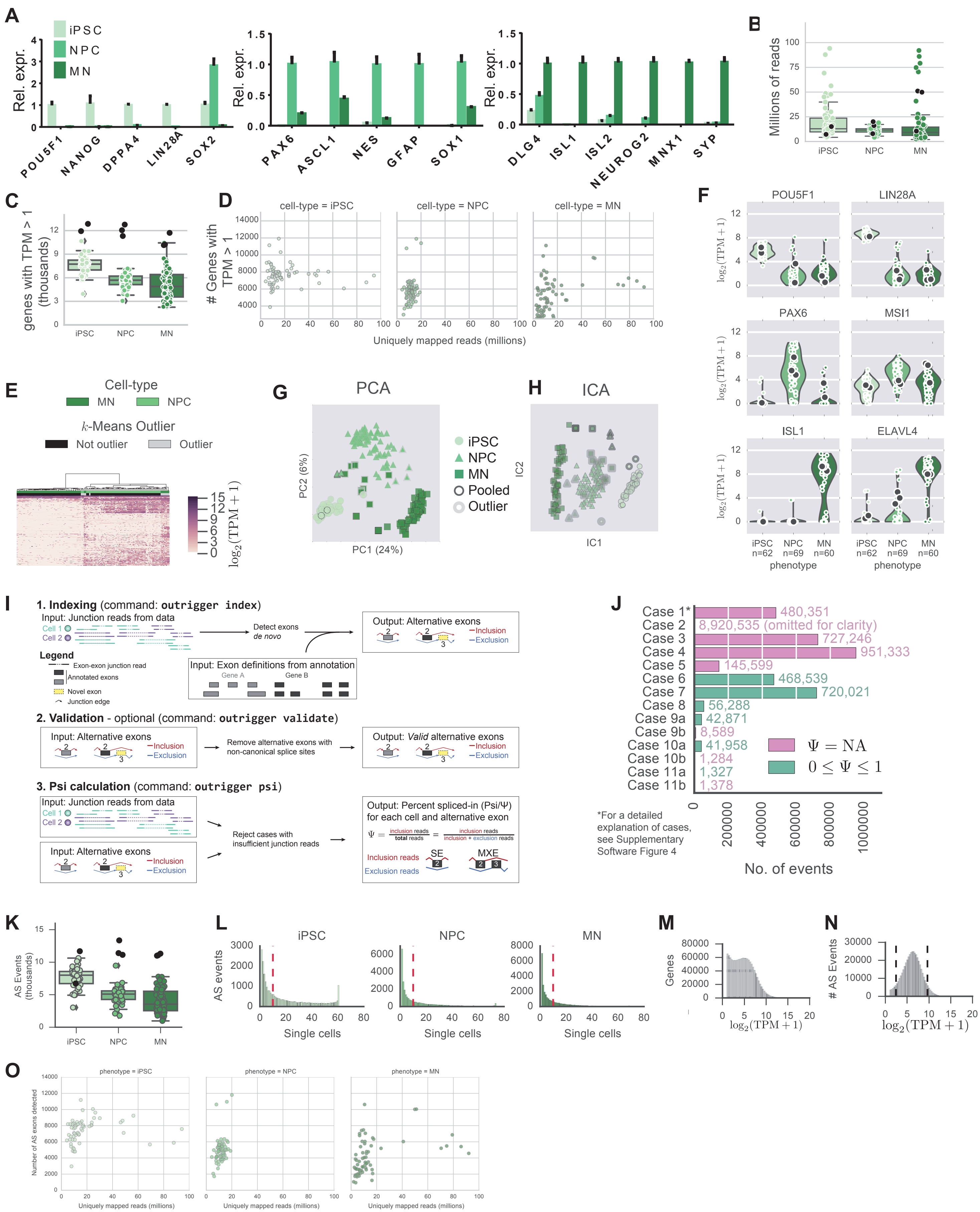
Figure S1



A
iPSC
NPC
MN
Rel. expr.
POU5F1
NANOG
DPPA4
LIN28A
SOX2
PAX6
ASCL1
NES
GFAP
SOX1
DLG4
ISL1
ISL2
NEUROG2
MNX1
SYP
B
Millions of reads
iPSC
NPC
MN
C
genes with TPM > 1 (thousands)
D
cell-type = iPSC
cell-type = NPC
cell-type = MN
Genes with TPM > 1
Uniquely mapped reads (millions)
E
Cell-type
MN
NPC
k-Means Outlier
Not outlier
Outlier
$\log_2(\mathrm{TPM}+1)$
F
POU5F1
LIN28A
PAX6
MSI1
ISL1
ELAVL4
$\log_2(\mathrm{TPM}+1)$
iPSC n=62
NPC n=69
MN n=60
phenotype
G
PCA
PC2 (6%)
PC1 (24%)
iPSC
NPC
MN
Pooled
Outlier
H
ICA
IC2
IC1
I
1. Indexing (command: outrigger index)
Input: Junction reads from data
Cell 1
Cell 2
Detect exons de novo
Output: Alternative exons
Inclusion
Exclusion
Legend
Exon-exon junction read
Annotated exons
Novel exon
Junction edge
Input: Exon definitions from annotation
Gene A
Gene B
2. Validation - optional (command: outrigger validate)
Input: Alternative exons
Remove alternative exons with non-canonical splice sites
Output: Valid alternative exons
3. Psi calculation (command: outrigger psi)
Input: Junction reads from data
Input: Alternative exons
Reject cases with insufficient junction reads
Output: Percent spliced-in (Psi/Ψ) for each cell and alternative exon
$\Psi = \frac{\text{inclusion reads}}{\text{total reads}} = \frac{\text{inclusion reads}}{\text{inclusion + exclusion reads}}$
Inclusion reads
Exclusion reads
SE
MXE
J
Case 1* 480,351
Case 2 8,920,535 (omitted for clarity)
Case 3 727,246
Case 4 951,333
Case 5 145,599
Case 6 468,539
Case 7 720,021
Case 8 56,288
Case 9a 42,871
Case 9b 8,589
Case 10a 41,958
Case 10b 1,284
Case 11a 1,327
Case 11b 1,378
$\Psi = \mathrm{NA}$
$0 \leq \Psi \leq 1$
*For a detailed explanation of cases, see Supplementary Software Figure 4
No. of events
K
AS Events (thousands)
L
iPSC
NPC
MN
AS events
Single cells
M
Genes
$\log_2(\mathrm{TPM}+1)$
N
AS Events
$\log_2(\mathrm{TPM}+1)$
O
phenotype = iPSC
phenotype = NPC
phenotype = MN
Number of AS exons detected
Uniquely mapped reads (millions)

Supplementary Figures

Figure S1: Related to Figure 1. Quality control of scRNA-seq libraries and overview of *outrigger*.

A. RT-qPCR validation of biomarker expression in the bulk populations of iPSCs (light green), NPCs (medium green), MNs (dark green). Relative expression of the indicated genes was normalized to housekeeping genes RPL27 and PGK.

B. Sequencing depth for single cell libraries. On average, 10-20 million reads of 100bp length was obtained.

C. Number of detected genes in single cells. Approximately 4,000-6,000 genes were detected at the cutoff of TPM > 1 in single cells.

D. Number of detected genes in contrast to sequencing depth for each population. x-axis, number of reads that mapped uniquely to the genome (fewer than 10 locations), y-axis, number of genes with TPM > 1 detected in each sample. Bulk samples are indicated with a black outline and outlier samples are indicated with a grey outline.

E. Outlier MN cells identified by K-means clustering exhibited a transcriptome resembling NPCs.

F. Expression of lineage-specific transcription factors (left) and RNA binding proteins (right). Specifically, POU5F1/OCT4 and LIN28A are specific to iPSCs, PAX6 and MSI1 are more highly expressed in NPCs, and ISL1 and ELAVL4 are only expressed in MNs.

G. PCA of highly variant gene expression. Highly variant is defined as two standard deviations away from mean gene-level variance across all samples.

H. ICA on highly variant gene expression. Highly variant is defined as two standard deviations away from mean gene-level variance across all samples.

I. Three steps of outrigger and associated commands: indexing (*outrigger index*), validation (*outrigger validate*) and percent spliced-in (Psi/Ψ) calculation (*outrigger psi*). In the first step of building an index, outrigger considers the entirety of junction reads from the user-input dataset to detect exons *de novo*, integrates annotated exons, then searches for alternative exons. In the second step (optional) of validation, outrigger removes alternative exons lacking consensus splice sites. In the third step of calculating Psi/Ψ, outrigger utilizes junction reads together with alternative exons defined in the indexing step to calculate Ψ for each event covered with sufficient reads. Only junction reads are used in outrigger. SE, Skipped Exon; MXE, Mutually Exclusive Exons.

