Supplementary Information for

Improved enzymatic labeling of fluorescent in situ hybridization probes applied to the visualization of retained introns in cells

Wen Xiao[a], Kyu-Hyeon Yeom[a], Chia-Ho Lin[a], Douglas L. Black[a,*]

[a]Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles, CA 90095.

* Corresponding Author

Email: dougb@microbio.ucla.edu

This PDF file includes:

SI Materials and Methods

Cell Culture

C2C12 cells were purchased from ATCC (cat.no CRL1772) and maintained in Dulbecco's Modified Eagle's Medium (DMEM) (Gibco, Invitrogen) supplemented with 10% fetal bovine serum (FBS). C2C12 cells were trypsinized and split onto coverslips in 12-well plates at a concentration of 0.1 million cells per well. After 24 hours growth, the cells were prepared for FISH assays. The E14 mouse embryonic stem cell (mESC) line was cultured as previously described (Yeom et al. 2021; Hooper et al. 1987). Briefly, E14 cells were seeded on 0.1 % gelatin coated 6-well plates with mitotically inactivated mouse embryonic fibroblasts (MEF) as a feeder layer (CF1, Applied Stem Cell, Inc.) in mESC media. The cells went through at least 2 passages from the initial thawing. The mESC cultures were then dissociated and transferred onto coverslips in 0.1 % gelatin-coated 12 well plates at a concentration of 0.1 million per well. The mESC media consisted of DMEM (Fisher Scientific) supplemented with 15 % ESC-qualified fetal bovine serum (Thermo Fisher Scientific), 1x non-essential amino acids (Thermo Fisher Scientific), 1x GlutaMAX (Thermo Fisher Scientific), 1x ESC-qualified nucleosides (EMD Millipore), 0.1 mM β-Mercaptoethanol (Sigma-Aldrich), and 10^3 units/ml ESGRO leukemia inhibitor factor (LIF) (EMD Millipore). 24 hours after seeding, the cells were prepared for FISH.

FISH combined with Immunofluorescence Assays

C2C12 cells were washed once with ice-cold PBSM (1xPBS, 5mM MgCl2), followed by fixation with 4% paraformaldehyde in PBSM for 10 min at room temperature (RT). After a 5-minute wash with ice-cold PBSM, the cells were permeabilized with 0.3% Triton X-

100 in PBSM for 7 min on ice. The cells were washed once with PBSM and blocked with 3% BSA (Fraction V) in PBSM for 0.5 h at RT. The coverslips were then incubated with primary antibody (SRSF2/SC35: Abcam-ab11826 , SRSF1: Thermo 32-4600) in 3% BSA in PBSM for 1 hr at RT. After 3 washes in PBST (1XPBS 0.1% tween 20), secondary antibody (goat-antimouse-Cy3: VWR- 95040-042) diluted in 3% BSA in PBSM was added for 40 min at RT. After 3 washes in PBST, cells were refixed in 4% paraformaldehyde in PBSM for 10 min at RT. After a brief wash with PBSM, cells were equilibrated in 10% formamide in 2xSSC for 30 min. DNA probes were hybridized to cells at a concentration of 1 ng/ul in Hybridization buffer (Biosearch: SMF-HB1-10, 10% formamide added freshly), and stored in a humidified box overnight. Cells were washed once with Wash-buffer A (Biosearch: SMF-WA1-60) at 37C for 30 min followed by washing with Wash-buffer A containing 0.5ug/ml DAPI at 37C for another 30 min. Cells were washed once with Wash-buffer B (Biosearch: SMF-WB1-20) at RT for 5 min and mounted with prolong anti-fade mounting media until completely dry.

Measurement of labeling efficiency by PAGE and Spectrophotometry. The concentrations of original and labeled probes were measured on a Nanodrop spectrophotometer and then adjusted to 10 ng/µL with nuclease-free H_2O. 1 µL of each probe (10 ng) was added to 4 µL loading buffer (30% glycerol), loaded on a 15% polyacrylamide gel, and run for 30 min in TBE buffer at 200 Volt. After washing with H_2O, the gel was stained with SYBR-Gold (Thermo Fisher, s-11494, 1:10,000 dilution) for 15 min at room temperature, washed 2 x 5 minutes with H_2O, and scanned on a Typhoon phosphorimager (Amersham™ Typhoon™ 5). SYBR-Gold was detected on the Cy2

channel, with other channels used according to the excitation of each fluorophore. The degree of labeling was calculated by PAGE using the equation in Figure S2A. The degree of labeling was also calculated for each labeled oligo from the UV-Vis absorbance of each fluorophore and DNA measured on a NanoDrop spectrophotometer, using the formula in Figure S2. Information on probe sequences and DOL determination by PAGE and spectrophotometry are presented in Tables S3-S9.

