Highly Specific Multiplexed RNA imaging with Split-Fluorescence *In Situ* Hybridization

Jolene Jie Lin Goh[1,2], Nigel Chou[1,2], Wan Yi Seow[1], Norbert Ha[1], Chung Pui Paul Cheng[1], Kok Hao Chen[1,*]

[1] Genome Institute of Singapore, Agency for Science, Technology and Research (A*STAR), 60 Biopolis Street, Singapore 138672, Singapore.

[2] These authors contributed equally

[*] Correspondence should be addressed to K. H. C. (chenkh@gis.a-star.edu.sg).

ABSTRACT (3 sentences, 70 words)

We present SPLIT-FISH, a highly specific multiplexed Fluorescent *In Situ* Hybridization method centered on an optimized split-probe design. SPLIT-FISH resulted in reduced non-specific background, lowered false detection rate, and enabled accurate profiling of un-cleared mouse tissue slices. With SPLIT-FISH, we measured the localization and abundance of 317 genes from the mouse brain, liver, kidney and ovary, revealing diverse transcript localization patterns.

Main text (1500 words including Abstract, no heading)

Multiplexed Fluorescent *In Situ* Hybridization (FISH)[1–5] allows combinatorial imaging of the transcriptome[6,7] and is an attractive approach to spatial transcriptomics, which promises to reveal structure-function relationships of single cells in the native tissues. A key challenge to make multiplexed FISH more broadly applicable to all tissue types is the accuracy of detecting individual RNA molecules in the complex tissue environment, which often have limited signal and suffer from tissue-dependent background. Many efforts focus on signal amplification to make RNA spots brighter[8–11]. However, signal amplification can only improve signal relative to tissue auto-fluorescence, it does not differentiate real RNA spots from non-specifically bound probes, since all probes are equally amplified.

Off-target binding of FISH probes generates background fluorescence as well as spurious spots, a problem amplified in multiplexed FISH due to the use of highly diverse (thousands of sequences) and concentrated probe solutions. The current strategy to tackle this problem is to use customized tissue clearing approaches to remove cellular proteins and lipids[6,12], minimizing non-specific probe binding. However, this method does not remove non-specific binding to other remaining RNAs and it involves lengthy protocols that may require further optimization to be adaptable to other tissues.

Another approach to suppress non-specific background is the use of split (or tandem) probes[13,14], where RNA fluorescence signal is detected only when two independent hybridizations are co-localized. Since the split-probe strategy has not been adapted to multiplexed FISH, we sought to further optimize the probe sequence at the single-molecule level and to design custom barcoded split sequences to allow combinatorial labelling.

First, we optimize the split probe sequence using single-molecule FISH on the MUC5AC transcripts (Fig. S1). We reasoned that the split readout sequence needed to be shorter in length than in conventional multiplexed FISH to prevent any unpaired off-target probes from binding to the fluorescent readout probe (Fig. 1a and 1d). So, we titrated the

length of the split readout sequence and found that it needed to be 9 nucleotides or less to produce no detectable background signal (Fig. S2). At the same length, when the adjoining encoding probe binds and the two parts of the readout sequences are brought into close proximity with each other, strong on-target signals should be detected. So, we further screened several pairing schemes, including circular, cruciform, double ‘C’, and double ‘Z’ (Fig. S1), and found the circular construct to have the brightest on-target signal – 38% higher intensity than the previously reported double 'Z' construct[13]. Importantly, it produced comparable signal intensity as the conventional readout scheme, indicating that RNA brightness was not compromised as a result of background discrimination. As a proof of concept, we tested the optimized split-probe by targeting the long non-coding RNA, XLOC_010514, for which one of the single-molecule FISH probes is known to bind to off-targets in the nucleus[15] (Fig. S3). The split-probe approach successfully quells the signals arising from non-specific binding, suggesting that there is no need to remove or even know the ‘rogue’ sequence *a priori*.