J. AS events identified by outrigger across all splicing events and all samples. Events are binned by cases (see supplementary software figure 4).

K. The number of AS exons (both SE and MXE event types) detected per single cell library.

L. Number of cells per each AS exon. Many AS exons were detected in only one cell. A event has to be detected in a minimum of 10 cells to be considered for following analysis, as indicated by a dashed red line.

M. Number of detected genes in contrast to their expression level in $\log_2(\text{TPM} + 1)$.

N. Number of detected AS events in contrast to gene expression in $\log_2(\text{TPM} + 1)$. 90% of the detected splicing events reside in transcripts expressed between 2.5-10 of $\log_2(\text{TPM} + 1)$, as indicated by a dashed black line.

O. Number of detected AS events in contrast to the sequencing depth for each population. x-axis, number of reads that mapped uniquely to the genome (fewer than 10 locations), y-axis, number of non-NA AS events detected in each sample. Bulk samples are indicated with a black outline and outlier samples are indicated with a grey outline.

Figure S2

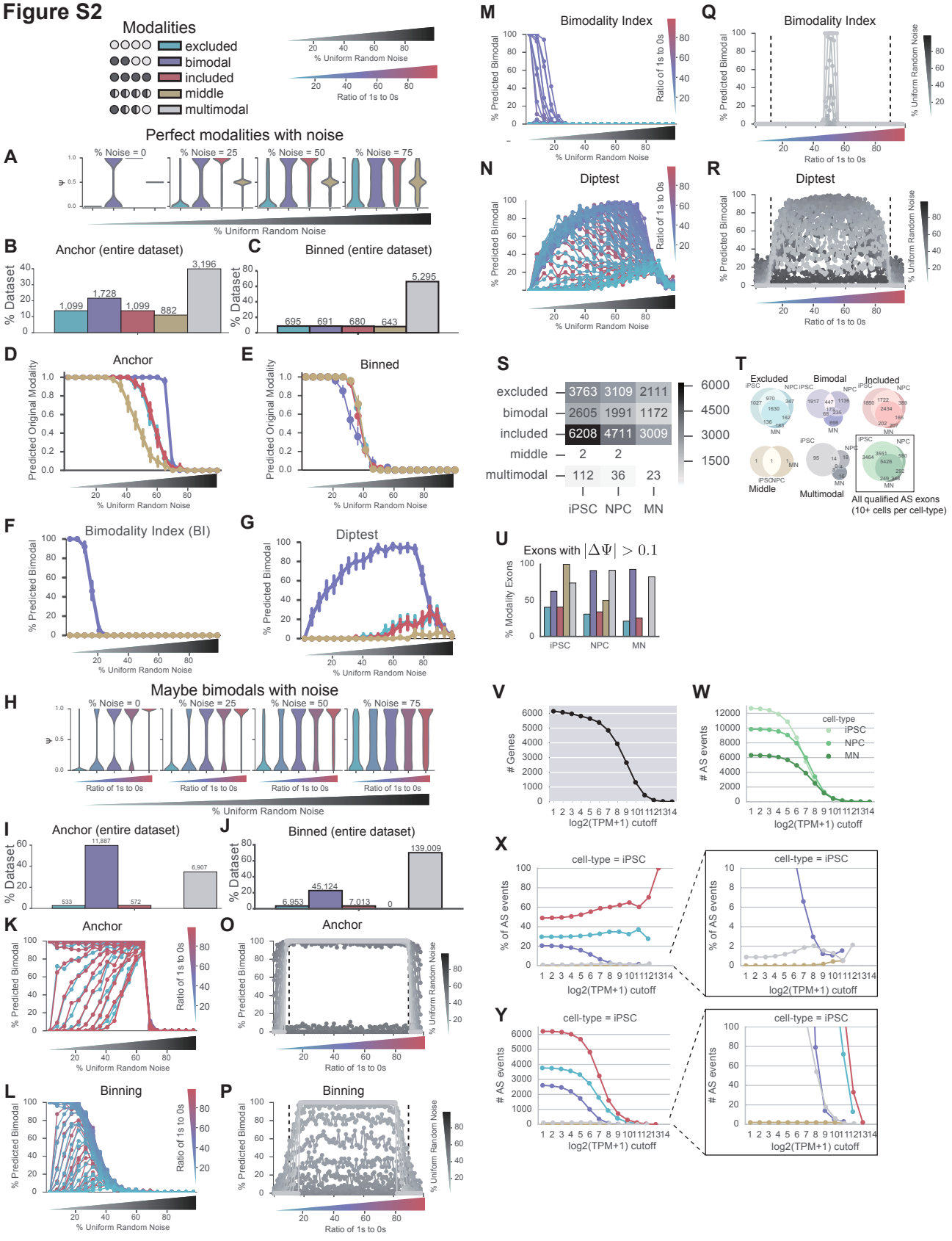

Modalities
excluded
bimodal
included
middle
multimodal
% Uniform Random Noise
Ratio of 1s to 0s
Perfect modalities with noise
A
% Noise = 0
% Noise = 25
% Noise = 50
% Noise = 75
Ψ
% Uniform Random Noise
B
Anchor (entire dataset)
% Dataset
1,099
1,728
1,099
882
3,196
C
Binned (entire dataset)
695
691
680
643
5,295
D
Anchor
Predicted Original Modality
E
Binned
F
Bimodality Index (BI)
% Predicted Bimodal
G
Diptest
H
Maybe bimodals with noise
I
Anchor (entire dataset)
533
11,887
572
6,907
J
Binned (entire dataset)
6,953
45,124
7,013
0
139,009
K
Anchor
Ratio of 1s to 0s
O
Anchor
L
Binning
P
Binning
M
Bimodality Index
N
Diptest
Q
Bimodality Index
R
Diptest
S
excluded 3763 3109 2111
bimodal 2605 1991 1172
included 6208 4711 3009
middle 2 2
multimodal 112 36 23
iPSC NPC MN
6000
4500
3000
1500
T
Excluded
Bimodal
Included
Middle
Multimodal
All qualified AS exons
(10+ cells per cell-type)
U
Exons with $|\Delta\Psi| > 0.1$
% Modality Exons
iPSC
NPC
MN
V
Genes
log2(TPM+1) cutoff
W
AS events
cell-type
iPSC
NPC
MN
X
cell-type = iPSC
% of AS events
Y
AS events

Figure S2: Related to Figure 2. Simulated datasets to test *anchor* performance.
A. Simulated modality dataset was created with increasing noise. The base dataset (% Noise = 0) is consisted of 100 samples of either all zeros (excluded), half zeros and half ones (bimodal), all ones (included), or all 0.5s (middle), exactly representing the four modalities. Uniform random noise was added in 5% increments, with 100 iterations at each noise level.
B. Percentage of events categorized as different modalities by *anchor* in the randomly generated test datasets, across all noise levels, as illustrated in (**A.**). Number of events in each modality is annotated on top of the barplots.
C. Percentage of events categorized as different modalities by binning in the randomly generated test datasets, across all noise levels, as illustrated in (**A.**). Number of events for each modality is annotated on top of the barplots.
D-G. Specificity of modality estimation. Recapitulation of the original modality as a function of additional noise, using anchor (**D.**), binning (**E.**), Bimodality index (**F.**), and diptest (**G.**) methods. The x-axis depicts the percent of uniform random noise added (visualized as a triangle gradient), and the y-axis depicts the fraction of times a noisy feature was categorized into each modality. The hue of the line is the modality.
H. Simulated dataset of "Maybe Bimodals" was created to test the robustness. The test set is consisted of potential bimodal events, each containing 100 samples of only zeros ($\Psi = 0$) and ones ($\Psi = 1$) in every combination, shown here as relative to the number of ones. Uniform random noise was added in increasing 5% levels for 100 iterations at each level. While each combination of 1s and 0s was created, only a subset are shown for brevity – 1:99, 25:75, 50:50, 75:25, and 99:1 ratios of 1:0 are shown, with added uniform random noise of 0% (original), 25%, 50%, and 75%.
I. Percentage of events categorized in modalities by anchor in the randomly generated bimodal test datasets, across all noise levels, as illustrated in (**H.**). Number of events for each modality is annotated on top of the barplots.
J. Percentage of events categorized in modalities by binning in the randomly generated bimodal test datasets, across all noise levels, as illustrated in (**H.**). Number of events for each modality is annotated on top of the barplots.
K-R. Accuracy of bimodality prediction, as a function of the noise added to the dataset.
K-N. Specificity of bimodality estimation upon addition of uniform random noise. The x-axis shows the percent added uniform random noise (visualized as a triangle gradient), and the y-axis indicates the fraction of time features in each noise percentage and proportion of $1:0$ was categorized as bimodal. Overall, all but the very extremes of the $1:0$ proportions were consistently categorized as bimodal until 70% noise, after which point nearly all events became multimodal. Modality estimations are shown using anchor (**K.**), binning (**L.**), Bimodality Index (**M.**), and Diptest (**N.**).
O-R. Sensitivity of bimodality detection. Percentage of events predicted as bimodal given different proportions of 0s and 1s, and increasing uniform random noise. Events are called as bimodal with approximately 9:1 ratio of 0s and 1s (and vice versa), shown with a dotted line at 10% ones and 90% ones. Bottom triangle gradient shows increasing ratio of ones to zeros, i.e. from exclusion to bimodal, to inclusion. Bimodality estimations are shown using anchor (**O.**), binning (**P.**), Bimodality Index (**Q.**), and Diptest (**R.**).
S. Summary of total number of AS events identified by outrigger and their modality identified by anchor for each cell type.
T. Venn diagrams of events shared in modalities between cell types. AS events in included and excluded modality are largely shared across the three cell types, but fewer bimodal events are shared across three cell types. Boxed, all AS events, regardless of modality.
U. Percentage of modality AS events inconsistent with pooled estimates, where the mean difference of psi between singles and pooled ($|\Delta\overline{\Psi}|$) is greater than 0.1.
V-Y. Effect of the expression level on modality estimation.
V. Number of detected genes at varying expression cutoffs.
W. Number of AS exons at varying expression cutoffs.
X. Percentage of modality estimated at different expression cutoffs (right, zoomed in panel).

