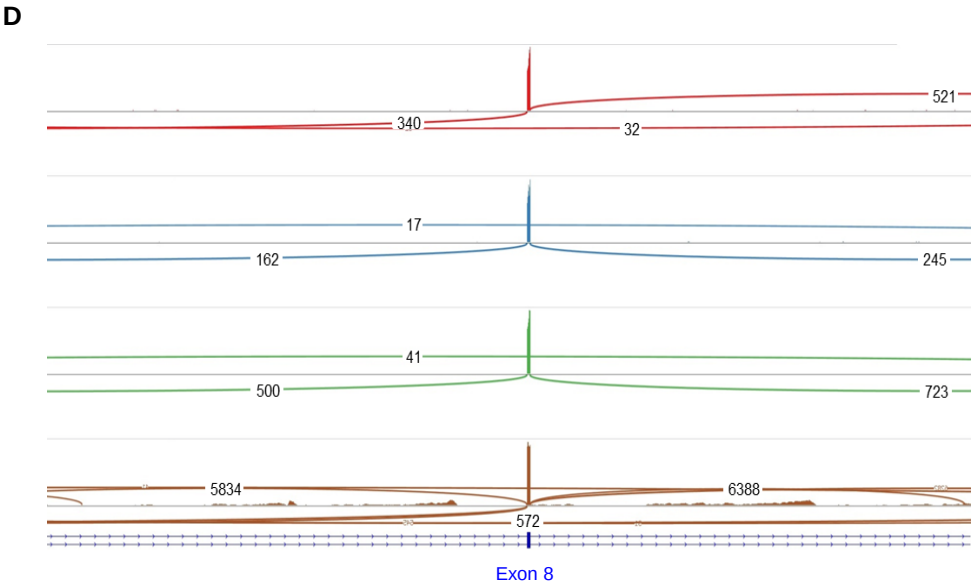
Supplementary Figure 3; related to figure 3

