Autonomous combinatorial color barcoding for multiplexing single molecule RNA visualization

Yong-Sheng Cheng[1*], Yue Zhuo[1], Katharina Hartmann[1], Peng Zou[1], Gözde Bekki[1], Heike Alter[1] & Hai-Kun Liu[1*]

[1]Division of Molecular Neurogenetics, German Cancer Research Center (DKFZ), DKFZ-ZMBH Alliance, Im Neuenheimer Feld 280, Heidelberg, 69120, Germany.
[*]Correspondence: yongsheng.cheng@dkfz.de and l.haikun@dkfz.de

Abstract

Single molecular fluorescence in situ hybridization (smFISH) detects RNA transcripts with spatial information and digital molecular counting. However, the broad usage of smFISH is still hindered by the complex chemical probe conjugation or microscopy set-up, especially for investigating multiple gene expression. Here we present a multiple fluorophore enzymatical labeling method (termed HuluFISH) for smFISH probes to achieve flexible combinatorial color barcoding in single hybridization step. The multiplex capacity of HuluFISH follows an exponential growth with the increase of the number of fluorophore types. We demonstrate that this method can be used to detect cellular heterogeneity in embryonic mouse brain on single cell level.

Introduction

Since the invention of *in situ* hybridization[1], it has been continuously advancing our understanding of gene expression with spatial information. The smFISH technology pioneered by Robert Singer's lab[2] and further developed by Raj et al.[3], brings *in situ* RNA quantification into single molecular and digital manner. Nevertheless, the limited choices of single fluorophore on probes cannot cope with the increasing demand of simultaneous multiple gene detection. Although sequential hybridization has been employed to achieve multiplex gene detection using smFISH[4,5], sophisticated experimental settings hinder its broad applications in the biomedical community. One alternative strategy for increasing the multiplexity beyond fluorophore limit is using combinatorial color barcoding via spectral or spatial separated groups of smFISH probe[6,7]. Current combinatorial barcoding either needs a long gene target for mRNA[6] or only targets intrinsically non-stable introns[7], and thus restricts their applications from detecting the majority of the transcripts (median mouse coding DNA sequence (CDS) length is 1026 bp). Therefore, a combinatorial color barcoding on individual smFISH probe will empower the conventional smFISH with massive color combinations with the same number of probes.

Results

Conventional smFISH probes or its derivatives are using chemical labeling for conjugating a fluorophore to the internal, 5' or 3' end of an unlabeled oligonucleotide pre-equipped with an amine group, which is readily reacting with fluorophores functionized with a N-Hydroxysuccinimide (NHS) ester[2,8,9]. We apply a novel enzymatic fluorophore labeling method, which is based on the usage of T4 DNA ligase (T4DL), for HuluFISH 1.0. It does not require any amine modification of unlabeled gene-specific oligonucleotides (GSO) for HuluFISH probe. As a consequence, it is now possible to cost-effectively synthesize single fluorophore labeled smFISH probes (Figure 1a). Comparing with other enzymatic labeling methods we have tested, T4DL has the most cost-effective design (Suppl. Figure1a). In the T4DL based labeling strategy, only a standard PCR primer quality oligonucleotide is required, and the free 3' hydroxyl group from the GSO is enzymatically conjugated with a common pre-fluorescently-labeled oligonucleotide (termed Hulu), mediated by an adaptor with 4 bp 3' degenerative sequence to facilitate duplex formation (Figure 1a). This new T4DL based chemistry also abolishes the necessity of HPLC purification of smFISH probe (Suppl. Figure1b and 1c). The polyacrylamide gel electrophoresis (PAGE) purified mouse Gapdh probe has comparable detection sensitivity with the commercially available one (Figure 1b).

Currently, the smFISH probe selection is based on melting temperature (Tm)[3] or Gibbs free energy[10], which are not very indicative of probe hybridization efficiency. We developed a pipeline based on Primer3[11] and DECIPHER[12] to design and filter for GSO with high hybridization efficiency, which is a more tangible indicator (Figure 1c). Comparing with the conventional Tm based method, our probe design has better signal-to-noise ratio (SNR) and higher contrast (Figure 1d and 1e). With this new approach, we still have short probe (17-21 bp) to minimize the off-target effect, and a good balance between hybridization capacity and the number of probes we could design for smaller RNA (minimally 24 GSOs). This can be used for customized probe design for any other smFISH methods.