Y. Number of modality events estimated at different expression cutoffs (right, zoomed in panel).

Figure S3

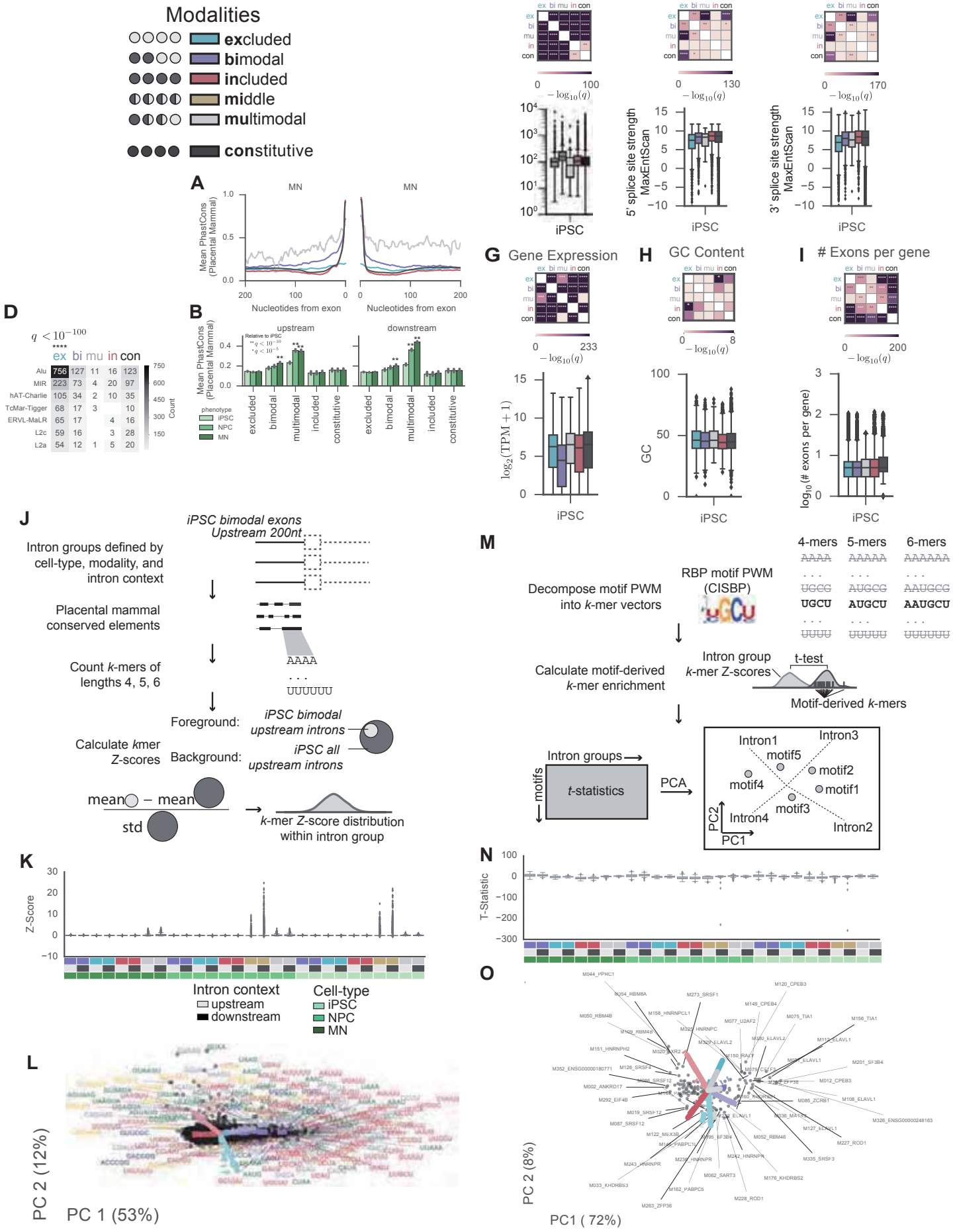

Modalities
excluded
bimodal
included
middle
multimodal
constitutive
A
MN
MN
Mean PhastCons (Placental Mammal)
Nucleotides from exon
Nucleotides from exon
B
upstream
downstream
Mean PhastCons (Placental Mammal)
phenotype
iPSC
NPC
MN
excluded
bimodal
multimodal
included
constitutive
C AS Exon length
E 5' Splice site
F 3' Splice site
5' splice site strength MaxEntScan
3' splice site strength MaxEntScan
iPSC
G Gene Expression
H GC Content
I # Exons per gene
GC
iPSC
D
ex bi mu in con
Alu 756 127 11 16 123
MIR 223 73 4 20 97
hAT-Charlie 105 34 2 10 35
TcMar-Tigger 68 17 3 10
ERVL-MaLR 65 17 4 16
L2c 59 16 3 28
L2a 54 12 1 5 20
Count
J
iPSC bimodal exons
Upstream 200nt
Intron groups defined by cell-type, modality, and intron context
Placental mammal conserved elements
Count k-mers of lengths 4, 5, 6
AAAA
UUUUUU
Calculate kmer Z-scores
Foreground:
iPSC bimodal upstream introns
Background:
iPSC all upstream introns
mean − mean
std
k-mer Z-score distribution within intron group
K
Z-Score
Intron context
upstream
downstream
Cell-type
iPSC
NPC
MN
L
PC 2 (12%)
PC 1 (53%)
M
Decompose motif PWM into k-mer vectors
RBP motif PWM (CISBP)
4-mers 5-mers 6-mers
AAAA AAAAA AAAAAA
UGCU AUGCU AAUGCU
UUUU UUUUU UUUUUU
Calculate motif-derived k-mer enrichment
Intron group k-mer Z-scores
t-test
Motif-derived k-mers
Intron groups
motifs
t-statistics
PCA
Intron1
Intron2
Intron3
Intron4
motif1
motif2
motif3
motif4
motif5
PC1
PC2
N
T-Statistic
O
PC 2 (8%)
PC1 (72%)