Next, we focused on optimizing split-probe for multiplexed FISH, which we term SPLIT-FISH. SPLIT-FISH can be divided into 3 stages – hybridization of the encoding probe, bridging of the split encoding probes, and fluorescent labelling with the readout probes (Fig. 2a). Because the bridging of split probes needs to be repeated over many rounds, we screened the ‘bridge’ sequences for fast binding kinetics to allow shorter hybridization time. We also found that the primers used for oligo library amplification impeded the circularization of the adjoining probe pairs, so we incorporated restriction sites adjacent to primer sequences, allowing the primers to be cleaved off by restriction digestion. With the optimized design, we were able to perform multiple iterations of hybridization and washing (up to at least 20 rounds) without the loss of FISH signal or RNA count (Fig. S4).

We proceeded to compare the performance of the conventional versus split-probe approaches in mouse cell lines and tissue slices. To demonstrate combinatorial labelling of RNAs, we randomly selected 317-genes as targets and designed twenty-six barcoded

'bridge' sequences. As one split-probe can only carry one barcode, we designed an N Choose 2 barcoding scheme (Supplementary Table 2) by splitting the encoding probes into two tiles, each carrying one of the two required barcodes. Compared to the samples stained with the conventional probe library, the samples stained with the split-probe library showed ~2 and ~6 fold reduction in non-specific background intensity (estimated as the median value of the raw images) in mouse hepatocyte cultures (AML12) and brain tissue slices respectively (Fig. 1b, c, e, f). Images taken after washing of readout probes also confirmed that off-target binding is the main contributor to background signal and tissue auto-fluorescence (in our detection channels) was less in comparison (Fig. S5). Most of the detected RNA spots were successfully decoded and they correlated strongly with RNA-seq (Fig. 2c, d, e, f). Despite employing a simpler barcoding scheme that has no error correction, the false positive rate (estimated using number of blank code-words detected per cell) was ~1.3 fold (AML 12 cell line), ~99 fold (brain tissue), and ~200 fold (liver tissue) lower in the split-probe scheme compared to the conventional scheme (Fig. 2b). Despite not using tissue clearing, the average cell-line false positive rate was 0.15 per cell comparable to the previously reported result of 0.08 per cell in a cleared cell-line sample[12].

To demonstrate that the split-probe scheme works robustly without any tissue-specific clearing, we used the same 317-gene probe set to image the mouse frontal cortex, liver, kidney, and ovary. The transcripts counts from these tissues correlated well with RNA-seq (Fig. S6), with log Pearson correlation values between 0.52 and 0.67. Furthermore, diverse localization patterns were observed in each tissue type (Fig 3a, b, c, d). For example, MAP4 was found to be highly enriched in the neuronal processors in the frontal cortex and Myh11 was found predominantly lining the portal veins in the liver. Distinct zonation patterns in the kidney (e.g. Osbpl8, Ppl, and Notch 3) suggested a spatial division of labor previously observed in liver[16]. Some transcripts, such as Slc12a7, Plxnc1, and Dsp, were highly compartmentalized in the mouse ovary, likely corresponding to the different

maturation stages of the follicles. SON and ABCC2 were also highly localized to the nucleus in liver tissue, highlighting the power of multiplexed FISH to distinguish subcellular features.

In conclusion, we demonstrated accurate multiplexed FISH of 317 genes in four different mouse tissues without using tissue clearing. SPLIT-FISH not only simplified multiplexed FISH tissue preparation protocol, but also potentially broadened the range of accessible tissue types. While 317 genes is mostly sufficient for many biological questions, one major limitation of our current split-probe implementation is its limited multiplexing capability - a result of using only one barcode per probe-pair (compared with up to four readout sequences in the conventional design). We envision that with further shortening of the barcoding sequence or creative refinement of the barcode structure, the number of transcripts probable by SPLIT-FISH can be further expanded, opening the way to *in situ* transcriptomics for all tissue types.

1. 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).

2. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).

3. Shaffer, S. M. *et al.* Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* **546**, 431–435 (2017).

4. Codeluppi, S. *et al.* Spatial organization of the somatosensory cortex revealed by osmFISH. *Nat. Methods* **15**, 932–935 (2018).

5. Mateo, L. J. *et al.* Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* **568**, 49–54 (2019).

6. Eng, C.-H. L. *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* (2019). doi:10.1038/s41586-019-1049-y

7. Xia, C., Fan, J., Emanuel, G., Hao, J. & Zhuang, X. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *Proc. Natl. Acad. Sci.* **116**, 201912459 (2019).

8. Wu, C. *et al.* RollFISH achieves robust quantification of single-molecule RNA biomarkers in paraffin-embedded tumor tissue samples. *Commun. Biol.* **1**, 1–8 (2018).

9. Rouhanifard, S. H. *et al.* ClampFISH detects individual nucleic acid molecules using click chemistry–based amplification. *Nat. Biotechnol.* **37**, 84–94 (2019).

10. Xia, C., Babcock, H. P., Moffitt, J. R. & Zhuang, X. Multiplexed detection of RNA using MERFISH and branched DNA amplification. *Sci. Rep.* **9**, 1–13 (2019).

11. Kishi, J. Y. *et al.* SABER amplifies FISH: enhanced multiplexed imaging of RNA and DNA in cells and tissues. *Nat. Methods* **16**, 533–544 (2019).

12. Moffitt, J. R. *et al.* High-performance multiplexed fluorescence in situ hybridization in

culture and tissue with matrix imprinting and clearing. *Proc. Natl. Acad. Sci.* **113**, 201617699 (2016).

13. Wang, F. *et al.* RNAscope: A novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *J. Mol. Diagnostics* **14**, 22–29 (2012).

14. Choi, H. M. T. *et al.* Third-generation in situ hybridization chain reaction: Multiplexed, quantitative, sensitive, versatile, robust. *Dev.* **145**, 1–10 (2018).

15. Cabili, M. N. *et al.* Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol.* **16**, 20 (2015).

16. Halpern, K. B. *et al.* Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. *Nature* **542**, 1–5 (2017).

Acknowledgements

We thank Prof. Shyam Prabhakar, Prof. Zhao Ziqing Winston, Mike Huang, Jiamin Toh, Dr. Yunching Chang, Dr. Christabelle Goh, Dr. Vipul Singhal, and Dr. Kalpesh Mehta for insightful discussions and technical assistance. We thank the Advanced Molecular Pathology Laboratory, IMCB, A*STAR, Singapore for mouse necropsy and cryo-sectioning services. We thank the GIS next generation sequencing platform, A*STAR, Singapore for sequencing services. We thank Applied Materials for in-kind contributions. This research was supported by the Singapore Ministry of Health's National Medical Research Council under its grant OFYIRG15nov017 and also by the Agency for Science, Technology, and Research under its grant I1801E0029. K. H. C. is grateful for support from AXA Post-Doctoral fellowship. N. C. is grateful for support from A*STAR Graduate Academy.

Author contributions

K.H.C. conceived the study. K.H.C., J.J.L.G., and N. C. designed the experiments. K.H.C. and N.C. built the instruments. N.C. and K.H.C. developed the computational software and performed the image analysis. J.J.L.G. and K.H.C. performed the experiments. W.Y.S., N. H., and C.P.P.C. assisted with the experiments and replicated the results. K.H.C., J.J.L.G., and N. C. analysed the data and interpreted the results. W.Y.S. assisted with data analysis. K.H.C., N. C., J.J.L.G., W.Y.S., and N. H. wrote the manuscript.

Competing financial interests

Authors have submitted a provisional patent application that is based on the technology described in this manuscript.