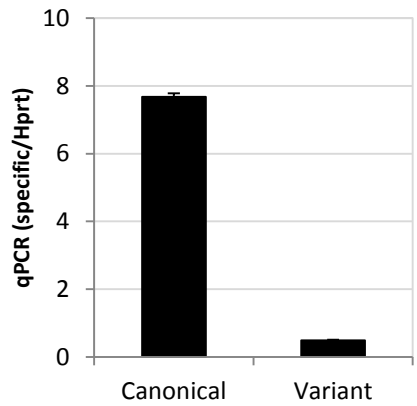




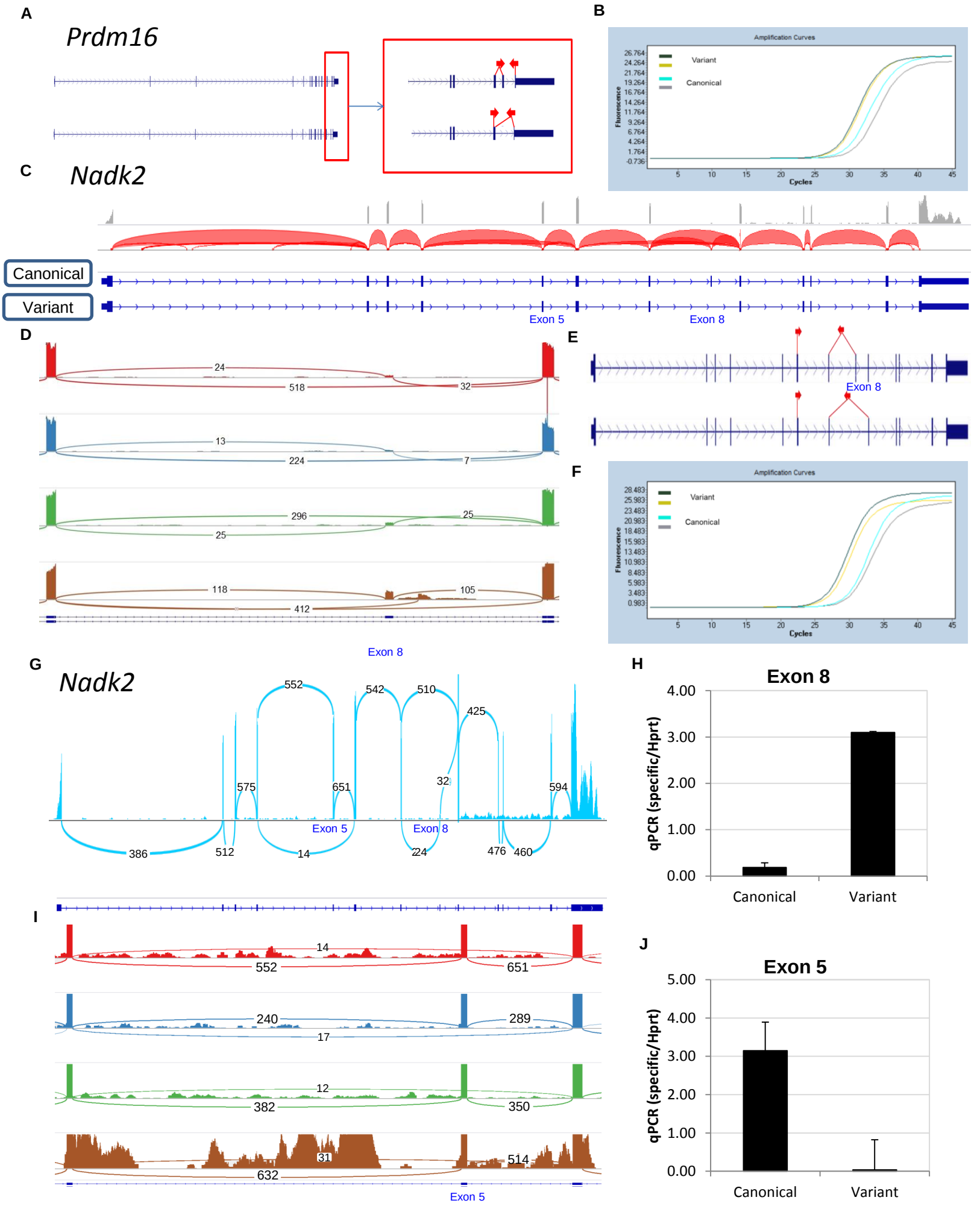




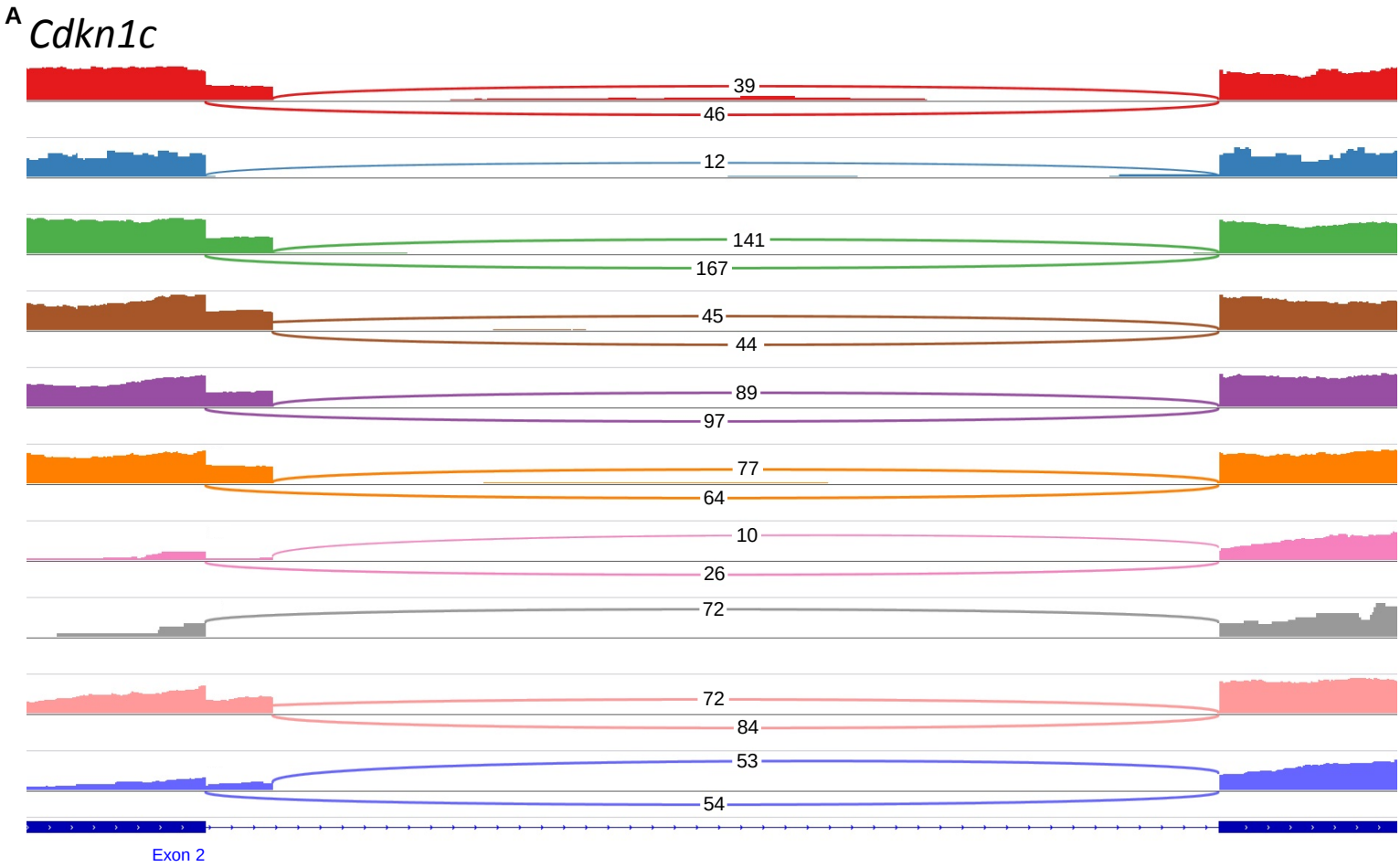


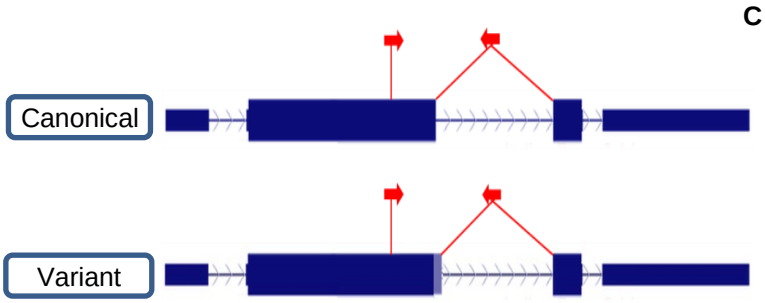