Figure S3: Related to Figure 3. Molecular features of each splicing modality.
A. Flanking intron sequence is more conserved in bimodal modality. Shown in motor neurons, intron conservation of bimodal events is slightly higher than excluded AS events.
B. Mean placental mammal PhastCons score in introns flanking exons in different modalities, across cell types. Bimodal and multimodal exons in motor neurons and NPCs are statistically enriched for higher conservation as compared to iPSCs (Kolmogorov-Smirnov test, Bonferroni-corrected).
C. Length of the alternative exons in different modalities. Comparisons are at the bottom and statistics are on the top. Bimodal exons are statistically enriched for longer exons, compared to excluded modality (Kolmogorov-Smirnov test, Bonferonni-corrected).
D. Number of AS events overlapping with repetitive elements in each modality, shown in iPSC. Excluded modality is statistically enriched for repetitive elements ($q < 10^{-50}$, Hypergeometric test).
E. $5'$ splice site scores of the exon in each modality measured by MaxEntScan. Bimodal and excluded exons have statistically significantly lower splice site scores than included exons (Kolmogorov-Smirnov test, Bonferonni-corrected).
F. $3'$ splice site scores of the exon in each modality measured by MaxEntScan. Bimodal and excluded exons have statistically significantly lower splice site scores than included exons (Kolmogorov-Smirnov test, Bonferonni-corrected).
G. Mean expression level of genes ($\log_2(\mathrm{TPM}+1)$, x axis) harboring corresponding AS events in each modality. While events from all five modalities are detected across entire range of gene expression, genes containing bimodal exons are statistically enriched for lower expression (Kolmogorov-Smirnov test, Bonferonni-corrected).
H. GC content of the alternative exons of different modalities. Excluded exons are statistically enriched for higher GC content, compared to included exons (Kolmogorov-Smirnov test, Bonferonni-corrected).
I. Number of exons per gene in each modality, measured by the maximum number of exons in any transcript of a given gene. Genes containing excluded exons are statistically enriched for fewer exons per gene (Kolmogorov-Smirnov test, Bonferonni-corrected).
J. Intron groups and the process of obtaining conserved k-mer Z-scores. Intron groups are defined by cell-type, modality, and intron context.
K. Z-scores of k-mer enrichment in the different intron groups, labeled with a color bar of modality, intron context, and cell-type.
L. PCA on k-mer Z-scores, with each point as a k-mer and the vector components as the intron groups. k-mers with PCA distance greater than 2.5 standard deviations away from zero were labeled with the color of the majority nucleotide. If there was a tie for the majority nucleotide, it was assigned the color grey. An interactive version of this plot can be viewed here: https://plot.ly/~OlgaBotvinnik/20/modality-k-mer-z-scores-background-phenotype/. Multimodal is not shown because its k-mer enrichment has a much larger range than the other modalities and overwhelms the plot.
M. Overview of motif enrichments calculated from intron groups using a t-test and their transformation into PCA for visualization.
N. Boxplots of the t-statistics of motif enrichment in different intron groups, labeled with colorbars of modality, intron context, and cell-type.
O. PCA on the t-statistics of the Motif enrichment, labeled with the motif ID and RPB name from CISBP v0.6. An interactive version of this plot is available at https://plot.ly/~OlgaBotvinnik/32/cisbp-motif-t-test-enrichments-background-phenotype/

Figure S4

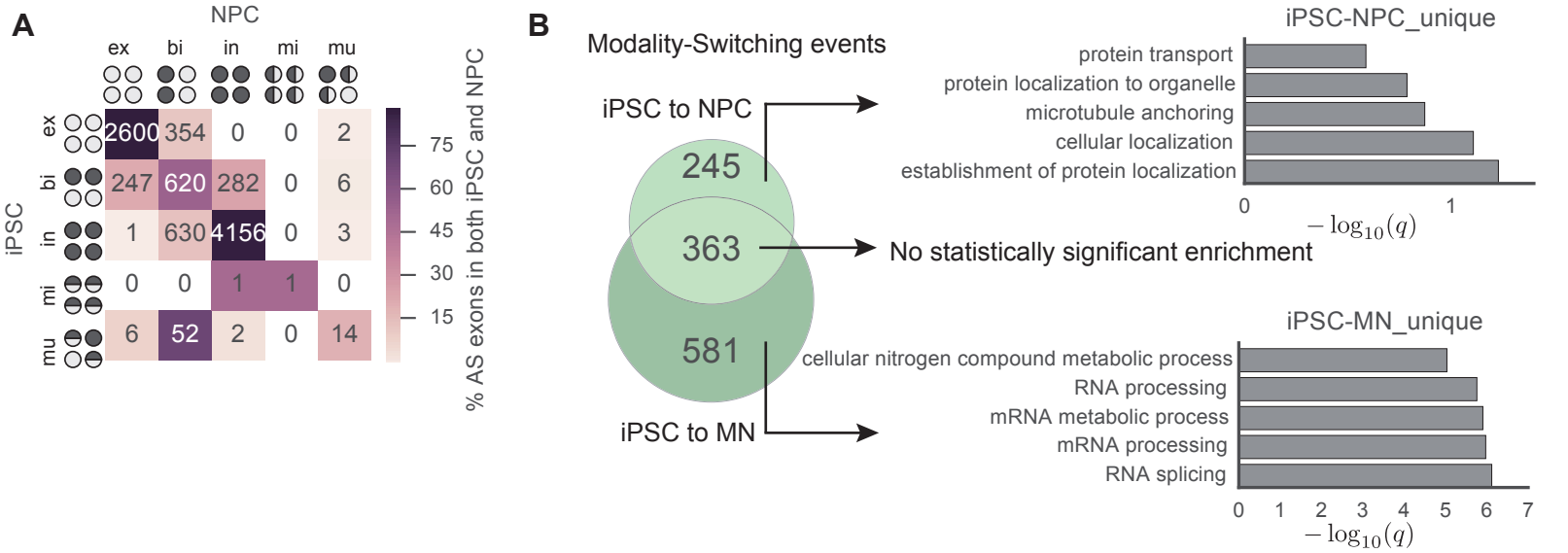

A
NPC
ex
bi
in
mi
mu
iPSC
2600 354 0 0 2
247 620 282 0 6
1 630 4156 0 3
0 0 1 1 0
6 52 2 0 14
75
60
45
30
15
% AS exons in both iPSC and NPC
B
Modality-Switching events
iPSC to NPC
245
363
581
iPSC to MN
iPSC-NPC_unique
protein transport
protein localization to organelle
microtubule anchoring
cellular localization
establishment of protein localization
0
1
$-\log_{10}(q)$
No statistically significant enrichment
iPSC-MN_unique
cellular nitrogen compound metabolic process
RNA processing
mRNA metabolic process
mRNA processing
RNA splicing
0 1 2 3 4 5 6 7
$-\log_{10}(q)$

Figure S4: Related to Figure 4. Switching AS events are enriched for transcriptome and post-transcriptional regulation GO terms.

A. AS events change modalities during iPSC to NPC transition. A total of 7,962 AS events was identified as overlapping events in both iPSCs and MNs. Notably, ≈ 82% of excluded events in iPSCs remained in excluded modality, and ≈ 84% of included events in iPSCs remained as included in NPCs. In contrast 42% of bimodal events in iPSCs switch to either included or excluded modalities in NPCs.

B. Overlapping events shared by all three populations, the events changing between iPSCs to NPCs (light green) and iPSCs to MNs (dark green). Venn diagram show the overlap between the two sets of switching AS events and GO function terms for each section of switching events.