In principle, our T4DL based labeling method also enables multiple fluorophore labeling if the Hulu oligonucleotide is pre-synthesized with multiple fluorophores. However, the technical complexity increases with the number of fluorophores to be incorporated into a single oligonucleotide. Therefore, we extended the T4DL based labeling to multiple-way ligation for incorporating multiple single fluorophore labeled Hulu oligonucleotides (HuluFISH 2.0, Figure 2a). Ligation control experiment shows that HuluFISH 2.0 has a specific ligation product for Gapdh HuluFISH probes, and higher yield compared with the HuluFISH 1.0 (Suppl. Figure 2a). The HuluFISH 2.0 is insensitive to ligation conditions (Suppl. Figure 2b), which demonstrates the

robustness of the its probe preparation over temperature, reaction time, etc. One critical challenge for using multi-colored probe is that when multiple fluorophores are close to each other, they could be quenched by multiple mechanisms, for example self-quenching and Förster resonance energy transfer (FRET)[13,14]. Considering the size limitation of the Hulu oligonucleotide, here we use 15 bp spacing for the individual dye, and an adaptor oligonucleotide annealed with the Hulu oligonucleotide in order to rigidify the ssDNA backbone for dyes (Figure 2b).

Gapdh probe staining without the adaptor masks the FISH signal by fusing dots with high background in all channels for Atto488, Atto565, and Atto647N (Figure 2c). And these dot-like signals 3 channels are not co-localized very well. With the stabilization by the adaptor oligonucleotide, individual clear dots can be obtained in all 3 channels and well co-localized within every channel for Gapdh probes (Figure 2d). Without GSO, the Hulu-adaptor duplex does not generate any dot like signal (Suppl. Figure 2c). With the multiple labeling capacity of our method, we could assign various color combinations to a panel of genes, and decode the dots by counting their appearance in channels (Suppl. Figure 2d). The evolved multiple fluorophore labeling capability with HuluFISH 2.0 extends the conventional smFISH with an autonomous combinatorial color barcoding mechanism. Flurophores in each color combination are covalently linked with individual probe, therefore

the fluorophore stoichiometry is invariable between probes. During imaging acquisition, the intensity ratio between fluorophores will be independent of the brightness of FISH dots. The barcoding capacity simply increases with the exponentials of the channel (fluorophore choice) number n (the theoretical number of combinations is the sum of all color combinations: $\sum_{k=1}^{n}\binom{n}{k} = 2^n - 1$). If the relative ratio of the maximal intensities of each FISH dot among channels can be precisely determined, the number of combinations can be higher.

One of the most interesting applications for smFISH is exploring the multiple gene expression patterns in tissue samples. Just with 3 base colors, the color combinations can be used to detect 7 genes in one round of hybridization. Here we use embryonic day 12.5 (E12.5) mouse telencephalon cryo-section samples to visualize the tissue heterogeneity of these 7 genes (Figure 3a). Simultaneous 7-gene detection shows the molecular heterogeneity of fetal brain neural progenitors *in vivo* (Figure 3b). Hierarchical cluster analysis reveals subgroups of mouse telencephalon neural progenitors on single cell resolution (Figure 3c and 3d).

Discussion

Here, we present the HuluFISH as a new framework for smFISH. HuluFISH has the capability to enzymatically ligate multiple fluorophores to probes, which are designed by filtering for their hybridization efficiency. And this new approach allows us to simultaneously detect genes with the multiplexity that increases exponentially with the number of available microscopy lasers and fluorophore types. With 4 to 5 color channels, it is possible to image 15 to 31 genes in one round of hybridization, which will fulfill a large number of experimental needs in detecting multiple RNA species, without resorting to multiple-step sequential hybridizations or super-resolution microscopy. HuluFISH labeling method is compatible with any other FISH related techniques. In particular, SeqFISH[4] or MERFISH[5] could employ HuluFISH labeling to either reduce the number of hybridization steps for fixed multiplexity or increase the multiplexity within their operational steps. Multiplexing *in situ* quatification of gene expression has become the next frontier in many fields for biomedical research. We believe the broad application of HuluFISH and its derivatives will greatly facilitate the discovery processes like cellular heterogeneity and precise gene expression regulation, in particular for project like the Human Cell Atlas Initiative.

