

Utilizing Split Fluorescent Proteins to Visualize Binary Cell Fate Decisions

Setu Mehta1

1*Harvard College*

## May 2021

**Abstract**

inary cell fate decisions serve at a cornerstone of cellular decision-making processes during embryonic development. Understanding and studying these decisions require an intimate knowledge of the spatial and temporal expression dynamics of critical genes. Split fluorescent proteins (sFP) can serve as a novel tool to study these binary cell fate decisions, with unique applications such as the potential to amplify weak genetic signals. Ultimately, sFPs can be utilized to revolutionize the study of protein-protein interactions during embryonic development and beyond.

BL

# Binary Cell Fate Decisions

During embryonic development, pluripotent stem cells are exposed to a variety of cellular signals that influence their cell fate. Arnold and Robertson demonstrated that cells undergo binary lineage decisions regulated by the activity of critical signaling pathways (Arnold and Robertson, 2009). The first binary cell fate decision pluripotent stem cells undergo is their specification into either the trophectoderm (TE) or the inner cell mass (ICM). This fate decision is determined by the selective expression of caudal-type homeobox protein 2 (CDX2) in the TE or octamere-binding transcription factor 3/4 (OCT3/4) in the ICM. Cells in the ICM are then exposed to further genetic signals that specify either primitive endoderm or epiblast by expression of GATA-binding factor 6 (GATA6) and Nanog (Arnold & Robertson, 2009). As embryonic development progresses, these cells continue to make binary cell fate decisions.

Various studies have relied on single-cell sequencing to identify critical genes associated with cell patterning and axis specification during embryonic development (Arnold & Robertson, 2009; Jang et al., 2017; Thomson et al., 2011; Yiangou et al., 2019). While these studies have elucidated potential genetic targets important for early embryonic development, these data function on the assumption that we can make conclusions about gene expression dynamics based on the genetic expression of fixed cells and apply them to live cell cultures undergoing similar cell fate transitions *in vitro* (Jang et al., 2017). It is challenging to track cell fate decisions and investigate how the *in vitro* environment and the spatial and temporal expression of relevant genes influence cell fate decisions in real-time (Thomson et al., 2011). Therefore, there is a need to develop an effective method to track live cells to understand how genes identified through single-cell sequencing are spatially and temporally expressed in a network of cells.

Specifically, it can be interesting to study the temporal and spatial role of transcription factors in binary cell fate decisions specifying important embryonic events such as mesoderm cell fate specification. Mesodermal-derived cell fate pathways are unclear due to the early specification of



mesodermal progenitors during gestation, as well as the increased duration of time required to generate heterogeneous mixtures of desired mesodermal cell types (Loh et al., 2016).

# Studying Binary Cell Fate Decisions

Traditionally, scientists have used fluorescent proteins (FPs) such as green fluorescent protein (GFP) isolated from the *Aequorea victoria* jellyfish to visualize the live expression of proteins of interest (Rodriguez et al., 2017). The fluorescent protein is inserted into the target gene sequence through homology-directed repair (HDR), a process where while repairing a double-stranded DNA break, the cell incorporates new DNA at that locus as the new sequence is flanked by homology arms that resemble the region of insertion (Liang et al., 1998). Therefore, when the DNA at that insertion loci is translated, the FP is also translated, making the protein product fluorescent and easy to track in real-time.

To increase the efficacy and targeting ability of HDR, zinc finger nucleases (ZFN), transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced short palindromic repeats (CRISPR-Cas9) gene-editing technologies were developed. These gene-editing technologies allow scientists to more accurately choose where they want to insert a new DNA sequence through HDR and allows for more targeted tagging of genes of interest with FPs. ZFNs are composed of zinc finger proteins fused to the FokI endonuclease and recognize a trinucleotide DNA sequence (Kim et al., 1996). TALENs are composed of the bacterial TALE protein fused to the FokI endonuclease, and each TALE molecule identifies one nucleotide, with an array of TALE’s recognizing more than one nucleotide (Christian et al., 2010). CRISPR-Cas9 is composed of a Cas9 nuclease, a trans-activating crRNA (tracrRNA), and a single guide RNA (sgRNA) that localizes the Cas9 nuclease to the correct region. This genetic region has to be followed by a protospacer adjacent motif (PAM) for accurate localization of Cas9 (Mali et al., 2013).