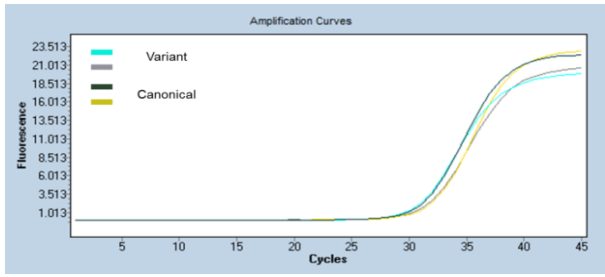
Supplementary Figure 4; related to figure 4



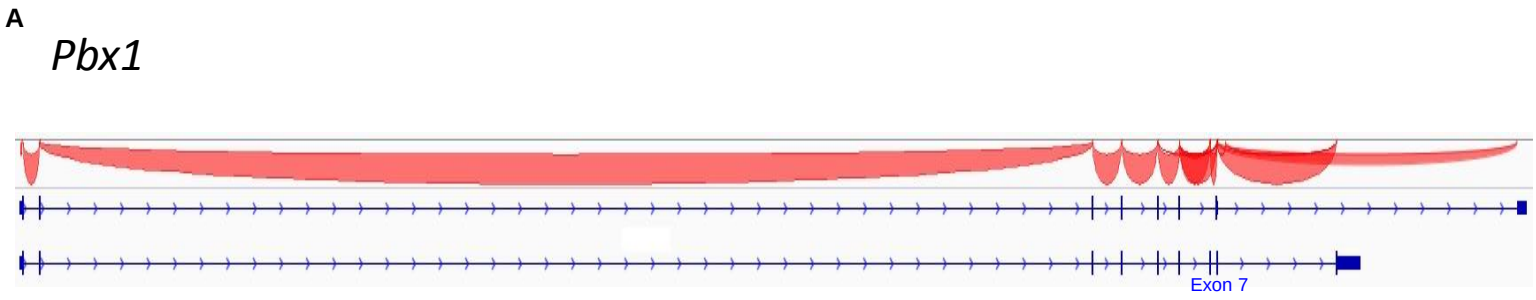
Supplementary Figure 5; related to figure 5

