Overview:
Measuring splicing using RNA-seq data

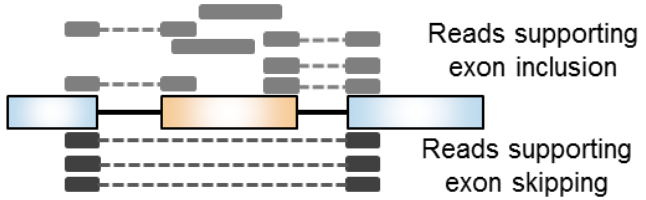
Read mapping to ref. transcriptome

Splicing quantification and paired differential analysis using rMTAS *(Shen et al. PNAS 2014)*

Filter significant events

De novo transcriptome reconstruction

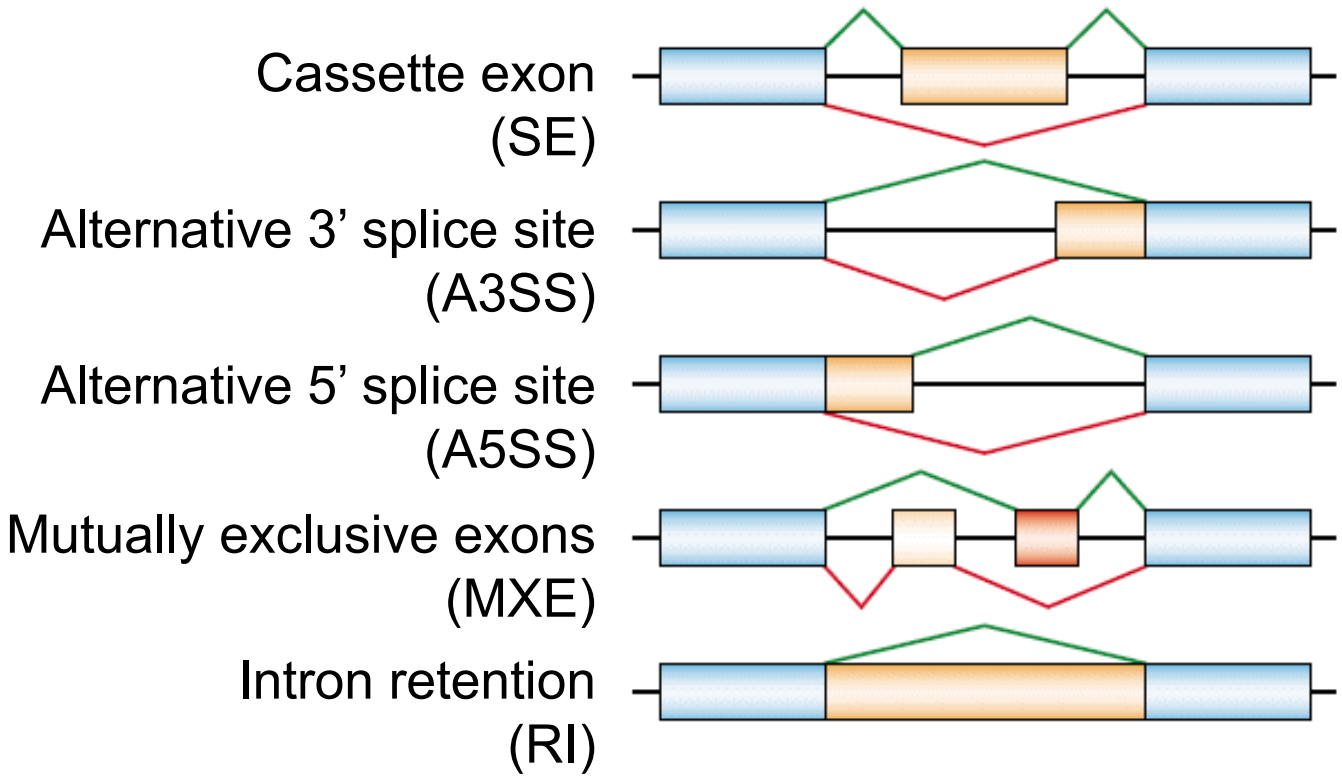
Percent Spliced In (PSI) calculation



$$PSI = \frac{\text{Inclusion reads}}{\text{Inclusion + skipping reads}}$$

- dPSI>|10%|
- > 5 inclusion or skipping reads per replicate

Types of alternative splicing events



Differential splicing analysis pipeline steps

1. Read trimming
2. Read mapping to ref transcriptome or de novo transcriptome
3. Assigns groups for rMATS comparison
4. Run rMATS
5. Filter significant events, correct gene names
6. Generate output tables

Differential splicing analysis pipeline output

Two outputs:

1. Using reference (Gencode) transcriptome *_Ref
2. Using de novo transcriptome *_Novel

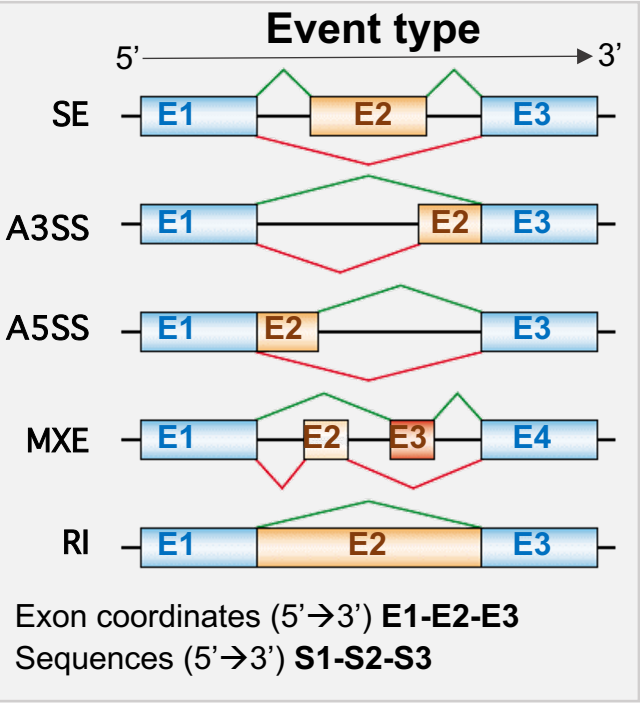
→ **Filtered files** contain only the significant events differentially spliced between case and control