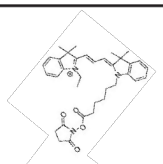
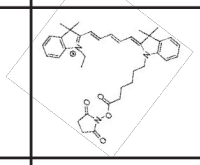
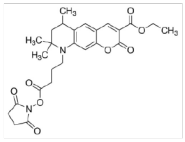
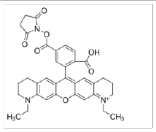
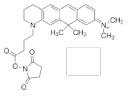
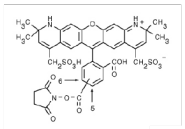
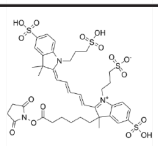
Microscopic image acquisition. Cells on mounted coverslips were imaged on a Leica TCS SP8 light-sheet microscope equipped with a 63x, 1.4 Numerial Aperture (NA) oil-immersion objective (Leica) and a Leica sCMOS-camera in the microscopy core at the California NanoSystems Institute, UCLA. Images were acquired in ~200 nm z-dimension axis steps across ranges of approximately 4 μm and 2 μm for the ES and C2C12 cells, respectively. Images were acquired using appropriate filter sets - DAPI: 358 nm excitation, 461 nm Emission; Quasar570: 548 nm excitation, 566 nm Emission; Quasar670: 647 nm excitation, 670 nm Emission; ATTO488: 500 nm excitation, 520 nm Emission, ATTO565: 564 nm excitation, 590 nm Emission; ATTO647N: 646 nm excitation, 664 nm Emission; Alexa568: 578 nm excitation, 603 nm Emission.

Table S1

Reagents	Amount	Price ($)	Total Reactions	Cost/Reaction ($)	Cost/ hybridization ($)
Oligos (48x)	200 uM each in ~100ul H20	120	>250	<0.48	~0.015
NHS-ester-Dye	1 mg	130-240	75	1.73-3.2	
Amino-11-ddUTP	1 umol	549	100	5.49	
TdT enzyme	2500 U	280	42	6.67	
Total		1079-1189		14.37-15.84	

Reagents and materials required for FISH probe labeling with approximate cost. See detailed information in methods section.

Table S2

Dye-NHS-ester	Structure	Cat.No.
Quasar570		Biosearch Technology FC-1063S-5
Quasar670		Biosearch Technology FC-1065S-5
ATTO488		Sigma-Aldrich 41698-1MG-F
ATTO565		Sigma-Aldrich 72464-1MG-F
ATTO647N		Sigma-Aldrich 18373-1MG-F
Alexa Fluor 568		Fisher Scientific A20003
Alexa Fluor 647		Fisher Scientific A20006

Chemical structures, catolog numbers and vendors of dyes used in this study.

Figure legends

Figure S1. **PAGE gel analysis and quantification of the DOL of dye coupled oligos. (A)** Sequences of oligonucleotides with variable GC content. Oligos containing from 40% GC to 60% GC were selected from an oligo set targeting the Clcn6 exons using the Stellaris probe design tool. **(B)** PAGE analysis of the oligos from A, before and

after the ddUTP and dye labeling steps. The upper panel shows the absorbance of SYBR-Gold. The bottom panel displays the composite absorbance in the SYBR (green) and Quasar 670 (red) channels. **(C)** Quantification of the DOL after the ddUTP and dye labeling steps for the oligos in B. **(D)** Sequences of oligonucleotides with 4 different 3' terminal nucleotides. **(E)** PAGE analysis after the ddUTP and dye labeling steps for the oligos in D. Upper panel: the absorbance of SYBR-Gold. Bottom panel: a composite of the absorbance of SYBR (green) and Quasar 670 (red). **(F)** Quantification of the DOL after the ddUTP and dye coupling steps for the oligos in E. **(G)** PAGE analysis of ATTO488 labeled Neat1 oligos using earlier method. **(H)** Quantification of the DOL by PAGE in G. **(I)** PAGE analysis of ATTO488 and Quasar570 labeled Neat1 oligos using our new method. **(J)** Quantification of the DOL by PAGE in I.

Figure S2. **Equations used to calculate the degree of probe labeling by spectrophotometry and PAGE.** Above: calculation of DOL by spectrophotometer. ε_{dye} stands for the extinction coefficient of dye and cf_{260} represents the relative absorbance of dye at 260 nm. ε_{oligo} is calculated by taking average of the extinction coefficient of oligos in the oligo mixture (see methods and Tables S3-S9).

Below: Calculation of DOL by PAGE. U stands for the remaining SYBR-Gold band intensity of the unlabeled oligos after dye coupling. SO stands for the SYBR-Gold band intensity of the starting oligos prior to dye coupling.