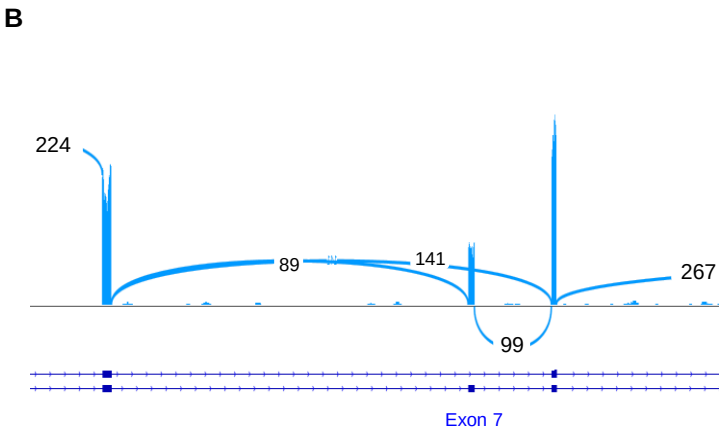


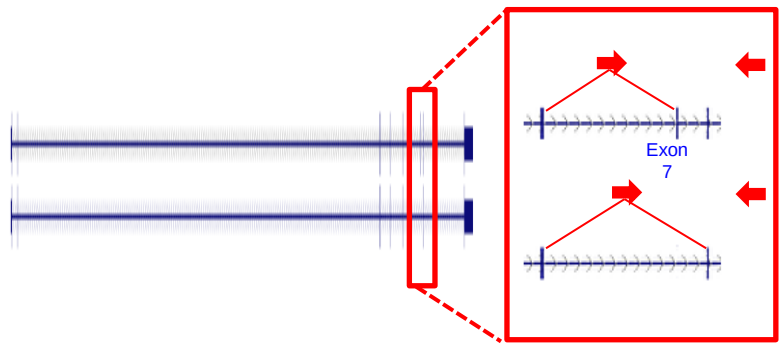


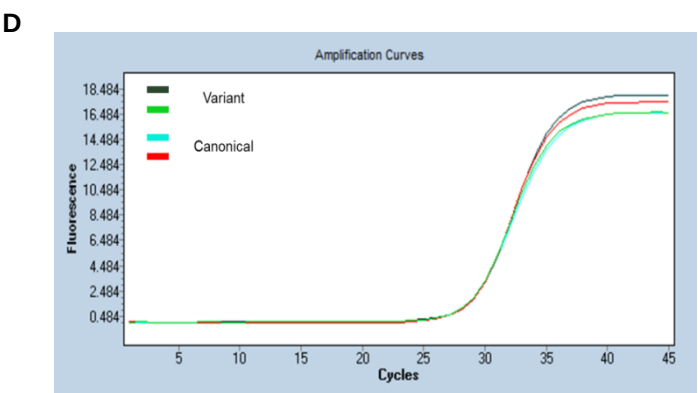


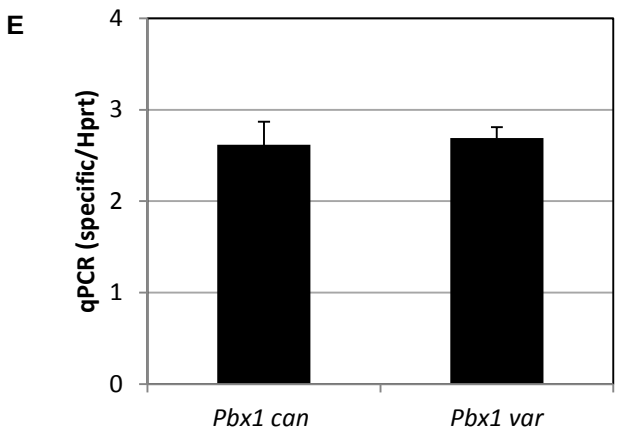
Supplementary Figure 6; related to figure 6











Supplementary Table 1; related to figure 1

Project	Sample	Short name	Markers	Link to paper
Sun	SRR892995*	B_1	SP Lin- Sca1+ cKit+ CD150+	http://www.sciencedirect.com/science/article/pii/S1934590914000964
	SRR892996	A_1		
	SRR892997	B_2		
	SRR892998*	A_2		
Cabezas	ERR420375*	A_3	Lin- Sca1+ cKit+ CD34- CD150+ CD135- CD48-	http://www.cell.com/cell-stem-cell/abstract/S1934-5909(14)00301-4
	ERR420376	B_3		
	ERR420380	A_4		
	ERR420384*	B_4		
Qian	ERR674987	B_0	Lin- Sca1+ cKit+ Flk2-	http://www.cell.com/cell-stem-cell/abstract/S1934-5909(15)00499-3
Venkatraman	ERR296708	A_0	Lin- Sca1+ cKit+ CD34- Flk2-	http://www.nature.com/nature/journal/vaop/ncurrent/abs/nature12303.html