Methods

Cell culture and tissue section preparation

Mouse Hepa 1-6 cells were cultured in DMEM medium with 10% fetal bovine serum and 1 x penicillin/streptomycin. Hepa 1-6 cells were directly grown on coverslip without coating. Embryonic mouse brain tissue cryo-sections were cut at 6 to 10 μm from embryonic day 12.5 C57BL/6J mouse embryo embedded in Tissue-Tek O.C.T. (Sakura, 4583). Adherent Hepa 1-6 cells or cryo-sections were fixed with 4% formaldehyde in PBS for 10 min and then quenched with 135 mM glycine in PBS for 10 min at room temperature. Fixed cells were then washed once with PBS and permeablized in 70% ethanol overnight at 4 ◦C. All water used for FISH related buffers was diethyl pyrocarbonate (DEPC) treated. After permeabilization cells were stored in cryo-protectant (25% glycerol, 25% ethylene glycol, 0.1M phosphate buffer, pH 7.4) at -20 ◦C until FISH staining.

Probe design

smFISH probes based on the conventional design[3] were implemented in a R script to select GSOs first with Primer3[11] to get all possible GSOs without strong secondary structure from the input mRNA sequence using the standard condition for selecting the right_primer in Primer3. Then non-overlapping GSOs were selected with minimally 2 bp gap. For HuluFISH 1.0 probes, all

GSOs from Primer3 were additionally calculated for their hybridization efficiency with DECIPHER package in R[12] under the condition used for staining. And the GSOs were filtered to have hybridization efficiency above 0.9 (maximally 1) and then non-overlapping HuluFISH GSOs were selected as before. For HuluFISH 2.0 GSOs, additional tag sequence was added to their 3' end after their selection. Adaptor, tag for GSO and Hulu sequences were randomly generated and controlled for strong secondary structure by UNAfold[15]. Passed sequences were blasted against a local mouse and human transcript database (ensemble release 87) for less than or equal to 15 bp exact match.

HuluFISH probe labeling and purification

HuluFISH was initially an acronym for Helix-stabilized, unbiased and ligated uni/multi-color probe for FISH. In search of a multicolor object such as rainbow and confetti to name this technology in an imagery fashion, we got the inspiration from a famous Chinese cartoon, Hulu Brothers (húlú is calabash in Chinese), where each of the seven protagonists was transformed from a calabash with a distinct rainbow color, much like the base color in HuluFISH multiplexing. Besides, calabash fruits on a winding vine resemble fluorophores on a helical HuluFISH probe. Additionally, húlú bears the image of life in Chinese culture: it is a container for elixir and a symbol for reproduction, coinciding with the intended use of HuluFISH in

205 medicine and life science.

206

207 FISH GSOs and adaptor oligonucleotides were synthesized from Sigma with
208 lowest quality for purification (desalting). For individual gene, GSOs were
209 pooled together to have 100 μM total oligonucleotide concentration.
210 Fluorescent Hulu oligonucleotides were purchased from Eurofins Genomics
211 with various dyes, including Atto dyes, Alexa dyes or Cy dyes. For HuluFISH
212 1.0, ligation was performed in T4 DNA ligase buffer (NEB, B0202S), with 30
213 μM adaptor for HuluFISH 1.0, 3 µM GSOs and Hulu oligonucleotide, 25%
214 PEG8000, 30 U/μL T4 DNA ligase (NEB, M0202M). Ligation reaction mix was
215 then incubated in a thermocycler, with 12 cycles of 37 ˚C 10 seconds / 16 ˚C 5
216 minutes. For HuluFISH 2.0, ligation reaction mix was prepared as HuluFISH
217 1.0 with some modifications, such as 16.7 μM of GSOs, adaptor for
218 HuluFISH 2.0 and Hulu 2.0 oligonucleotides, 50 U/μL T4 DNA ligase. Then
219 the ligation mix was left in the dark at room temperature for 2 hours. The
220 ligation product was concentrated with 9 volumes of butanol and centrifuged
221 as pellet at 20,000 g, 15 minutes at 4 ˚C. colorful labeled oligonucleotide
222 pellet was washed once with 100% ethanol and spin down to remove
223 ethanol, then resolubilized in loading buffer (8M Urea, 1 x TBE (Carl Roth,
224 A118.1), 0.01% bromophenol blue and xylene cyanol). With 5 minute dena-
225 turing at 90 ˚C, oligonucleotides were loaded onto 15% Urea-PAGE gel (8M