Figure S3. PAGE analysis of probe sets coupled to distinct fluorophores: **(A)** ATTO488, **(B)** ATTO565, **(C)** ATTO647N, **(D)** Alexa568, and **(E)** Alexa647. Upper panels: the fluorescence of SYBR-Gold (grey). Bottom panels: composites of the SYBR-Gold fluorescence in green with that of each individual dye in red. Notably, for ATTO488 **(A)**,

the absorbance of ATTO488 and SYBR-Gold are similar and can be detected with the same excitation wavelength in the upper panel. Absorbance of the other dyes (ATTO565, ATTO647N, Alexa568, Alexa647) does not overlap with SYBR-Gold.

Figure S4. Scatter plots correlating the DOL measurements by spectrophotometer and PAGE for ATTO488 **(A)**, ATTO565 **(B)**, ATTO647N **(C)**, Alexa568 **(D)** and Alexa647 **(E** and **F)** for multiple probe sets. Each dot represents a probe set containing up to 48 oligos (Table S5-S9). The R^2 values and fitted linear equations are shown..

Figure S5. Specificity of FISH probes. **(A)** Scatter plot of fluorescence intensity per pixel between Odd and Even FISH images of Polr2a (See Figure 4A). **(B)** Summary of FISH probes and their performance in ES cells. Detectible intron signal coincident with exon signal in the nucleus was required to be considered as good performance. **(C-L)** FISH images in ES cells using fluorescently labeled FISH probes targeting exons and introns in 5 genes, Cstf1, Neil3, Rab9b, Slc25a28, and Znrf3 (See Fig. S5B).

Figure S6. Correlation of intron and nuclear exon spot numbers for retained and highly spliced introns. Scatterplots correlating the nuclear spot counts of an intron and the exons of the same gene. **(A)** Retained Noc2l intron 12. **(B)** Highly spliced Noc2l intron 4. **(C)** Retained Gabbr1 intron 5. **(D)** Highly spliced Gabbr1 intron 11. **(E)** The PIR (Percent Intron Retained) of the Noc2l and Gabbr1 introns in fractions of ESC determined by RNAseq (chr = chromatin fraction, nuc = soluble nucleoplasmic fraction, cyt = cytoplasmic fraction) (Yeom et al. 2021).

Figure S7. Malat1 colocalizes with speckles. **(A)** Combined FISH and immunofluorescence assays using Malat1 FISH probes (green) and SRSF2 (SC35)

antibody (red) in C2C12 cells. **(B)** Line plot profiles of fluorescent intensity for nuclei in A showing overlap between the Malat1 and SRSF2 signals. **(C)** Combined FISH and immunofluorescence assays using Malat1 FISH probes and SRSF1 antibody in C2C12 cells. **(D)** Line plot profile of fluorescent intensity for the nucleus in C showing overlap between the Malat1 and SRSF1 signals.