→ **Unfiltered files** contain all the events detected

→ **Summary table** – number of significant events per event type

→ **Bigwig files** for read coverage for visualization using UCSC browser or IGV

Differential splicing analysis output (Filtered files)



Exon coordinates (5'→3') **E1-E2-E3**
Sequences (5'→3') **S1-S2-S3**

Long ID for AS events*

SE; chr;strand; **[E2]** Start;End; **[E1]** Start;End; **[E3]** Start;End

A3SS; chr;strand; **[E2-E3]** Start;End; **[E1]** Start;End; **[E3]** Start;End

A5SS; chr;strand; **[E1-E2]** Start;End; **[E1]** Start;End; **[E3]** Start;End

MXE; chr;strand; **[E2]** Start;End; **[E3]**Start;End; **[E1]** Start;End; **[E4]** Start;End

RI; chr;strand; **[E2]** Start;End; **[E1]** Start;End; **[E3]** Start;End

**Exon coordinates are from left to right. regardless of gene orientation*

short ID for AS events*

SE; chr;strand; **[E2]** Start;End; **[E1]** End; **[E3]** Start

A3SS; chr;strand; **[E2]** Start;End; **[E1]** End; **[E3]** Start

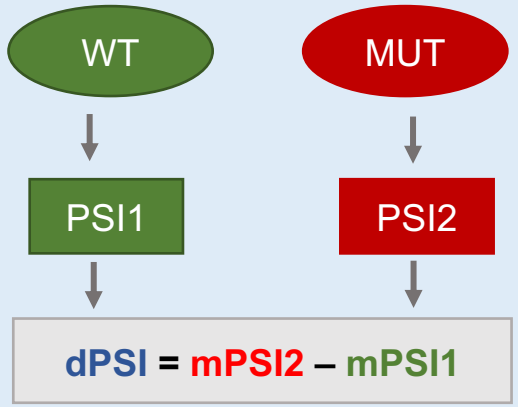
A5SS; chr;strand; **[E2]** Start;End; **[E1]** End; **[E3]** Start