Figure S5

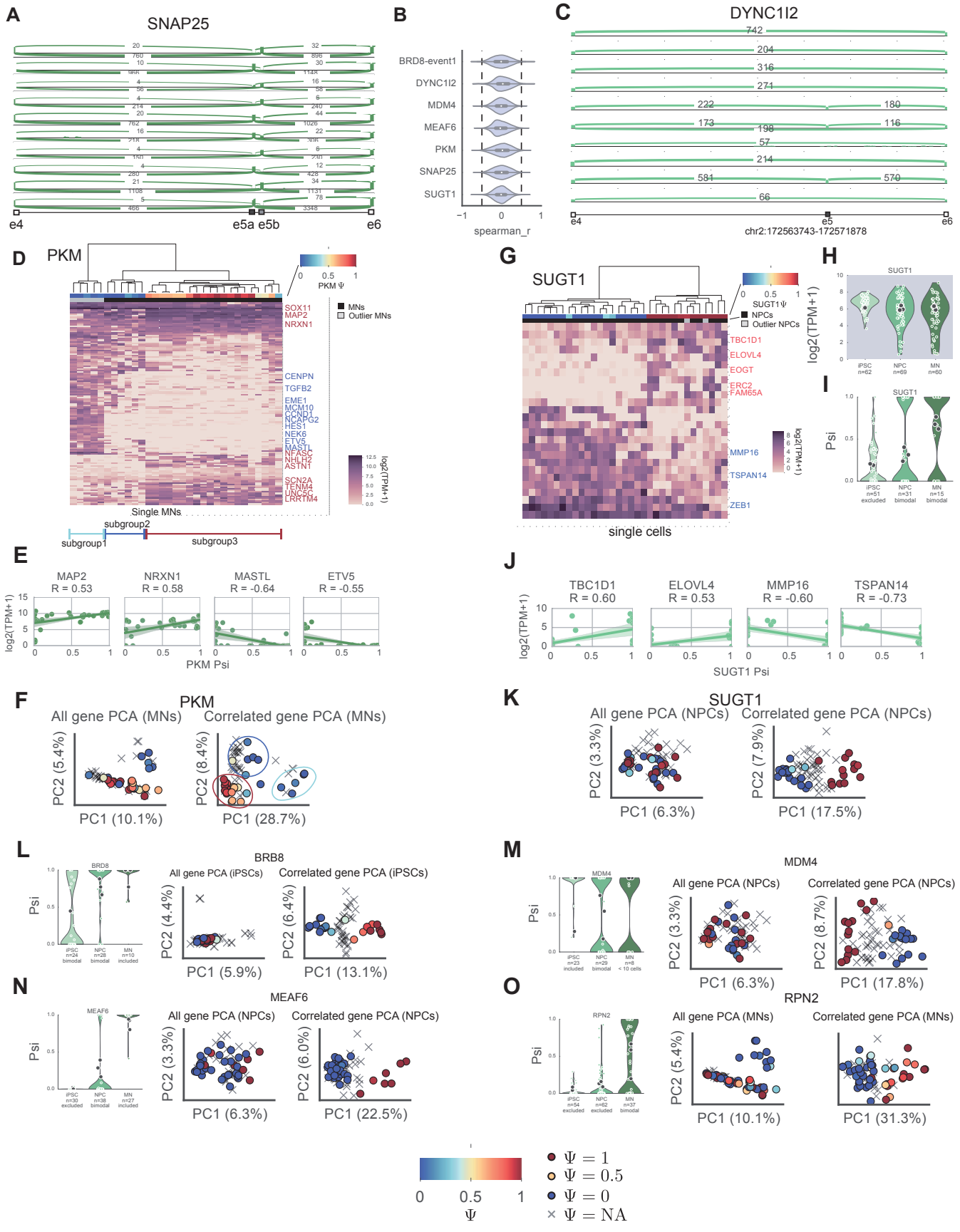

A
SNAP25
e4
e5a e5b
e6
B
BRD8-event1
DYNC1I2
MDM4
MEAF6
PKM
SNAP25
SUGT1
spearman_r
C
DYNC1I2
chr2:172563743-172571878
D
PKM
PKM Ψ
MNs
Outlier MNs
SOX11
MAP2
NRXN1
CENPN
TGFB2
EME1
MCM10
CCND1
NCAPG2
HES1
NEK6
ETV5
MASTL
NFASC
NHLH2
ASTN1
SCN2A
TENM4
UNC5C
LRRTM4
log2(TPM+1)
Single MNs
subgroup2
subgroup1
subgroup3
E
MAP2
R = 0.53
NRXN1
R = 0.58
MASTL
R = -0.64
ETV5
R = -0.55
log2(TPM+1)
PKM Psi
F
PKM
All gene PCA (MNs)
Correlated gene PCA (MNs)
PC2 (5.4%)
PC1 (10.1%)
PC2 (8.4%)
PC1 (28.7%)
G
SUGT1
SUGT1 Ψ
NPCs
Outlier NPCs
TBC1D1
ELOVL4
EOGT
ERC2
FAM65A
MMP16
TSPAN14
ZEB1
log2(TPM+1)
single cells
H
SUGT1
log2(TPM+1)
iPSC n=62
NPC n=69
MN n=60
I
SUGT1
Psi
iPSC n=51 excluded
NPC n=31 bimodal
MN n=15 bimodal
J
TBC1D1
R = 0.60
ELOVL4
R = 0.53
MMP16
R = -0.60
TSPAN14
R = -0.73
log2(TPM+1)
SUGT1 Psi
K
SUGT1
All gene PCA (NPCs)
Correlated gene PCA (NPCs)
PC2 (3.3%)
PC1 (6.3%)
PC2 (7.9%)
PC1 (17.5%)
L
BRD8
Psi
iPSC n=24 bimodal
NPC n=26 bimodal
MN n=10 included
All gene PCA (iPSCs)
Correlated gene PCA (iPSCs)
PC2 (4.4%)
PC1 (5.9%)
PC2 (6.4%)
PC1 (13.1%)
M
MDM4
Psi
iPSC n=23 included
NPC n=29 bimodal
MN n=8 < 10 cells
All gene PCA (NPCs)
Correlated gene PCA (NPCs)
PC2 (3.3%)
PC1 (6.3%)
PC2 (8.7%)
PC1 (17.8%)
N
MEAF6
Psi
iPSC n=30 excluded
NPC n=38 bimodal
MN n=27 included
All gene PCA (NPCs)
Correlated gene PCA (NPCs)
PC2 (3.3%)
PC1 (6.3%)
PC2 (6.0%)
PC1 (22.5%)
O
RPN2
Psi
iPSC n=54 excluded
NPC n=62 excluded
MN n=37 bimodal
All gene PCA (MNs)
Correlated gene PCA (MNs)
PC2 (5.4%)
PC1 (10.1%)
PC1 (31.3%)
0
0.5
1
Ψ
Ψ = 1
Ψ = 0.5
Ψ = 0
Ψ = NA

Figure S5: Related to Figure 5. Highly variant AS events reveal intricacies of cell states.
A. Read coverage tracks for SNAP25 in MNs. Numbers indicate observed junction reads.
B. Spearman correlation values of a gene's alternative splicing score (Ψ) to gene expression values, with a dotted line at the threshold of R > 0.5.
C. Tracks from NPCs were shown to illustrate the bimodal inclusion of exon 5. Numbers indicate observed junction reads covering this SE in DYNC1I2.
D-F. A multimodal MXE event in PKM as an example to dissect MNs into three subgroups.
D. Genes correlating with Psi of the MXE event containing exon 9 and exon 10 (Figure 1) is able to cluster the MNs into three subgroups. Subgroup 1, mostly composed of outliers identified by k-means clustering (Figure S2), contain characteristic genes for progenitors. Subgroup 2 and 3 are enriched for neuronal genes. Rows represent the genes and columns represent single cells in MNs. Genes detected in MNs and correlated with the Psi, using an emipircally-defined threshold of Spearman's R greater than two standard deviations away from the mean permuted correlation values. Psi/Ψ ranged from 0 (blue) to 0.5 (yellow) to 1 (red). Black and grey: cells designated as qualified cells versus outlier-cells based on k-means clustering. Representative genes enriched in two of the subgroups are highlighted in blue (high with exon 10 inclusion) or red (high with exon 9 inclusion).
E. Example genes enriched in two of the subgroups of MNs. MAP2 and NRXN1 are more highly expressed in cells with Ψ ≈ 1; ETV5 and MASTL are more highly expressed in cells with Ψ ≈ 0. Psi scores of the MXE in PKM is plot on x-axis and $\log_2$(TPM + 1) of indicated genes is plot on y-axis.
F. Genes correlating with Psi is able to separate the three subgroups in MNs. Left, PCA using all detected genes in MNs. Right, PCA using genes correlating with Psi.
G-K. A bimodal SE event in SUGT1 as an example to dissect NPCs into two subgroups.
G. Genes correlating with Psi of the SE event cluster the NPCs into two subgroups. Genes detected in NPCs and correlated with the Psi. Blue: cells with Psi around 0. Red: cells with Psi around 1. Light Blue to yellow: cells with Psi around 0.5. Black and grey: cells designated as qualified cells versus outlier cells based on k-means clustering. Representative genes enriched in two of the subgroups are highlighted in blue (high upon exon exclusion) or red (high upon exon inclusion).
H. Expression of SUGT1 in the three populations.
I. Psi distribution of a SE event (lower) in SUGT1 in the three populations. This event is excluded in iPSCs, and bimodal in both NPCs and MNs.
J. Example genes enriched in the two subgroups of NPCs. TBC1D1 and ELOVL4 are more highly expressed in cells with Psi ≈ 1; MMP16 and TSPAN14 are more highly expressed in cells with Psi 0. Psi scores of the SE event in SUGT1 is plot on x-axis and $\log_2$(TPM + 1) of indicated genes is plotted on y-axis.
K. Only genes correlating with Psi is able to separate the two subgroups in NPCs. Left: PCA using all detected genes in NPCs. Right: PCA using genes correlating with Psi.
L-O. PCA using all detected genes in perspective population fail to identify substructures of seemingly homogenous cells (left panel). PCA using gene correlating with each AS event (right panel) is able to identify the delicate substructures of cells.
L. Bimodal SE event in BRD8 reveals iPSC substructure.
M. Bimodal SE event in MDM4 reveals NPC substructure.
N. Bimodal SE event in MEAF6 reveals NPC substructure.
O. Bimodal SE event in RPN2 reveals MN substructure.