Urea, 1 x TBE, 15% Rotiphorese Gel 30 (Carl Roth, 3029.2), 0.05% ammonium persulfate, 0.05% tetramethylethylenediamine) pre-run at 300 V for 30 minutes. Running condition was usually 300 V, 30 minutes, or until the bromophe- nol blue reached the end. Gel bands with fluorescent dye-oligonucleotide conjugates were excised under the ambient light. Gel pieces were homogenized manually by microtube pestle (Sigma, Z359947-100EA), and then extracted with 500 μL 10 mM TE buffer (pH 8.5, 10 mM tris(hydroxymethyl)aminomethane (Tris), 1 mM Ethylenediaminetetraacetic acid (EDTA)) at room temperature overnight, protected from light by wrapping in aluminium foil. The extracted oligonucleotides in TE were concentrated again by butanol, and washed once by ethanol like before. The final pellets were dried in dark at room tempera- ture for 5-10 minutes, and then resolubilized in H_2O. The concentration was determined by nanodrop one (ThermoFisher) as ssDNA.

FISH probe staining and imaging

HuluFISH probe mix was adjusted to 10 nM for each single oligonucleotide in hybridization buffer (2 x SSC (saline-sodium citrate), 10 % (w/v) dextran sulfate, 10% (v/v) formamide, 1 mg/mL tRNA (Roche, 10109541001), 2 mM ribonucleoside vanadyl complex (NEB, S1402S), 0.2 mg/mL BSA). Gapdh-Quasar570 probe was purchased from Biosearch Technology, resuspended and used for the staining as instructed from the manufacturer. Hybridization

was performed in a water bath at 30 ˚C overnight, with the sample faced down on the parafilm. Cells on coverslip or tissue sections on glass slide were washed with washing buffer (2 x SSC, 10 % (v/v) formamide, 0.1 % (w/v) Tween-20) at 37 ˚C for 6 × 10 minutes. The last washing step included 0.5 μg/mL DAPI (4',6-diamidino-2-phenylindole) for nuclei staining. The sample was mounted in

ProLong Gold Antifade (ThermoFisher, P10144), and cured overnight. The sample then was either imaged on a widefield microscope (Zeiss Cell Observer) with 200 ms, 950 ms and 5000 ms for 405nm, 488 nm and 561 nm channel, or on a confocal microscope with Airyscan (Zeiss LSM800, equipped with 405, 488, 561, and 640 nm laser) with maximal laser power (c.a. 5%) in each channel. The sample was scanned with Airyscan technology with the optimal settings provided by ZEN software.

Image analysis

Except for the nuclear outline manually defined in ImageJ, all the image analysis was performed in R, and majorly based upon the package EBImage[16]. All intensity threshold values were based on the arbitrary units generated from Zeiss Airyscan and thus not specified in the following description. FISH dot identification relied on 2D local maxima identification and alignment. Initially for each frame, 2D maxima above a low threshold value were identified. Each 2D local maximum regarded its projection on the neighboring z-slices for alignment: those that fall within 0.08 um were

assigned to the same FISH dot. The pixels with maximal intensities (pseudo-3D-maxima) for identified FISH dots were extracted for further analysis.

Signal-to-noise ratio (SNR) and contrast were generated adaptively for each individual FISH dot. To this end, pixel values (local background) were taken from a square centered around the pseudo-3D-maxima, excluding all circular regions covering the PSF (point spread function) for 2D maxima on the same plane. Contrast is defined as the ratio of the maximal intensity and the mean of its local background values; SNR, as traditionally defined, equals to maximal intensity divided by the standard deviation of local background values.