MXE; chr;strand; **[E2]** Start;End; **[E3]**Start;End; **[E1]** End; **[E4]** Start;

RI; chr;strand; **[E2]** Start;End; **[E1]** End; **[E3]** Start;

**Exon coordinates are from left to right. regardless of gene orientation*

PSI Calculation



dPSI (+) Higher inclusion in MUT
dPSI (-) Lower inclusion in MUT

Differential splicing analysis output (Filtered files)

1. Type of alternative event

3. Gene name

16. Average PSI control

17. Average PSI case

18. ΔPSI (Case-Control)
>0 more included, <0 more skipped

Event_I	ShortID	LongID	GeneN	strand	IC1	SC1	IC2	SC2	PSI1	PSI2	mIC1	mSC1	mIC2	mSC2	mPSI1	mPSI2	dPSI_2	PValue	FDR	E1	E2	E3	E4	S1	S2	S3	S4	Genom	BLAT
MXE	MXE;GAS	MXE;GAS5;1;-;17386	***GAS5***	-	26;30;26;4	1;0;4;10	4;3;28;88	36;14;138;	0.982;1.0;(	0.189;0.30	32	4	31	161	95%	27%	-68%	0.00E+00	0.00E+00	1:1738654	1:1738652	1:1738644	1:1738639	CTCCTGG	GTATGGA	GACTCAC	ggtcttgcct(	GRCh38	>MXE;GAS
SE	SE;C20orf	SE;C20orf96;20;-;289	***C20orf96***	-	4;13;8;27	0;0;2;1	5;16;9;28	5;11;14;20	1.0;1.0;0.6	0.313;0.39	13	1	15	13	89%	33%	-56%	5.5E-10	5.79E-07	20:290259	20:289559	20:283963-284081		ACCCAAC	GATTATG	TGTTTCACTTCAAGC	GRCh38	>SE;C20or
SE	SE;PWRN	SE;PWRN1;15;+;249	***SNHG14***	+	5;3;27;22	0;0;0;0	63;8;32;12	21;4;20;58	1.0;1.0;1.0	0.574;0.47	14	0	56	26	100%	49%	-51%	5.11E-15	1.26E-11	15:249817	15:249983	15:24998970-2499907		TGTCATA(	CCATTAT(	GCAATCAGGATGAC	GRCh38	>SE;PWR
RI	RI;ASB1;2	RI;ASB1;2;+;2384463	***ASB1***	+	12;15;0;1	0;0;20;22	2;0;2;0	32;53;27;2	1.0;1.0;0.0	0.044;0.0;(	7	11	1	33	51%	2%	-48%	3.11E-05	0.000997	2:2384463	2:2384510	2:238451046-2384522		GTGTTCC(	CTCTCAG	CTCTCAGTGTGGAA	GRCh38	>RI;ASB1;
SE	SE;PWRN	SE;PWRN1;15;+;249	***PWARSN***	+	14;9;119;5	0;0;0;0	180;23;67;	21;4;20;58	1.0;1.0;1.0	0.646;0.55	49	0	144	26	100%	54%	-47%	0	0	15:249819	15:249981	15:24998970-2499907		GGTTCAT(	GAAGAAT	GCAATCAGGATGAC	GRCh38	>SE;PWR
A5SS	A5SS;nan;	A5SS;nan;1;-;201370	***TNNT2***	-	2264;1892	27;17;22;2	928;827;7(	63;139;93;	0.802;0.84	0.416;0.22	1982	24	1080	98	80%	35%	-46%	0.00E+00	0.00E+00	1:2013720	1:2013700	1:201368162-2013682		AAGCAG(	GTATGTG	AGCAGGAGGAGGC	GRCh38	>A5SS;na