Figure S6

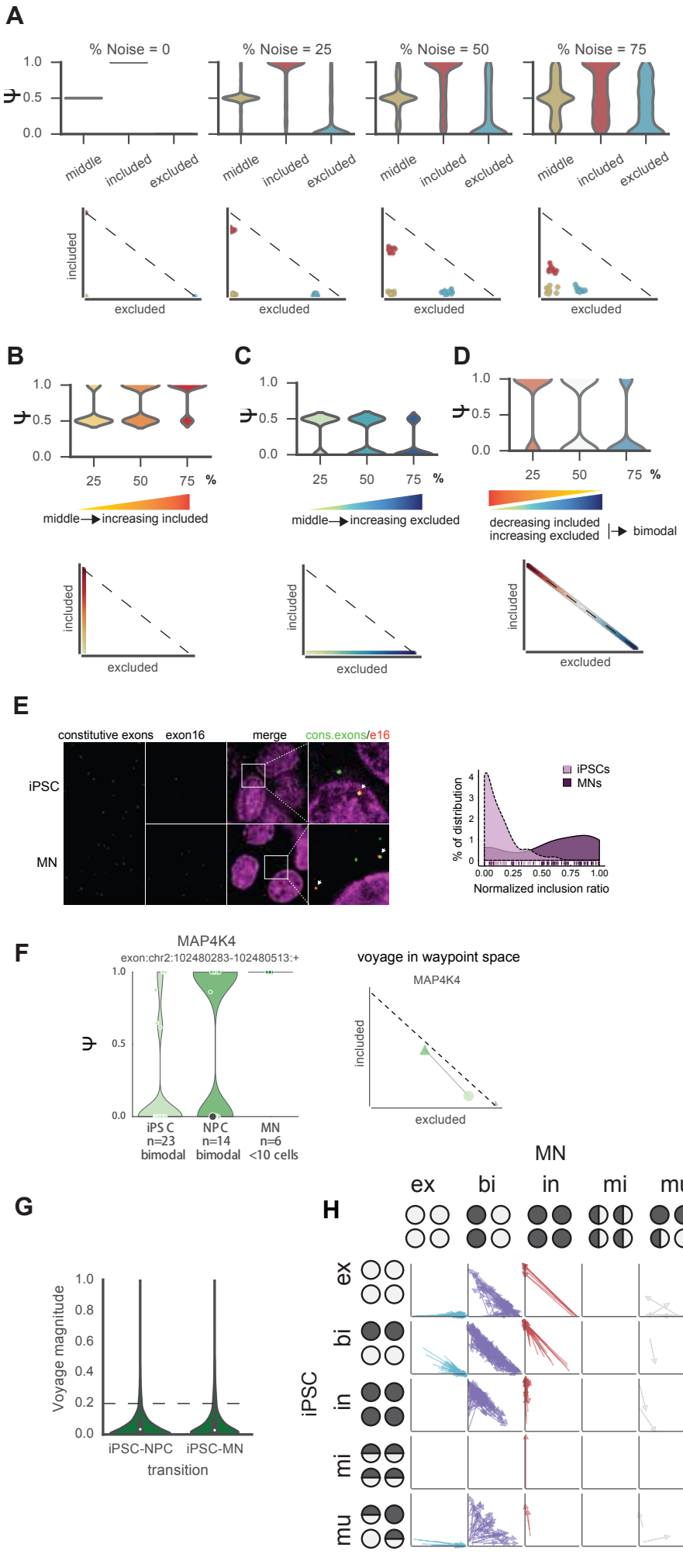

A
% Noise = 0
% Noise = 25
% Noise = 50
% Noise = 75
middle
included
excluded
B
C
D
25
50
75
%
middle→increasing included
middle→increasing excluded
decreasing included
increasing excluded
bimodal
E
constitutive exons
exon16
merge
cons.exons/e16
iPSC
MN
iPSCs
MNs
% of distribution
Normalized inclusion ratio
F
MAP4K4
exon:chr2:102480283-102480513:+
iPSC
n=23
bimodal
NPC
n=14
bimodal
MN
n=6
<10 cells
voyage in waypoint space
G
Voyage magnitude
iPSC-NPC
iPSC-MN
transition
H
MN
ex
bi
in
mi
mu
iPSC

Figure S6: Related to Figure 6. Overview of *bonvoyage*.

A-D. Datasets used to test *bonvoyage*. Uniform random noise was added in 5% intervals to all datasets, up to 95% noise, for 100 iterations at each noise level.

A. Perfect middle, included, and excluded modalities, with added noise. Only 0%, 25%, 50% and 75% noise levels are shown for brevity. Top, averaged violinplots for all features at a given level of noise. Bottom, waypoint space of all features at the specified noise level.

B. Maybe middle-included modalities, created with every combination of 0.5 and 1.0 values. Only the 0% noise dataset is shown for brevity. Top, violinplots, bottom, waypoint plots.

C. Maybe excluded-middle modalities, created with every combination of 0.0 and 0.5 values. Only the 0% noise dataset is shown for brevity. Top, violinplots, bottom, waypoint plots.

D. Maybe bimodal modalities, created with every combination of 0 and 1 values. Only the 0% noise dataset is shown for brevity. Top, violinplots, bottom, waypoint plots.

E-F. Validation of a SE event in MAP4K4 by smRNA-FISH.

E. MAP4K4 smRNA-FISH. Left, probe sets are designed for constitutive exons and alternative exon 16. Exon 16 is excluded in iPSCs (n = 113, light purple with dashed line) and become more included in MNs (n = 68, dark purple with solid outline. Arrows point out foci overlapped for both constitutive and exon 16 probes. Normalized inclusion ratio is calculated by percentage of e16 probes co-localized with constitutive probes/constitutive probes, and resulting percentage is normalized by 95 percentage of the maximal percentage. Right, quantitation of normalized inclusion of exon 16.

F. MAP4K4 single-cell RNA-Seq. Left, violinplots percent spliced-in inclusion values. Right, waypoint space of exon 16.

G. Magnitude of change in waypoint space (voyages) from iPSC to NPC, and iPSC to MN, with a cutoff shown as a black dashed line at 0.2.

H. Global splicing dynamics between iPSC and MN modalities, visualized as vectors from iPSC to MN in waypoint space. Underlying data is the same as Figure 4a. Color of arrows are coded based on event modalities in MNs.