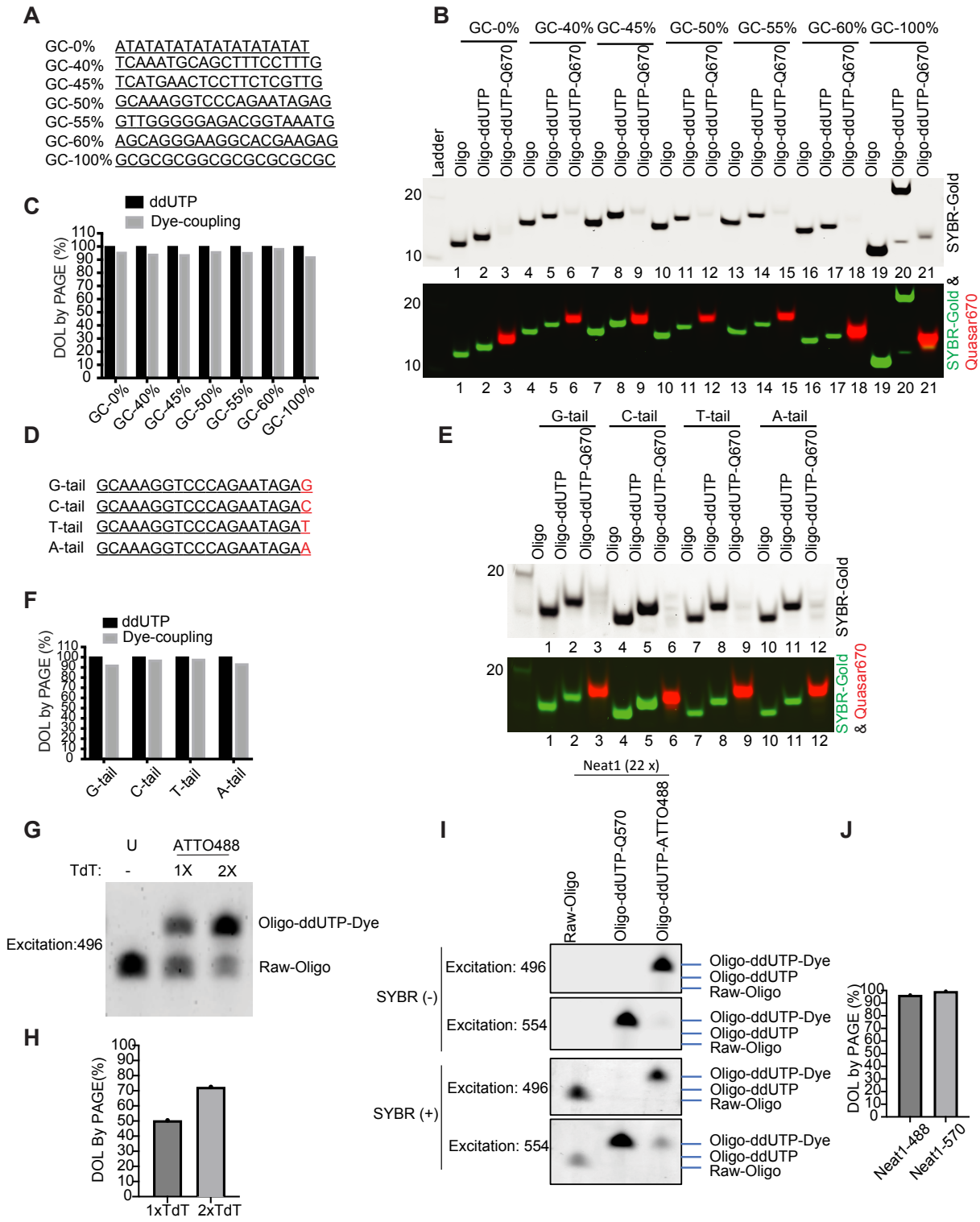
References

Hooper M, Hardy K, Handyside A, Hunter S, Monk M. 1987. HPRT-deficient (Lesch–Nyhan) mouse embryos derived from germline colonization by cultured cells. *Nature* **326**: 292–295.

Mueller F, Senecal A, Tantale K, Marie-Nelly H, Ly N, Collin O, Basyuk E, Bertrand E, Darzacq X, Zimmer C. 2013. FISH-quant: automatic counting of transcripts in 3D FISH images. *Nat Methods* **10**: 277–278.

Yeom K-H, Pan Z, Lin C-H, Lim HY, Xiao W, Xing Y, Black DL. 2021. Tracking pre-mRNA maturation across subcellular compartments identifies developmental gene regulation through intron retention and nuclear anchoring. *Genome Res* **31**: 1106–1119.

Fig. S1


A
GC-0% ATATATATATATATATATAT
GC-40% TCAAATGCAGCTTTCCTTTG
GC-45% TCATGAACTCCTTCTCGTTG
GC-50% GCAAAGGTCCCAGAATAGAG
GC-55% GTTGGGGGAGACGGTAAATG
GC-60% AGCAGGGAAGGCACGAAGAG
GC-100% GCGCGCGGCGCGCGCGCGC
B
GC-0%
GC-40%
GC-45%
GC-50%
GC-55%
GC-60%
GC-100%
Ladder
Oligo
Oligo-ddUTP
Oligo-ddUTP-Q670
SYBR-Gold
SYBR-Gold & Quasar670
C
ddUTP
Dye-coupling
DOL by PAGE (%)
GC-0%
GC-40%
GC-45%
GC-50%
GC-55%
GC-60%
GC-100%
D
G-tail GCAAAGGTCCCAGAATAGAG
C-tail GCAAAGGTCCCAGAATAGAC
T-tail GCAAAGGTCCCAGAATAGAT
A-tail GCAAAGGTCCCAGAATAGAA
E
G-tail
C-tail
T-tail
A-tail
F
ddUTP
Dye-coupling
DOL by PAGE (%)
G-tail
C-tail
T-tail
A-tail
G
U
ATTO488
TdT:
-
1X
2X
Excitation:496
Oligo-ddUTP-Dye
Raw-Oligo
H
DOL By PAGE(%)
1xTdT
2xTdT
I
Neat1 (22 x)
Raw-Oligo
Oligo-ddUTP-Q570
Oligo-ddUTP-ATTO488
SYBR (-)
SYBR (+)
Excitation: 496
Excitation: 554
Oligo-ddUTP-Dye
Oligo-ddUTP
Raw-Oligo
J
DOL by PAGE (%)
Neat1-488
Neat1-570