A5SS	A5SS;nan;	A5SS;nan;1;-;201369	***TNNT2***	-	2286;1917	27;17;22;2	984;880;7(	63;139;93;	0.784;0.82	0.401;0.21	2013	24	1183	98	79%	34%	-45%	0.00E+00	0.00E+00	1:2013720	1:2013698	1:201368162-2013682		AAGCAG(	GTATGTG	AGCAGGAGGAGGC	GRCh38	>A5SS;na
SE	SE;KANK	SE;KANK1;9;+;54910	***KANK1***	+	181;206;2(	1;6;2;7	96;111;74;	18;14;15;9	0.962;0.82	0.425;0.52	203	4	83	14	88%	45%	-44%	7.66E-15	1.77E-11	9:540508-5	9:549107-5	9:676890-677009		tgtccttgCT	AAAAACA	GTTGAATGCCTTTGA	GRCh38	>SE;KANK
MXE	MXE;GAS	MXE;GAS5;1;-;17386	***GAS5***	-	25;30;26;4	27;27;31;4	5;3;28;88	53;18;165;	0.659;0.69	0.165;0.25	32	33	31	195	67%	23%	-43%	0.00E+00	0.00E+00	1:1738654	1:1738652	1:1738644	1:1738642	CTCCTGG	GTATGGA	GACTCAC	GGTTTTTA	GRCh38	>MXE;GAS
MXE	MXE;nan;1	MXE;nan;12;-;562916	***CS***	-	10;6;9;18	5;0;4;0	8;1;2;9	9;3;7;8	0.742;1.0;(	0.561;0.32	11	2	5	7	88%	45%	-43%	4.96E-05	0.004158	12:563001(	12:562986	12:562916	12:562865	AGTGGG(	GCACCCA	GAGCTGA	AATGCAT	GRCh38	>MXE;nan;
RI	RI;nan;1;-;	RI;nan;1;-;201368161	***TNNT2***	-	2398;1996	27;17;22;2	1188;1065	63;139;93;	0.693;0.74	0.324;0.16	2171	24	1536	98	70%	28%	-42%	0.00E+00	0	1:2013720	1:2013682	1:201368162-2013682		AAGCAG(	GTATGTG	AGCAGGAGGAGGC	GRCh38	>RI;nan;1;
MXE	MXE;GAS	MXE;GAS5;1;-;17386	***GAS5***	-	26;30;26;4	70;61;77;1	4;3;28;88	105;37;298	0.57;0.637	0.12;0.225	32	79	31	425	59%	20%	-39%	0.00E+00	0	1:1738654	1:1738652	1:1738642	1:1738639	CTCCTGG	GTATGGA	GACTCAC	ggtcttgcct(	GRCh38	>MXE;GAS
A5SS	A5SS;DSF	A5SS;DSP;6;+;75792	***DSP***	+	15;26;199;	0;3;28;6	123;46;24;	52;27;7;15	1.0;0.602;(	0.292;0.22	83	9	69	25	72%	35%	-38%	5.21E-04	0.016895	6:7579275	6:7579773	6:7582642-7586572		CTGAAAA	GTAAGAA	GCATCTAATAGGATT	GRCh38	>A5SS;DS
A3SS	A3SS;nan;	A3SS;nan;6;-;123381	***TRDN***	-	5;1;4;16	0;0;3;0	2;3;3;17	5;3;1;13	1.0;1.0;0.5	0.28;0.493	7	1	6	6	89%	52%	-37%	1.00E-05	1.39E-03	6:1233821	6:1233813	6:123381370-1233813		attccaagaa	tAG	AACAACCCAAGgga	GRCh38	>A3SS;na