Figure S7

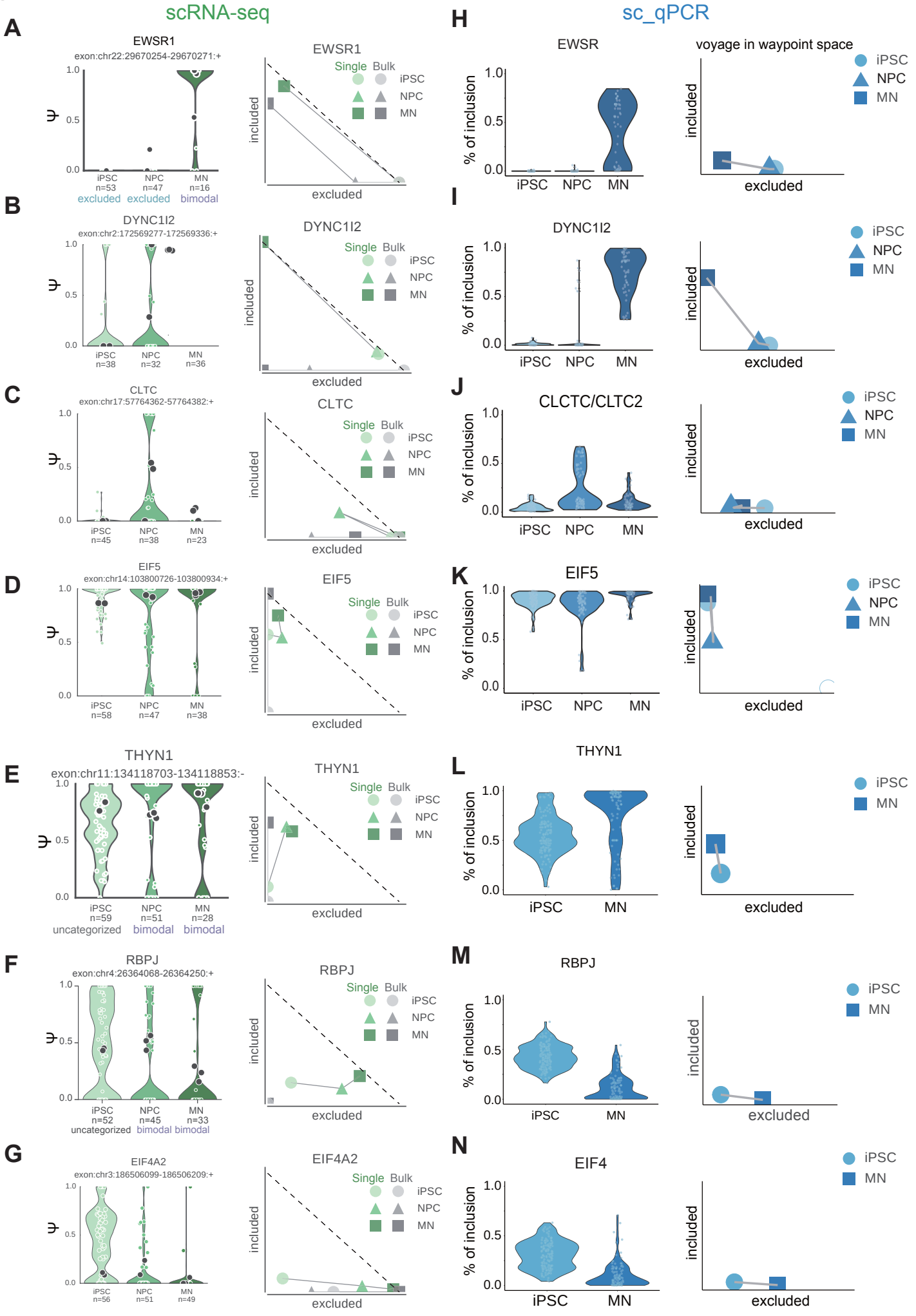

scRNA-seq
sc_qPCR
A
EWSR1
exon:chr22:29670254-29670271:+
iPSC n=53 excluded
NPC n=47 excluded
MN n=16 bimodal
B
DYNC1I2
exon:chr2:172569277-172569336:+
iPSC n=38
NPC n=32
MN n=36
C
CLTC
exon:chr17:57764362-57764382:+
iPSC n=45
NPC n=38
MN n=23
D
EIF5
exon:chr14:103800726-103800934:+
iPSC n=58
NPC n=47
MN n=38
E
THYN1
exon:chr11:134118703-134118853:-
iPSC n=59 uncategorized
NPC n=51 bimodal
MN n=28 bimodal
F
RBPJ
exon:chr4:26364068-26364250:+
iPSC n=52 uncategorized
NPC n=45 bimodal
MN n=33 bimodal
G
EIF4A2
exon:chr3:186506099-186506209:+
iPSC n=56
NPC n=51
MN n=49
Ψ
Single Bulk
iPSC
NPC
MN
included
excluded
H
EWSR
voyage in waypoint space
% of inclusion
I
DYNC1I2
J
CLCTC/CLTC2
K
EIF5
L
THYN1
M
RBPJ
N
EIF4

Figure S7: Related to Figure 7. Validation of alternative splicing events by sc-qPCR.
A-G. Distribution of alternative exon inclusion by scRNA-Seq for indicated events in EWSR1 (A.), DYNC1I2 (B.), CLTC/CLCT2 (C.), EIF5 (D.), THYN1 (E.), RBPJ (F.), and EIF4A2 (G.), shown in violin plots (left) and in waypoint plots (right). Percent spliced-in (Psi/Ψ) is calculated based on single cell RNA-seq data, illustrated in green. Black dots indicate bulk samples (1,000 cells) for each cell type.
H-N. Distribution of percentage of inclusion by sc-qPCR of indicated events EWSR1 (H.), DYNC1I2 (I.), CLTC/CLCT2 (J.), EIF5 (K.), THYN1 (L.), RBPJ (M.), and EIF4A2 (N.), shown in violin plot (left) and waypoint plot (right), illustrated in blue.

Protocols

Protocol 1: scRNA-seq protocol is adapted based on manufacture instruction (Fluidigm).

1. Prepare single cell suspension
   a. iPSCs, NPCs and MNs are washed with PBS, treated with accutase (Stem Cell Biotech) for 5-10 mins at 37°C until cells dislodge from the plate with gentle tapping.
   b. Carefully pipette 3-5 times to generate single cell suspension.
   c. Neutralize the reaction with culture medium and filter cell suspension with 40μm cell strainer (Fisher Scientific).
   d. Centrifuge at 200g for 5 min to collect cells in 15ml Falcon tube (Fisher Scientific), resuspend in cell culture media and filter with 40μm cell strainer
   e. Count cells and dilute cells to approximately 400 cells/μl.

2. Prime the IFC chip and load the cells
   a. Prime the IFC based on manufacture instruction, which takes 10 minutes.
   b. Prepare cell loading mix by gently pipette 3 times.

Single cell suspension (~400 cells/ul)	30ul
Cell suspension reagent (Fluidigm)	20ul

   c. Add 10μl of cell loading mix into the IFC after priming. Note, only 5ul of mix will go into the chip.
   d. Start the cell load and optional live/dead staining step. Run the script **mRNA Seq: Cell Load& Stain**, which takes 60 minutes.
3. Prepare master mixes and load the chip
   a. Cell lysis mix

Reagents	**Volume (μl)**
Loading Reagent (Fluidigm)	1.0
RNase Inhibitor (Takara Bio USA)	0.5
3' SMART CDS Primer IIA (Takara Bio USA)	7.0
Dilution Buffer (Takara Bio USA)	11.5
Total	20

   Pipet the cell lysis mix a few times and keep on ice.

   b. RT mix

Reagents	**Volume (μl)**
Loading Reagent (Fluidigm)	1.2
5X First-Strand Buffer (Takara Bio USA)	11.2
DTT (Takara Bio USA)	1.4
dNTP mix (each NTP at 10mM)	5.6
SMARTer IIA Oligonucleotide (Takara Bio USA)	5.6
RNAse Inhibitor (Takara Bio USA)	1.4
SMARTScribe Reverse Transcriptase (Takara Bio USA)	5.6
Total	**32**

   Vortex the RT mix for 3 seconds and centrifuge briefly, then keep on ice.
   c. cDNA amplification mix

Reagents	**Volume (μl)**
PCR-Grade Water	63.5
10X Advantage 2 PCR Buffer	10.0
50X dNTP Mix	4.0

IS PCR primer	4.0
50X Advantage 2 Polymerase Mix	4.0
Loading Reagent (Fluidigm)	4.5
Total	**90.0**

Vortex the cDNA amplication mix for 3 seconds and centrifuge briefly, then keep on ice.

4. Image cells and load the reagent mixes.
   a. Cells on IFC are imaged in phase contrast using X20 objective. Non-single cells are noted and will be excluded from analysis.
   b. Load the reagent mixes according to manufacture instruction.
   c. Run the script of **mRNA Seq: RT&Amp**
5. Thermal cycling protocols
   a. Lysis

Temperature	Time (min)
72°C	3
4°C	10
25°C	1

   b. RT

Temperature	Time (min)
42°C	90
70°C	10

   c. PCR amplification

Temperature	Time	Cycle
95°C	1 min	1
95°C	20 sec	5
58°C	4 min	
68°C	6 min	
95°C	20 sec	9
64°C	30 sec	
68°C	6 min	
95°C	30 sec	7
64°C	30 sec	
68°C	7 min	
72°C	10 min	1

6. Harvest cDNA on the next day
   a. Aliquot 10 µl of C1 DNA Dilution reagent into each well of 96 well plate
   b. Transfer ~3 µl of amplicons into 96 well plate according to manufacture instruction.
   c. Harvested cDNA can be stored at -20°C.
7. Single cell library preparation
   a. cDNA can be quantitated by PicoGreen and diluted with C1 Harvest Reagent to 0.10-0.3ng/ul. Choose the dilution ratio that will make most of the samples into the desired concentration.
   b. Transfer 2ul of cDNA in to a new 96-sample plate, then add appropriate amount of C1 Harvest Reagent. Seal with adhesive film. Vortex briefly then centrifuge at 1,500 rom for 30 second.
   c. Thaw Tagment DNA Buffer and NT Buffer (Nextera XT DNA Sample Preparation Kit) to room temperature. Make sure there is no precipitate in NT buffer, otherwise vortex until it's clear.

d. Prepare tagmentation mix:

Reagent	Vol per sample (µl)	Vol to prep for 96 samples
Tagment DNA Buffer	2.5	300
Amplification Tagment	1.25	150

e. Vortex at low speed for 20 seconds and spin down.
f. Pipet 3.75 µl of tagmentation mix into a new 96-well plate.
g. Pipet 1.25 µl of the diluted sample into the same 96-well plate.
h. Seal the plate, vortex at medium speed for 20 seconds. Centrifuge at 4,000 rpm for 5 minutes to remove bubbles. Total volume in each well is 5 µl.
i. Run the following tagmentation program:

Temperature	Time
55°C	10 min
10°C	Hold

j. Prepare 1.25 µl of NT Buffer for each well. Thus aliquot 150 µl for 96 wells plus overage.
k. Once sample reach 10 °C, add 1.25 µl NT Buffer to neutralize the reactions. Total volume in each well is 6.25 µl.
l. Seal the plate, vortex and centrifuge at 4,000 rpm for 5 minutes.
m. Add 3.75 µl of Nextera PCR Master Mix (NPM) Buffer in each well. Thus prepare 450 µl for 96 wells plus overage.
n. Pipet 1.25 µl of N7XX primers into each row of the 96-well.
o. Pipet 1.25 ul of S5XX primers into each column of 96 well.
p. The total volume is 12.5 µl. Seal the plate, vortex and centrifuge at 4,000rpm for 2 minutes.
q. Run the following amplification program:

Temperature	Time	Cycles
72°C	3 min	1
95°C	30 sec	1
95°C	10 sec	12
55°C	30 sec	
72 °C	60 sec	
72 °C	5 min	1
10 °C	Hold	1

r. Amplified Libraries can be stored at -20°C for long-term storage.