FIGURES (limit to 3)

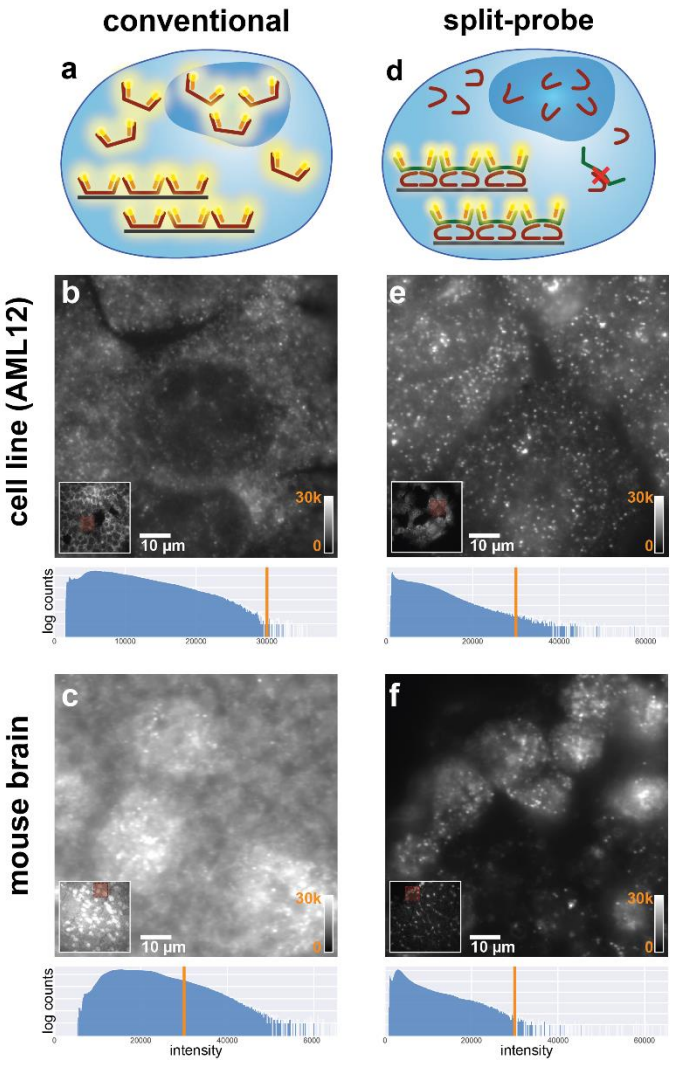

Figure 1: Comparison of split-probe to conventional multiplexed Fluorescent *In Situ* Hybridization (a, d) Schematic comparison between schemes. Cellular RNA in black, encoding probes in red, dye-labeled readout probes in orange. Bridge probes (split-scheme only) are in green, which bind only when 2 matching encoding probes are coincident at the same location. (b, c, e, f) Unprocessed images from AML12 cell line and mouse brain scaled to the same camera intensity value (30k). Inset shows a full field of view, main image shows cropped regions, position indicated by red box in inset. Histograms are calculated from the entire field of view. (b,c) Representative images prepared using conventional scheme. (e,f) Images prepared using split-probe scheme.