# Self-complementing Split Fluorescent Proteins

Incorporating FPs through traditional HDR requires the FP to be introduced through large plasmids containing the homology arms to the DNA locus of insertion flanking the FP DNA. Self-complementing split fluorescent proteins (sFPs) are engineered to allow more robust visualization of protein expression and localization (Feng et al., 2017). Each FP is composed of 11 beta-sheets and is split into two separate domains, a large and small domain, that fluoresce upon self-association. The larger domain is composed of the first 1-10 beta-sheets and associates with the small domain, which has the 11th beta-sheet. **Figure 1** demonstrates the self-complementation process of sFPs, describing that the domains cannot fluoresce until they are bound together.





To fluorescently tag a gene with sFPs, scientists can incorporate the large domain of the sFP into the genome. Then, instead of using HDR to integrate the entire FP, scientists can use HDR to incorporate only the small domain of the sFP into the region of interest, targeted to that region via ZFNs, TALENs, or CRISPR-Cas9. Therefore, when that DNA region is translated, both domains will self-complement and cause the translated protein to fluoresce, allowing scientists to live track the movement and expression of that protein. Using sFPs to fluorescently tag genes can be more effective than using FP proteins as HDR with sFPs only require small single-stranded oligo DNA nucleotides (ssODNs) to be incorporated into the genome, whereas HDR with FPs involves the incorporation of a large plasmid containing the entire FP sequence. Additionally, the generation of sFP small domain-containing ssODNs is much more convenient, as it is an oligo that can be ordered, whereas the FP-containing plasmid needs to be engineered. Furthermore, adding higher concentrations of the sFP small domain will amplify weak fluorescent signals, allowing scientists to visualize low throughput proteins (Feng et al., 2017). Currently, two sFPs, mNG2 and sfCherry3C, are the brightest available proteins of the sFP family. The mNG2 construct was created by Feng et al. (Feng et al., 2017). The sfCherry3C was created by Feng et al. (Feng et al., 2018).

Despite these benefits, there are some limitations to the use of sFPs. A major concern with using sFPs is that the large and small domains may dissociate, negatively impacting levels of fluorescence. However, **Figure 2** demonstrates that while the initial binding of sFPs is reversible (Step 1), upon maturation and subsequent fluorescence of the compound (Step 2), the binding is irreversible (Feng et al., 2018).



Some other most prominent limitations include a slightly smaller quantum efficiency than their FPs counterparts, a delay between the expression of the sFP-tagged proteins of interest and the fluorescent readout due to time required for sFP complementation, proper protein folding, and chromophore maturation, and the inherent affinity for sFP fragments with one another which can lead to the creation of false-positive signals (Köker et al., 2018; Romei & Boxer, 2019). Romei and Boxer outline multiple areas of innovation designed to address these limitations, including



introducing mutations in the sFP genome that will allow for protein folding and complementation at relevant physiological temperatures. It will be essential to continue following the large body of research dedicated to addressing these limitations to ensure that any sFP utilized in a research study is of the highest quality and efficiency.

# Conclusions

Ultimately, sFPs can be utilized as a novel tool and viable alternative in studying binary cell fate decisions. Incorporating the large domains of both the mNG2 and sfCherry3C sFPs, among other sFPs, at specific loci in human embryonic stem cells (hESC), can allow for the simultaneous live tracking of two proteins of interest. The small domains of the sFPs can then incorporated into the genetic loci of genes of interest to create stem cell lines to study and more effectively track gene expression. However, it will be important to continue modifying sFP structure and technology to address significant limitations to their use.