For color decoding in samples with Hulu-probe for multiple genes, the presence of fluorophore on each channel was initially separately determined. Dual or triple color coding was assigned when FISH dots from different channels co-localized within 0.08 um. Single color assignment required thresholding with a higher intensity, given there were three copies of fluorophores in the single-color Hulu-probe. Nuclei were manually segmented on the maximum intensity projected image in ImageJ. Without the assistance of membrane immunostaining, each identified FISH dot was assigned to its closest nuclei.

Statistical analysis

Wilcoxon two samples test was used for evaluating the significance of our

295 probe design based on HybEff and the conventional one based on Tm. p-
296 value is indicated in corresponding figure legend. Single cell gene set
297 expression data were hierarchically clustered using Euclidean distance and
298 shown as heatmap. 5 clusters were retrieved from 58 cells by cutting the
299 dendrogram tree.

300 ## Author Contributions

301 Y.S.C and H.K.L conceived and designed the project. Y.S.C performed
302 most of the experiments. Y.Z helped with Image acquisition and analysis,
303 K.H, P.Z. G.B and H.A helped with FISH staining H.K.L supervised the
304 project. Y.S.C, Y.Z, and H.K.L analyzed the results and wrote the
305 manuscript.

306 ## Acknowledgement

307 We thank Sven Poppelreuther Carl Zeiss Imaging center in the DKFZ for
308 providing critical suggestions in using Airyscan technology. We thank Bernd
309 Fisher for generously provide R code for local maxima. This work was
310 supported by the Helmholtz Association (VH-NG-702), the Deutsche
311 Forschungsgemeinschaft (LI 2140/1-1), the Deutsche Krebshilfe (110226),
312 the Helmholtz Alliance Preclinical Comprehensive Cancer Center (PCCC),
313 the Chica Heinz Schaller foundation, the EMBO young investigator program,
314 and the ERC (European Research Council) consolidator grant (647055) to
315 H-K.L.

316

Competing Financial Interests

Y.S.C and H.K.L are inventors on two provisional patent applicatons which present the HuluFISH.

Material & Correspondence

Correspondence and requests for materials should be addressed to Y.S.C (yongsheng.cheng@dkfz.de) or H.K.L (email: l.haikun@dkfz.de).

References

1. Gall, J. G. & Pardue, M. L. Formation and Detection of Rna-Dna Hybrid Molecules in Cytological Preparations. *Proc. Natl. Acad. Sci.* 63, 378–383 (1969).
2. Femino, A. M., Fay, F. S., Fogarty, K. & Singer, R. H. Visualization of Single RNA Transcripts in Situ. *Science* 280, 585–590 (1998).
3. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Meth* 5, 877–879 (2008).
4. Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. & Cai, L. Single-cell in situ RNA profiling by sequential hybridization. *Nat. Methods* 11, 360–361 (2014).
5. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* aaa6090 (2015). doi:10.1126/science.aaa6090
6. Lubeck, E. & Cai, L. Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nat. Methods* 9, 743–748 (2012).
7. Levesque, M. J. & Raj, A. Single-chromosome transcriptional profiling reveals chromosomal gene expression regulation. *Nat. Methods* 10, 246–248 (2013).
8. Raj, A. & Tyagi, S. in *Methods in Enzymology* 472, 365–386 (Elsevier, 2010).
9. Lyubimova, A. *et al.* Single-molecule mRNA detection and counting in mammalian tissue. *Nat. Protoc.* 8, 1743–1758 (2013).

10. Tsanov, N. *et al.* smiFISH and FISH-quant – a flexible single RNA detection approach with super-resolution capability. *Nucleic Acids Res.* gkw784 (2016). doi:10.1093/nar/gkw784

11. Untergasser, A. *et al.* Primer3—new capabilities and interfaces. *Nucleic Acids Res.* 40, e115–e115 (2012).

12. Wright, E. S., Yilmaz, L. S., Corcoran, A. M., Ökten, H. E. & Noguera, D. R. Automated Design of Probes for rRNA-Targeted Fluorescence In Situ Hybridization Reveals the Advantages of Using Dual Probes for Accurate Identification. *Appl. Environ. Microbiol.* 80, 5124–5133 (2014).