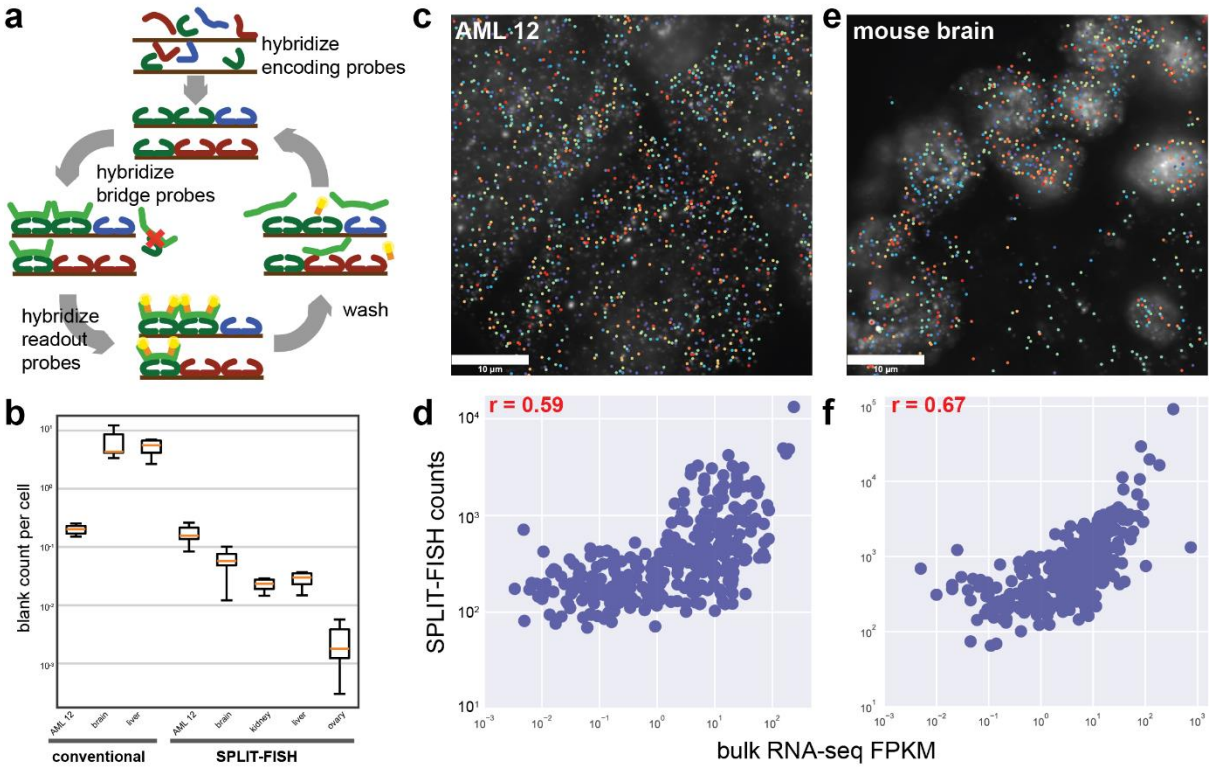

Figure 2: Split-probe based multiplexed Fluorescent *In Situ* Hybridization (a) Schematic of multiplexed SPLIT-FISH imaging protocol. Encoding probes are hybridized first. At each round of imaging, bridge probes are first introduced and allowed to hybridize, followed by fluorescently labelled readout probes. After imaging, both bridge and readout probes are washed out in preparation for the next round. (b) Comparison of blank counts per cell for AML 12, mouse brain and liver. Kidney and ovary blank counts per cell for SPLIT-FISH are also shown. 7 blank sequences were tested for each sample in both SPLIT-FISH (317 genes) and conventional (133 genes) schemes. (c, d, e, f) Decoded gene locations from multiplexed SPLIT-FISH. Regions displayed are identical to Figure 1. Here, maximum intensity projections across all bits are shown with decoded gene locations overlaid. Colours represent different genes. Scatterplots of total counts per gene (whole tissue) vs bulk RNA-sequencing FPKM values, with Pearson correlation in red, are shown below. (c, d) AML12 cell-line. (e, f) Mouse frontal cortex tissue slice.