# References

Arnold, S. J., & Robertson, E. J. (2009). Making a commitment: Cell lineage allocation and axis patterning in the early mouse embryo. *Nature Reviews Molecular Cell Biology*, *10*(2), 91–103. https://doi.org/10.1038/nrm2618

Christian, M., Cermak, T., Doyle, E. L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A. J., & Voytas, D. F. (2010). Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. *Genetics*, *186*(2), 757–761. https://doi.org/10.1534/genetics.110.120717

Feng, S., Sekine, S., Pessino, V., Li, H., Leonetti, M. D., & Huang, B. (2017). Improved split fluorescent proteins for endogenous protein labeling. *Nature Communications*, *8*(1), 370. https://doi.org/10.1038/s41467-017-00494-8

Feng, S., Varshney, A., Villa, D. C., Modavi, C., Kohler, J., Farah, F., Ali, N., Mueller, J. D., VanHoven, M., & Huang, B. (2018). *Bright split red fluorescent proteins with enhanced complementation efficiency for the tagging of endogenous proteins and visualization of synapses* [Preprint]. Cell Biology. https://doi.org/10.1101/454041

Jang, S., Choubey, S., Furchtgott, L., Zou, L.-N., Doyle, A., Menon, V., Loew, E. B., Krostag, A.-R., Martinez, R. A., Madisen, L., Levi, B. P., & Ramanathan, S. (2017). Dynamics of embryonic stem cell differentiation inferred from single-cell transcriptomics show a series of transitions through discrete cell states. *ELife*, *6*, e20487. https://doi.org/10.7554/eLife.20487

Kim, Y. G., Cha, J., & Chandrasegaran, S. (1996). Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(3), 1156–1160.

Köker, T., Fernandez, A., & Pinaud, F. (2018). Characterization of Split Fluorescent Protein Variants and Quantitative Analyses of Their Self-Assembly Process. *Scientific Reports*, *8*(1), 5344. https://doi.org/10.1038/s41598-018-23625-7

Liang, F., Han, M., Romanienko, P. J., & Jasin, M. (1998). Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *Proceedings of the National Academy of Sciences*, *95*(9), 5172–5177. https://doi.org/10.1073/pnas.95.9.5172

Loh, K. M., Chen, A., Koh, P. W., Deng, T. Z., Sinha, R., Tsai, J. M., Barkal, A. A., Shen, K. Y., Jain, R., Morganti, R. M., Shyh-Chang, N., Fernhoff, N. B., George, B. M., Wernig, G., Salomon, R. E. A., Chen, Z., Vogel, H., Epstein, J. A., Kundaje, A., … Weissman, I. L. (2016). Mapping the Pairwise Choices Leading from Pluripotency to Human Bone, Heart, and Other Mesoderm Cell Types. *Cell*, *166*(2), 451–467. https://doi.org/10.1016/j.cell.2016.06.011

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Güell Cargol, M., DiCarlo, J. E., Norville, J. E., & Church, G. M. (2013). *RNA-guided human genome engineering via Cas9*. http://dx.doi.org/10.1126/science.1232033



Rodriguez, E. A., Campbell, R. E., Lin, J. Y., Lin, M. Z., Miyawaki, A., Palmer, A. E., Shu, X., Zhang, J., & Tsien, R. Y. (2017). The growing and glowing toolbox of fluorescent and photoactive proteins. *Trends in Biochemical Sciences*, *42*(2), 111–129. https://doi.org/10.1016/j.tibs.2016.09.010

Romei, M. G., & Boxer, S. G. (2019). Split Green Fluorescent Proteins: Scope, Limitations, and Outlook. *Annual Review of Biophysics*, *48*, 19–44. https://doi.org/10.1146/annurev-biophys-051013-022846

Thomson, M., Liu, S. J., Zou, L.-N., Smith, Z., Meissner, A., & Ramanathan, S. (2011). Pluripotency Factors in Embryonic Stem Cells Regulate Differentiation into Germ Layers. *Cell*, *145*(6), 875–889. https://doi.org/10.1016/j.cell.2011.05.017

Yiangou, L., Grandy, R. A., Osnato, A., Ortmann, D., Sinha, S., & Vallier, L. (2019). Cell cycle regulators control mesoderm specification in human pluripotent stem cells. *Journal of Biological Chemistry*, *294*(47), 17903–17914. https://doi.org/10.1074/jbc.RA119.008251