8. Pool and clean up the library
a. Bring Agencourt AMPure XP beads to room temperature and vortex well.
b. Pool the desired libraries together in a 1.5 ml eppendorf tube. Libraries from dead cells or multi-cells can be eliminated at this step.
c. Add AMPure XP beads in X0.9 volume of the mix libraries. Mix well by pipetting 5 times.
d. Incubate at room temperature for 5 minutes. Place the tube on a magnetic stand for 3 minutes. Carefully remove the supernatant.
e. Add 200 µl of freshly prepared 80% ethanol and incubate for 30 seconds.
f. Remove ethanol. And repeat e-f once.
g. Let the beads dry on magnetic stand for 3 minutes.
h. Elute libraries by adding 50 µl of C1 DNA Dilution Reagent. Vortex for 3 seconds. Incubate at room temperature for 2 minutes.
i. Place the tube on a magnetic stand for 2 minutes. Transfer the supernatant to a new 1.5 ml tube.
j. Repeat b-I for one more time.
k. After final elution, libraries can be inspected on Agilent D1K high sensitivity Screen Tape. The library size distribution is usually around 300-1000bp.

Protocol2: Sc-qRT-PCR protocol is adapted based on manufacture instruction (Fluidigm)

1. Prepare PreAmp primer mix
   a. Resuspend each primer to 200 μM, vortex and spin.
   b. Combine forward and reverse primer (1:1) to make primer set at 100 μM.
   c. Pool PreAmp primers, and final concentration will be 500nM.

Reagents	**Volume**
100 μM primer set	1 μl X 96 = 96 μl
C1 DNA dilution reagent (Fluidigm)	104

2. Prepare single cell suspension
   a. iPSCs, NPCs and MNs were washed with PBS, treated with accutase (Stem Cell Biotech) for 5-10 mins at 37°C until cells dislodged from the plate with gentle tapping.
   b. Carefully pipette 3-5 times to generate single cell suspension.
   c. Neutralize the reaction with culture medium and filter cell suspension with 40μm cell strainer (Fisher Scientific).
   d. Centrifuge at 200g for 5 min to collect cells in 15ml Falcon tube (Fisher Scientific), resuspend in cell culture media and filter with 40μm cell strainer.
   e. Count cells and dilute cells to approximately 400 cells/μl.

3. Prime the IFC chip and load the cells
   a. Prime the IFC based on manufacture instruction. Run the script **Prime**, which takes 10 minutes.
   b. Prepare cell loading mix by gently pipette 3 times.

Single cell suspension (~400 cells/ μl)	30 μl
Cell suspension reagent (Fluidigm)	20 μl

   c. Add 10μl of cell loading mix into the IFC after priming. Note, only 5ul of mix will go into the chip.
   d. Start the cell load and optional live/dead staining step. Run the script **PreAmp: Cell load & Stain**, which takes 60 minutes.
4. Prepare the following reaction mixes:
   a. Prepare lysis Mix

Reagents	**Volume (μl)**
Singe-Cell Lysis Solution (Takara Bio USA)	12.75
C1 Lysis Plus Reagent (Fluidigm)	4.35
Total	17.2

   b. Prepare RT Mix

Reagents	**Volume (μl)**
Stop Solution (Takara Bio USA)	1.94
Single-Cell ViLO RT Mix (Takara Bio USA)	5.84
Single-cell SuperScript RT (Takara Bio USA)	3.62
C1 loading Reagent (Fluidigm)	0.6
Total	12.0

   c. Prepare PreAmp Mix

Reagents	**Volume (μl)**
Single-Cell PreAmp Mix	12.0
C1 PreAmp Dilution Reagent (Fluidigm)	42.0
500nM pooled PreAmp Primer Mix	6.0

Total	60.0

5. Image cells and load the reagent mixes.
   a. Cells on IFC are imaged in phase contrast using X20 objective. Non-single cells are noted and will be excluded from analysis.
   b. Load the reagent mixes according to manufacture instruction.
   c. Run the script of **PreAmp: Preamp.**
6. Thermal cycling protocols
   a. RT

Temperature	Time
25 °C	10 min
42 °C	60 min
85 °C	5 min

   b. Pre-amplification

Stage	Temp	Time	Cycles
Enzyme activation/RT inactivation	95 °C	10 min	1
Denature	95 °C	25 sec	18
Anneal/Extend	60 °C	4 min 10 sec	
Hold	4 °C		

7. Harvest the amplicons on the next day
   a. Pipe 25 µl of C1 DNA Dilution Reagent into a new 96-well plate.
   b. Transfer the amplicons from C1 chip to 96-well plate according to manufacture instruction.
   c. Harvested cDNA can be stored at -20°C.
8. Prepare Dynamic Array for BioMark HD
   a. Inject control line fluid into Dynamic Array according to manufacture instruction.
   b. Load the chip onto BioMark load to prime. This takes approximately 10 minutes.
9. Prepare the following reaction mixes
   a. Prepare 10X assays

Reagents	**Vol. for one reaction with overage µl**
100 µM primer sets	0.2
1X DNA suspension Buffer (Teknova)	1.8
2X Assay Loading Reagent (Fluidigm, PN 100-7611)	2.0
Total	4.0

   b. Prepare sample mix
   Mix SsoFast EvaGreen Supermix with Delta Gene Sample Reagent first for intended number of samples (eg. 96). Prepare master mix with extra overage of 1.2 time. Aliquot 2.2µl of master mix in a 96 well and then add 1.8 µl of PreAmp cDNA.

Reagents	**Vol. for one reaction with overage µl**
2X SsoFast EvaGreen Supermix with low ROX (Bio-Rad)	2.0
Delta Gene Sample Reagent (Fluidigm PN 100-6653)	0.2
PreAmp cDNA	1.8
Total	4.0

   c. Load 3 µl of sample mix and 3 ul of assay mix into Dynamic array according to manufacture instruction. Make sure not to introducing bubbles. Do not go pass the 1st stop on the pipette.

10. Load the Dynamic array onto Biomark HD loader and run the load script, which takes 90 minutes.
11. Load the Dynamic array onto BioMark HD to Data collection.

Protocol 3: smRNA-FISH protocol is adapted based on instruction from BioSearch Technologies

1. Fixation of cells
   a. iPSCs and MNs are grown on matrigel-coated coverglass inside of a 12-well plate.
   b. Aspirate growth medium and wash cell with PBS.
   c. Add 1ml of fixation buffer. Incubate at room temperature for 10 minutes.

37% Formaldehyde solution	1 ml
10XPBS (RNase-free)	1 ml
Nuclease-free water	8 ml
Total	10 ml

   d. Wash twice with 1ml PBS
   e. Immerse cells in 1ml of 70% ethanol for at least 1 hour at 4°C to permeabilize cells.
   f. Cells can be store in 70% ethanol at 4°C for up to a week.
2. Hybridization with probe sets
   a. Probe set is re-dissolved in TE buffer at a stock of 12.5μM.
   b. To prepare hybridization buffer containing probe, add 1μl of probe stock to 100μl of hybridization buffer, and the working probe solution is of 125nM. If doing co-staining, add 1μl of each probe to 100μl of hybridization buffer.
   c. Aspirate the 70% ethanol off the coverglass
   d. Add 1ml of Wash Buffer A and incubate at room temperature for 2-5 minutes. Make fresh before each experiment.

RNA FISH Wash Buffer A (Biosearch Technologies Cat# SMF-WA1-60)	2 ml
Nuclease-free water	7 ml
Deionized	1 ml
Total	10 ml

   e. Assemble humidified chamber in 150mm tissue culture plate. Line a water- saturated Whatman paper at the bottom of the dish. Place a single layer of Parafilm on top of the paper.
   f. Add 100ul hybridization buffer containing probe onto the Parafilm. Gently transfer the coverglass, cells side down, onto the drop of hybridization buffer.
   g. Incubate in the dark at 37°C for 4 hour to overnight.
   h. Transfer the coverglass, cells side up, to a fresh 12-well plate containing 1ml of Wash Buffer A.
   i. Incubate in dark at 37°C for 30 minutes.
   j. Aspirate Wash Buffer A, and then add 1ml of DAPI nuclear stain (5ng/ul DAPI in wash buffer A) to counterstain the nuclei.
   k. Incubate in the dark at 37°C for 30 minutes.
   l. Aspirate the DAPI staining, and then add 1ml of Wash Buffer B. Incubate at room temperature for 2-5 minutes.
   m. Wash once with PBS.

n. Mount the coverglass with Vectashield Mounting Medium, and seal the perimeter with clear nail polish. Once dry, the coverglass is ready for image.