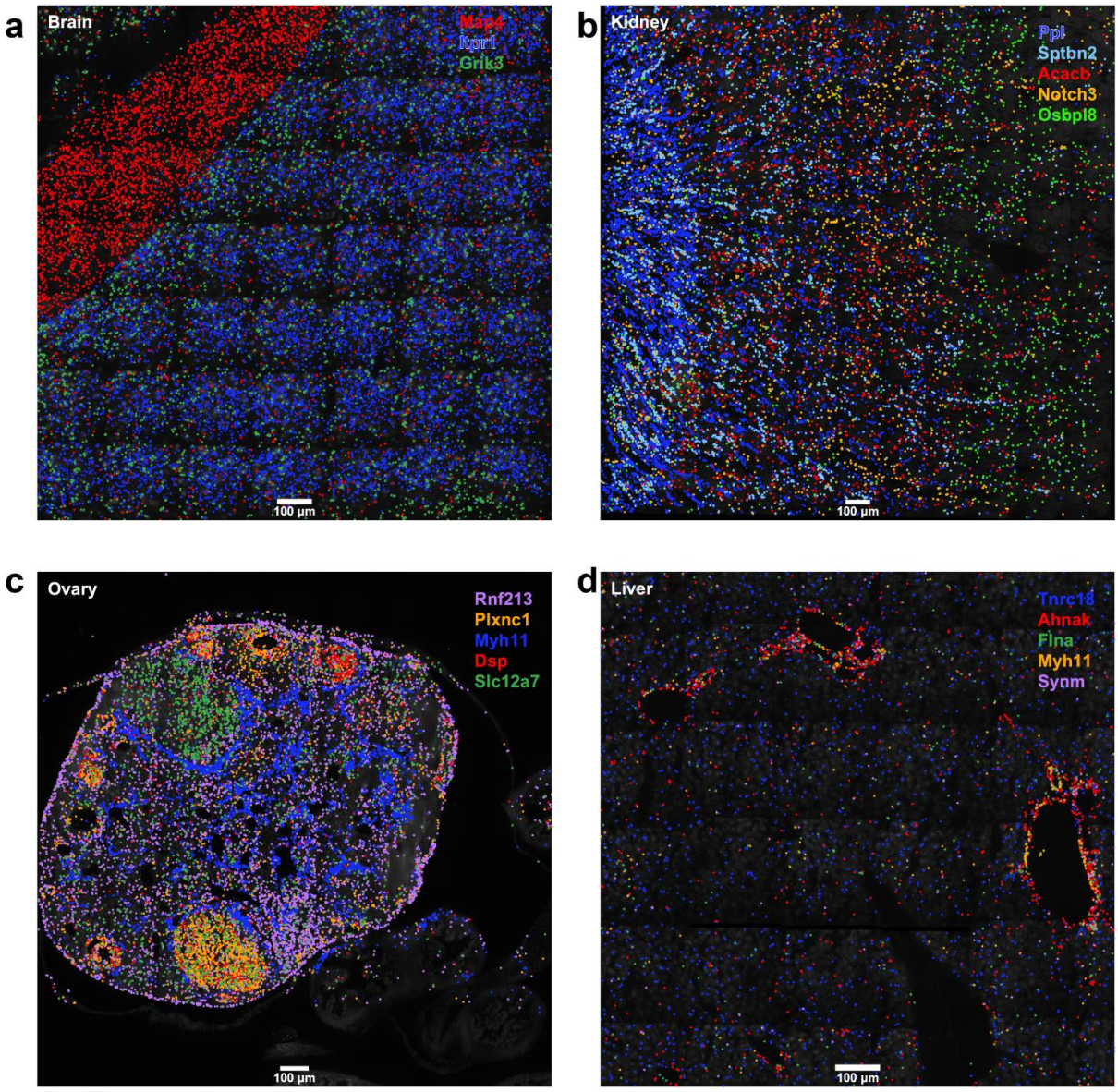

Figure 3: Illustrative transcript localization patterns in 4 mouse tissues (a) Brain tissue showing differential localization of transcripts in neuronal processes (Map4) and regions containing cell bodies (e.g. Itpr1). (b) Inner to outer zonation patterns of 5 genes (Ppl, Sptbn2, Acacb, Notch3, and Osbpl8) in a kidney section. (c) Compartmentalized localization of genes within large ovarian follicles or corpus luteum, localization of genes surrounding follicles (Myh11) and at the outer surface of the ovary (Rnf213) (d) Localization of genes around portal veins of the liver section.