5-8. Reads supporting inclusion or skipping for control (1) and case (2) for each replicate

9-10. PSI for control (1) and case (2) for each replicate

30. Event ID and sequences for BLAST

Header	Description
Event type	Type of event: SE (skipped exon or cassette exon), RI (retained intron), A3'SS (alternate 3' splice site or AA), A5'SS (alternate 5' splice site or AD) or MXE (mutually exclusive exon or MX)
Short ID	Short ID of DSEs, nomeclature include event type and 5'-3' coordinate of spliced sequence.
Long ID	Unique ID of DSEs, nomeclature include event type and 5'-3' coordinate of spliced sequence.
GeneName	Official gene symbol
ENSEMBL_ID	ENSEMBL ID
strand	strand (+) or (-)
IC1	Read counts that support inclusion of a spliced region for each biological replicate separated by ";" in control
SC1	Read counts that support skipping of a spliced region for each biological replicate separated by ";" in control
IC2	Read counts that support inclusion of a spliced region for each biological replicate separated by ";" in case
SC2	Read counts that support skipping of a spliced region for each biological replicate separated by ";" in case
PS1	Percent Spliced In or inclusion level for each biological replicate separated by ";" in control
PS2	Percent Spliced In or inclusion level for each biological replicate separated by ";" in case
mIC1	Mean IC1
mSC1	Mean SC1
mIC2	Mean IC2
mSC2	Mean SC2
mPSI1	Mean PSI1
mPSI2	Mean PSI2
dPSI (2-1)	difference in PSI (PSI2 - PSI1)
P-value	rMATS P-value
FDR	rMATS false discovery rate
E1	genomic coordinates for the 5' end of the upstream exon; 3' end of the upstream exon
E2	genomic coordinates for the 5' end of the spliced region; 3' end of the spliced region (for MXE event, this is the first mutually exclusive exon)
E3	genomic coordinates for the 5' end of the downstream exon; 5' end of the downstream exon (for MXE event, this is the second mutually exclusive exon)
E4	genomic coordinates for the 5' end of the downstream exon; 3' end of the downstream exon for MXE event
S1	upstream exon sequence
S2	spliced exon sequence
S3	downstream exon (for MXE this is the second exon)
S4	downstream exon sequence for MXE
MIN_RECIPRO	Filtering (1 Yes, 0 No) for min 5 reads for IC1 OR IC2 AND 5 reads for SC1 OR SC2
genome	version of reference genome assembly
BLAT	fasta formatted event ID and sequences of E1-E2-E3-E4

Visualizing RNA-seq bw files on UCSC genome browser

1) **Create bw files with UCSC coordinates** (e.g., "chr1" instead of "1")

2) Transfer files to JAX FTP server
- connect using WinSCP via SCP or SFTP protocol
- server: jaxbhftp02.jax.org
- login: anczukow-lab
- password: 5DSgc4c0aYaGk24gQOLAqgZb
- copy files to folder
- check that you see files here: ftp://ftp.jax.org/anczukow-lab and right click to copy link location

3) Load files to UCSC genome browser
- Login to UCSC- create login if needed
- MyData>CustomTracks
- For each bw file add the name and ftp location:

track type=bigWig name="Sample_Name" description="Sample_description" color=0,51,204
bigDataUrl=ftp://ftp.jax.org/anczukow-lab/file_name.bw

Visualizing RNA-seq bw files on UCSC genome browser

- Color examples:
  - blue: color=0,51,204
  - red: color=179,0,0
  - green: color=0,102,0
  - purple: color=89,0,179
  - other https://www.w3schools.com/colors/colors_picker.asp

- Submit and add custom track
- More here: https://genome.ucsc.edu/goldenpath/help/bigWig.html

4) Save and share the tracks

- MyData>MySession
- Add name and save session
- Share session with collaborators.

5) BLAT the sequences from filtered table