166 **Fig. S2**

A

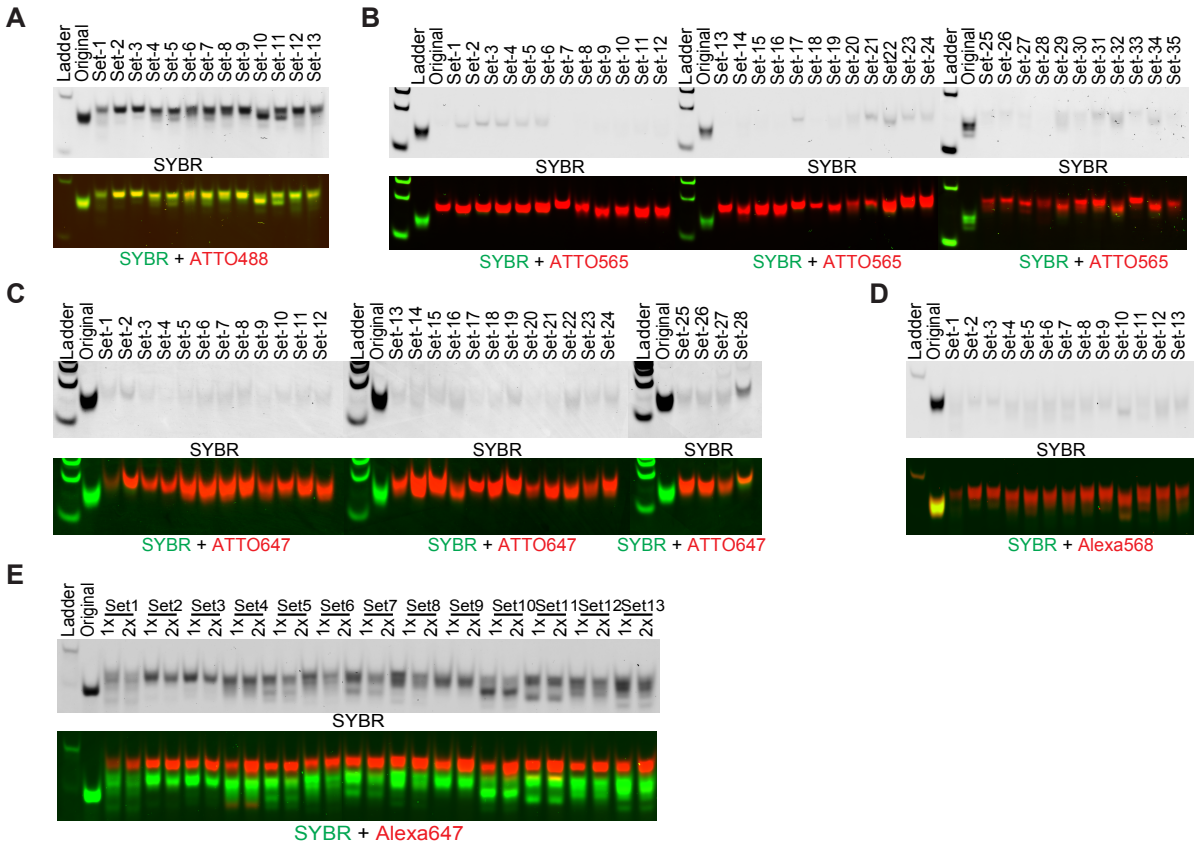
$$\text{DOL(UV-Vis)} = \frac{OD_{dye}}{\varepsilon_{dye}} \times \frac{(\varepsilon_{oligo} + 9000)}{(OD_{oligo} - cf_{260} \times OD_{dye})}$$