13. Roy, R., Hohng, S. & Ha, T. A practical guide to single-molecule FRET. *Nat. Methods* 5, 507–516 (2008).

14. Zhuang, X. *et al.* Fluorescence quenching: A tool for single-molecule protein-folding study. *Proc. Natl. Acad. Sci.* 97, 14241–14244 (2000).

15. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415 (2003).

16. Pau, G., Fuchs, F., Sklyar, O., Boutros, M. & Huber, W. EBImage—an R package for image processing with applications to cellular phenotypes. *Bioinformatics* 26, 979–981 (2010).

Figures and Figure Legends

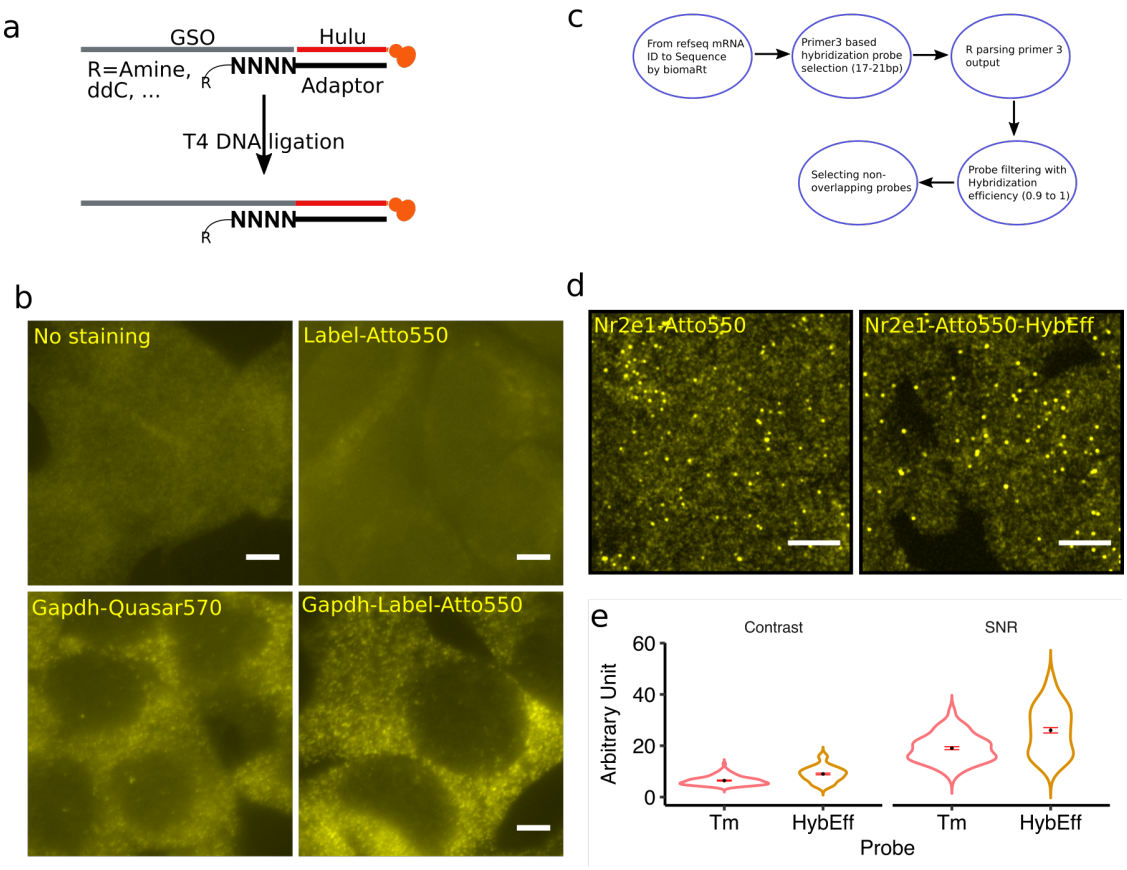

Figure 1: HuluFISH 1.0 probe's enzymatic labeling and improved probe design. (**a**) T4DL based labeling scheme for HuluFISH probe with standard 3' end -OH group of GSO. (**b**) *in situ* staining with HuluFISH probe (Gapdh-Label-Atto550) and commercial mouse Gapdh probe (Gapdh_Qusar570) in Hepa 1-6 cell. (**c**) smFISH probe selection pipeline used for all following probes in this paper. Scale bar: 10 μm. (**d**) smFISH detection of low-expressing gene, Nr2e1 in embryonic brain tissue with conventional Tm based or our new hybridization efficiency based probe design. Scale bar: 5 μm. (**e**) Contrast and SNR analysis for Nr2e1's conventional (Tm) and our

(HybEff) design. Between Tm and HybEff, welch two sample t-test's p-value for contrast and SNR are 2.6×10^{-10} and 3.0×10^{-8}.