Supplementary table 2; related to figures 3-6

qPCR primers for RNA-seq prediction validation

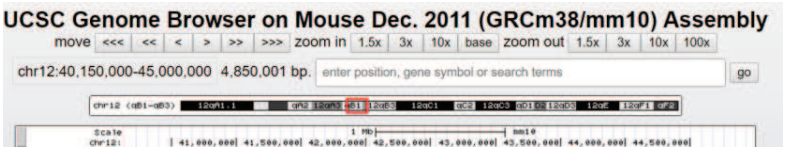
Gene	Sequence (5’ to 3’)
HoxA9 canonical	ccctgactgactatgcttgtgg cgagtggagcgagcatgtag
HoxA9 variant	ctggtgttttgtgtaggggca ccggacggcagttgatagag
Meis1 canonical	agcagtgagcaaggtgatgg ttgtgccaactgctttttctg
Meis1 variant	gcagtgagcaagcaccctta tgggctgcactattcttctcc
Prdm16 canonical	gaccatacccggaggtgtgt agagggacagcatcattgcat
Prdm16 variant	attgcatatgcctccgggta gaggaggaagaggaggagctg
Nadk2 canonical	tcatagtaggaagccagtcccttt cacagccgagcctttaacatc
Nadk2 variant	agtaggaagccctggatgacag cacagccgagcctttaacatc
Cdkn1c canonical	gcttggcgaagaagtcgttc tcgaggagcaggacgagaat
Cdkn1c variant	gcttggcgaagaagtcggag tcgaggagcaggacgagaat
Pbx1 canonical	aactcagcgggttcttccagt ctgtatcctcctgtctggctga
Pbx1 variant	gccttctgtaggggaggtcac ccaactcagcgggtggatac

Supplementary manual

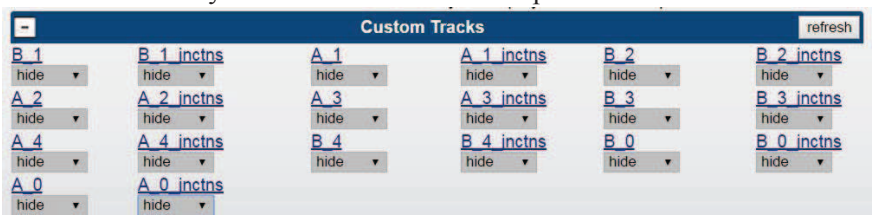
USCS alternative splicing track user guide

Goldstein Oron et al., 2016

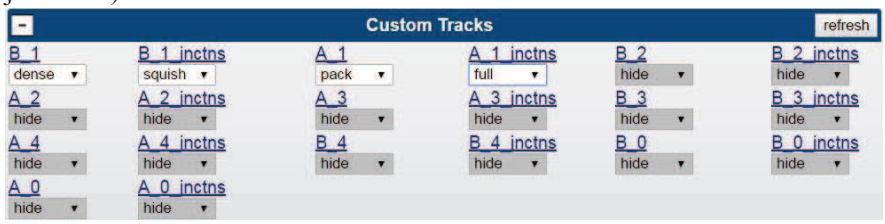
1. Click the link to the track of choice (HSC splice junctions)
2. Enter your gene of interest name (note-some genes have multiple names, if not found in UCSC browser might try NCBI-Gene and use the official gene symbol):



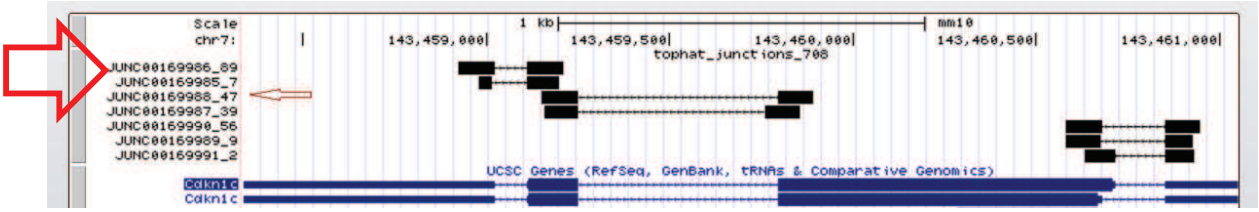
3. Scroll down until you reach the custom tracks panel:



4. Choose the sample(s) you wish to view, and change their setting according to your interest (reads and\or splice junctions)



5. Splice junctions are now displayed and quantified (the number to the right of the lower dash, e.g. JUNC00169988 has 47 counts)



6. Good luck!

Notes:

The junction number represents all the RNA-seq reads that continued from one exon to another and are assembled from aligned reads of both strands.
If you also wish to see exactly how many reads originated from which strand, return to step 4 and mark full on the left most column, then click refresh and all the individual reads will be available.