$$\text{DOL(PAGE)} = 1 - \frac{U}{SO}$$

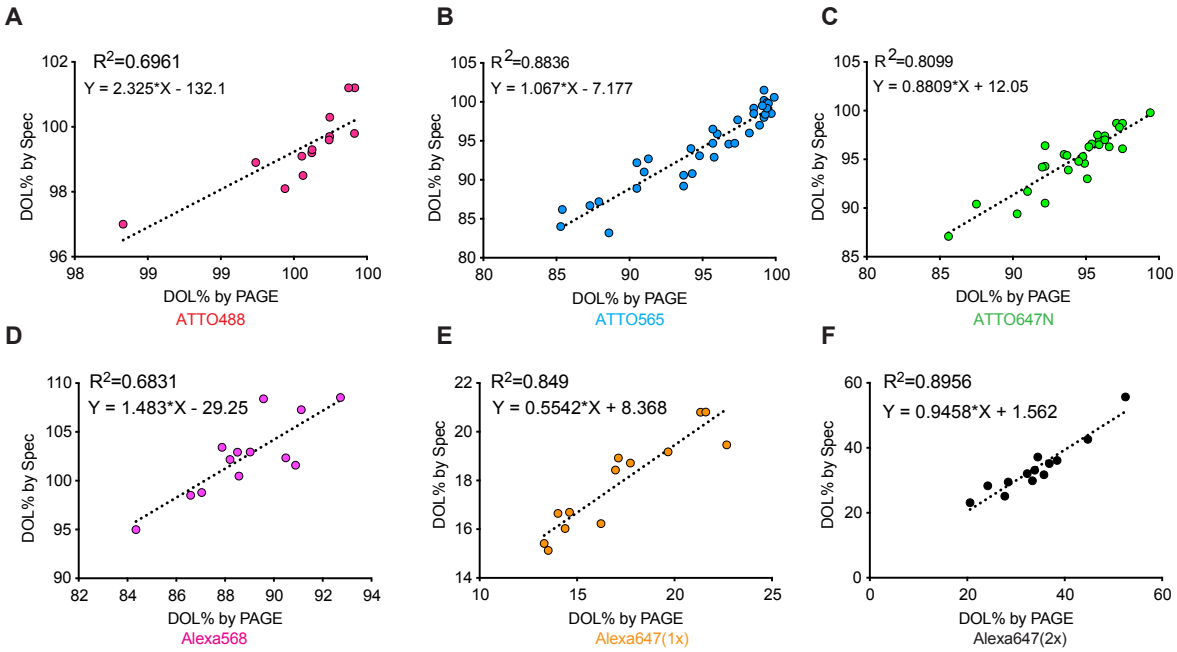
167

168

Fig. S3


A
Ladder Original Set-1 Set-2 Set-3 Set-4 Set-5 Set-6 Set-7 Set-8 Set-9 Set-10 Set-11 Set-12 Set-13
SYBR
SYBR + ATTO488
B
Ladder Original Set-1 Set-2 Set-3 Set-4 Set-5 Set-6 Set-7 Set-8 Set-9 Set-10 Set-11 Set-12
Ladder Original Set-13 Set-14 Set-15 Set-16 Set-17 Set-18 Set-19 Set-20 Set-21 Set22 Set-23 Set-24
Ladder Original Set-25 Set-26 Set-27 Set-28 Set-29 Set-30 Set-31 Set-32 Set-33 Set-34 Set-35
SYBR
SYBR + ATTO565
SYBR + ATTO565
SYBR + ATTO565
C
Ladder Original Set-1 Set-2 Set-3 Set-4 Set-5 Set-6 Set-7 Set-8 Set-9 Set-10 Set-11 Set-12
Ladder Original Set-13 Set-14 Set-15 Set-16 Set-17 Set-18 Set-19 Set-20 Set-21 Set-22 Set-23 Set-24
Ladder Original Set-25 Set-26 Set-27 Set-28
SYBR
SYBR + ATTO647
SYBR + ATTO647
SYBR + ATTO647
D
Ladder Original Set-1 Set-2 Set-3 Set-4 Set-5 Set-6 Set-7 Set-8 Set-9 Set-10 Set-11 Set-12 Set-13
SYBR
SYBR + Alexa568
E
Ladder Original Set1 1x 2x Set2 1x 2x Set3 1x 2x Set4 1x 2x Set5 1x 2x Set6 1x 2x Set7 1x 2x Set8 1x 2x Set9 1x 2x Set10 1x 2x Set11 1x 2x Set12 1x 2x Set13 1x 2x
SYBR
SYBR + Alexa647