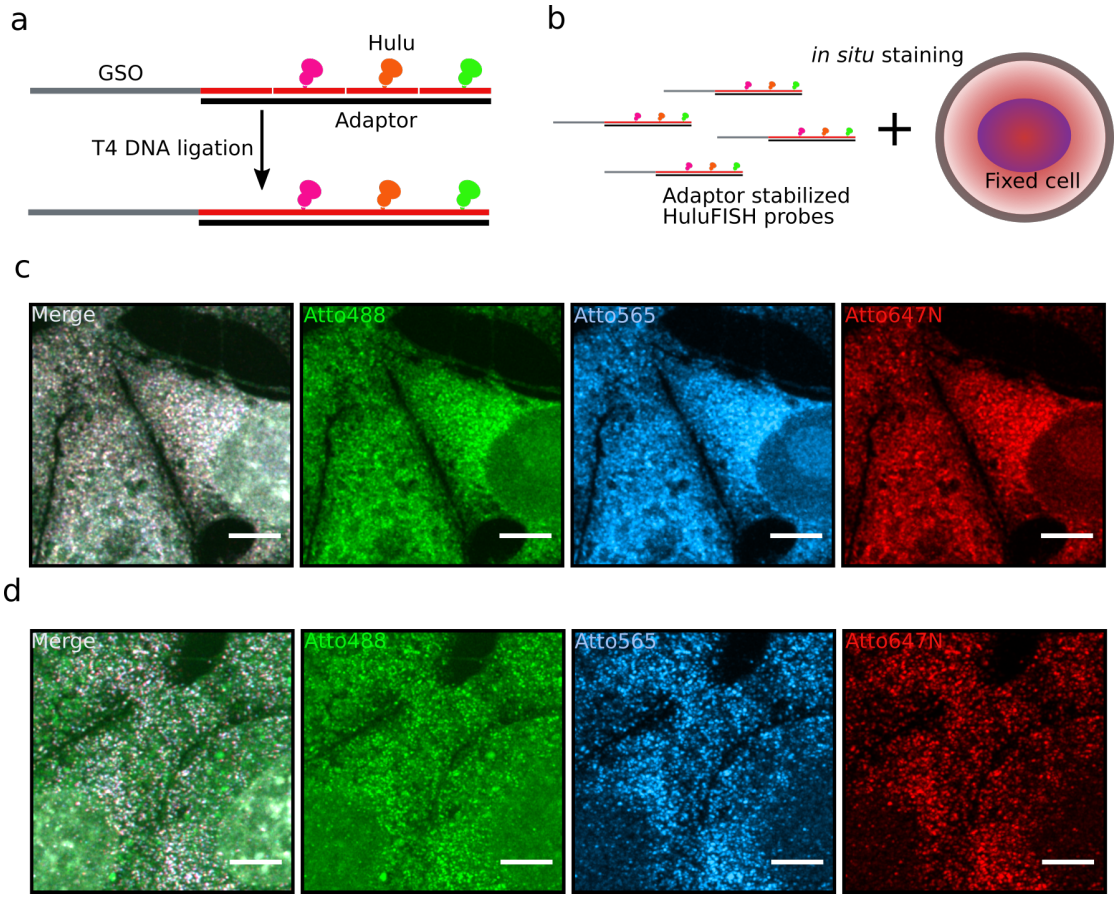

Figure 2: Multiple fluorophore labeling based on HuluFISH 2.0.

(**a**) Multiple-way ligation based fluorophore labeling of HuluFISH probes. (**b**) *in situ* staining with HuluFISH probes pre-annealed with the adaptor to avoid multiple fluorophore quenching. (**c**) Gapdh expression in Hepa 1-6 with the Gapdh HuluFISH probe conjugated with Atto488, Atto565, and Atto647N, without adaptor stabilization. (**d**) Gapdh mRNA visualized as individual dots by adaptor pre-annealed HuluFISH probe. Scale bar is 5 μm.