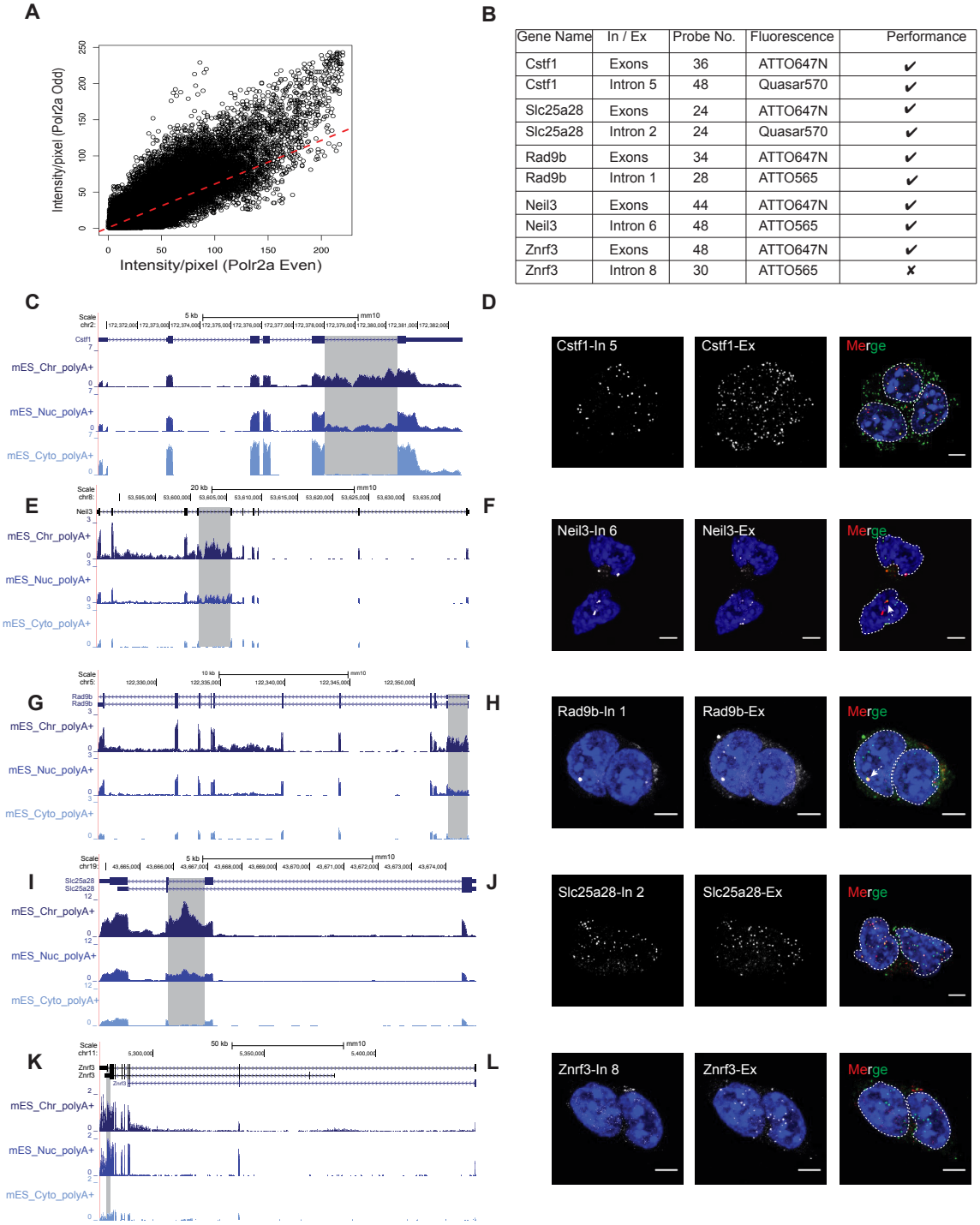
178 **Fig. S4**



179

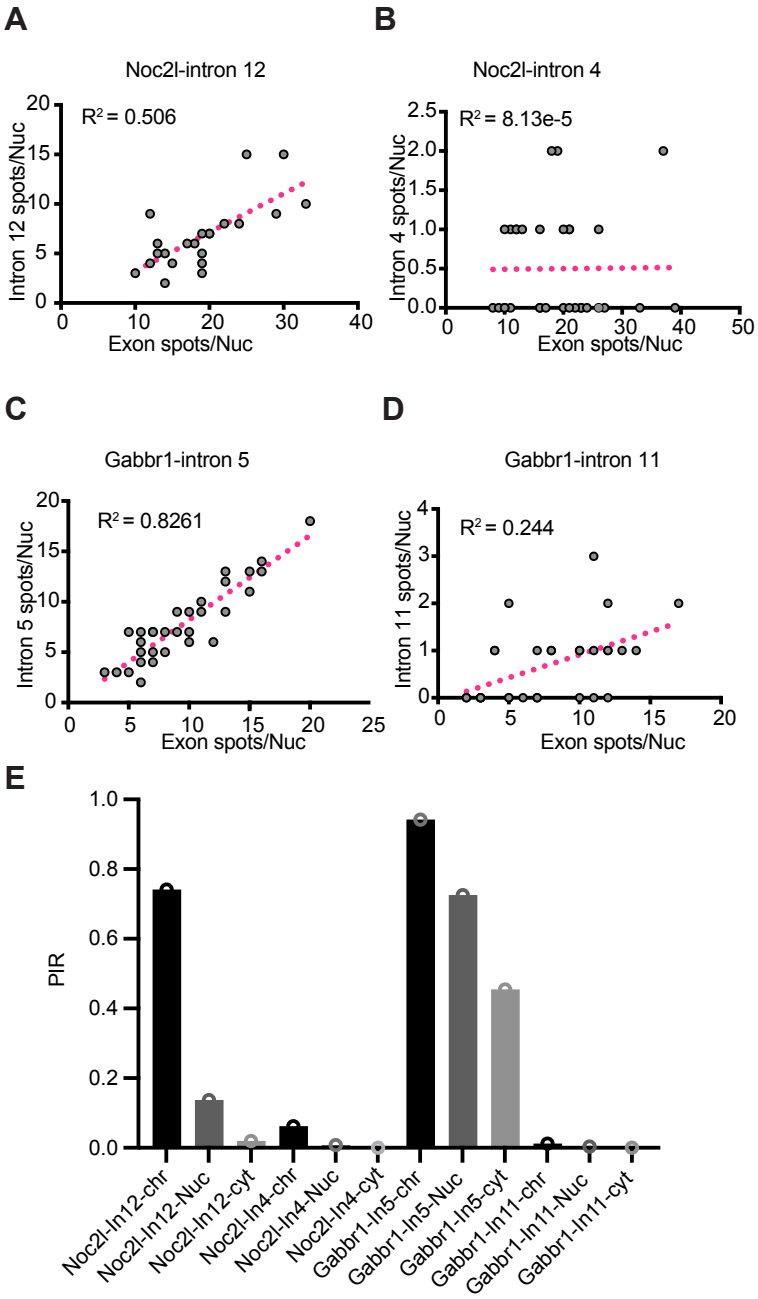
180

Fig. S5

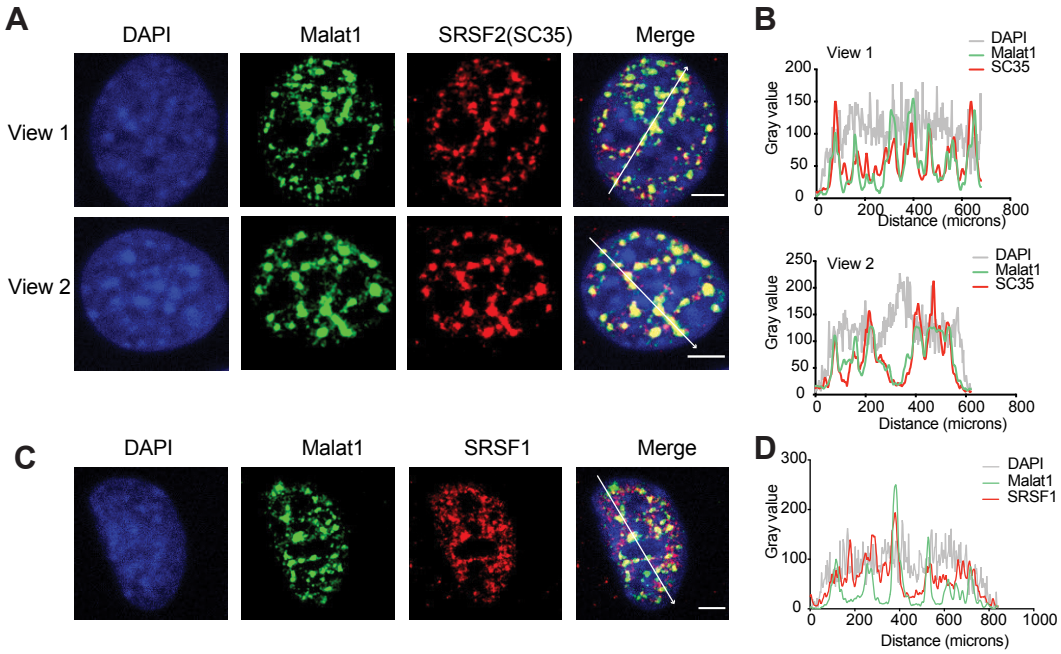


Gene Name	In / Ex	Probe No.	Fluorescence	Performance
Cstf1	Exons	36	ATTO647N	✔
Cstf1	Intron 5	48	Quasar570	✔
Slc25a28	Exons	24	ATTO647N	✔
Slc25a28	Intron 2	24	Quasar570	✔
Rad9b	Exons	34	ATTO647N	✔
Rad9b	Intron 1	28	ATTO565	✔
Neil3	Exons	44	ATTO647N	✔
Neil3	Intron 6	48	ATTO565	✔
Znrf3	Exons	48	ATTO647N	✔
Znrf3	Intron 8	30	ATTO565	✘

Fig. S6


A
Noc2l-intron 12
R² = 0.506
Intron 12 spots/Nuc
Exon spots/Nuc
B
Noc2l-intron 4
R² = 8.13e-5
Intron 4 spots/Nuc
Exon spots/Nuc
C
Gabbr1-intron 5
R² = 0.8261
Intron 5 spots/Nuc
Exon spots/Nuc
D
Gabbr1-intron 11
R² = 0.244
Intron 11 spots/Nuc
Exon spots/Nuc
E
PIR
Noc2l-In12-chr
Noc2l-In12-Nuc
Noc2l-In12-cyt
Noc2l-In4-chr
Noc2l-In4-Nuc
Noc2l-In4-cyt
Gabbr1-In5-chr
Gabbr1-In5-Nuc
Gabbr1-In5-cyt
Gabbr1-In11-chr
Gabbr1-In11-Nuc
Gabbr1-In11-cyt

188 **Fig. S7**



189

190