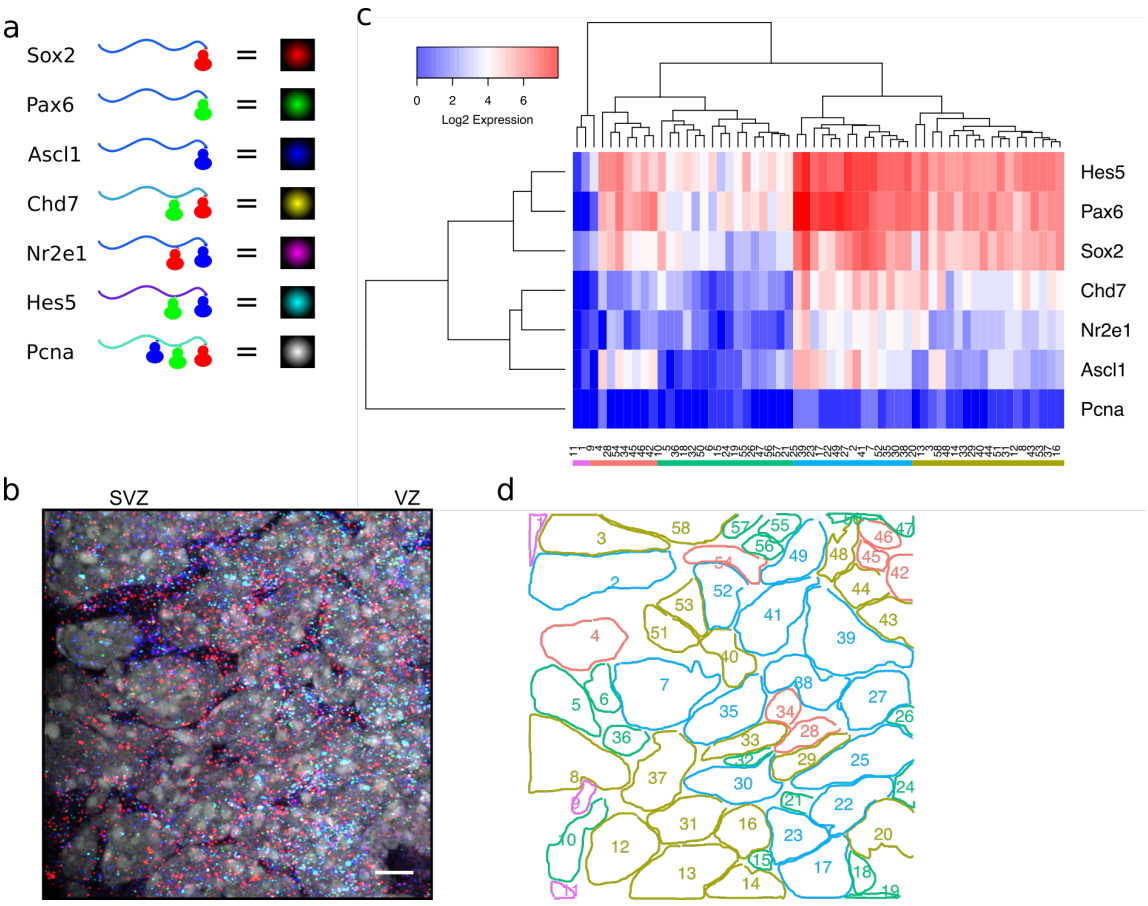

Figure 3: HuluFISH detection of 7 genes in mouse embryonic brain.

(**a**) Color coding scheme for HuluFISH from 3 base colors. (**b**) 7-gene detection in E12.5 mouse embryonic brain ventricular zone (VZ) and subventricular zone (SVZ). Scale bar 5 μm. (**c**) Hierarchical clustering of all single cells in (b) based on the Log_2 transformation of mRNA transcript counts for these 7 genes. (**d**) Spatial illustration of molecular subgroups in mouse telencephalon neural progenitors identified in (c). Cluster color scheme is the same as it in (c).