Perspective

Combinatorial genetics methods for discovering high-order regulatory combinations and engineering genetic drivers for neural differentiation

Dawn G. L. Thean, Alan S. L. Wong[*]

Diving into the search for effective cell differentiation factors: Researchers are still striving to find better therapeutics to revert or slow down the progression of neurological disorders such as Parkinson's disease, Alzheimer's disease, and Huntington's disease. These disorders are the result of neuronal cell death in different parts of the brain. Medications or treatments used to relieve the symptoms in patients have not been well established. They do not recover the damaged neural tissues and can result in unwanted side effects. Therefore, an increasing number of studies look to stem cells as a promising therapeutic, because of their self-renewal capabilities and flexibility of differentiation into desired cell lineages for engraftment into the patient to recover the lost neural tissues. However, before stem cells can be clinically used in treating neurological disorders, there are still areas that require a better understanding to unlock their full potential.

Stem cells are regenerative and malleable, with the propensity to become any cell-type given the right concoction of factors to drive differentiation. However, there is a depth of complexity in the large network of factors. It is important to determine the combinations of transcription factors (TF), small molecules, and/or growth factors to achieve the optimal synergism for differentiating stem cells including embryonic stem cells (ESCs) and induced pluripotent stem cells at the fastest rate, resulting in the purest population of a neuronal lineage. However, there could be a strategic way to screen for these factors in a more cost-efficient and time-effective manner. This would help optimize the already established protocols that are still laborious or for further characterization of neuronal subtypes such as medium spiny neurons, sensory neurons, and serotonergic neurons. It remains a challenge to profile the drivers of neuronal fate comprehensively and define how these elements interact with one another in the regulatory network.

A few groups have used new high-throughput screening approaches to elucidate the regulatory network of cell fate and provided informative insights on factors that drive neuronal conversion and survival. Liu et al. (2018) used CRISPR activation (CRISPRa) to search for TFs or DNA-binding factors that promote the neuronal fate of ESCs. However, the work was limited to study single factors or pairwise combinations. Another systematic screen was conducted by Tsunemoto et al. (2018), which investigated the relationship between TFs. Most of their selected single TF showed no effects in the study, however, they identified 76 pairwise TF combinations to promote differentiation of mouse fibroblasts to induced neurons. Notably, their results lacked some consistency with previous studies, such as not detecting the dopamine active transporter Slc6a3 in induced neurons, possibly due to missing out on

additional factors. Another study determined the essential genes to maintain or improve the survival of neurons by conducting a CRISPR interference (CRISPRi)-based screening (Tian et al., 2019). And again, they were limited to the screening of single sgRNAs. Further studies are required for an unbiased screening for other genetic factors, and to increase the number of factor combinations, as it has been shown that a cocktail of more than two factors may be required in different contexts to efficiently drive differentiation and reprogramming. A good example is the combination of Oct4, Sox2, Klf4, and c-Myc that is necessary for regulating the developmental signaling network for the pluripotency of ESCs (Takahashi and Yamanaka, 2006). In such complex systems, there is a need to comprehensively characterize the functions of high-order genetic combinations in a high-throughput fashion, and Combinatorial Genetics *En Masse* (CombiGEM) method may just be able to do that.

CombiGEM method for a high-throughput, high-order, and systematic screen of factor combinations: CombiGEM offers not only a systematic way of conducting large-scale pooled screens but also has the capability of scalable assembly of high-order combinatorial genetic libraries (Wong et al., 2015, 2016; Zhou et al., 2020). The one-pot process allows the user to screen for a multitude of genetic combinations; 1-way, 2-way, 3-way, and in theory, n-way libraries. This provides a rapid alternative to the conventional process of building and testing individual candidate combinations of interest. While several other combinatorial CRISPR screening strategies were also developed for studying pairwise genetic combinations (Han et al., 2017; Shen et al., 2017; Najm et al., 2018; Truong et al., 2019; DeWeirdt et al., 2020), CombiGEM offers a unique opportunity to evaluate interactions between three or more genetic combinations. For instance, if the user aims to use CRISPRa or CRISPRi for overexpressing or repressing, respectively, a list of candidate TF combinations, one could screen a barcoded library of TF-targeting sgRNA combinations using CombiGEM-CRISPR v2.0 (Wong et al., 2016; Zhou et al., 2020) as illustrated in **Figure 1**. By using one-pot ligation steps, the library of sgRNAs and their respective barcodes are incorporated into a lentiviral-destination vector that reports the expression of a fluorescent protein upon activation of a neuron cell type-specific promoter, such as tubulin α1. The library of sgRNAs and a separate lentiviral vector harboring the enzymatically deficient Cas9 either fused with a transcriptional activator or repressor can be delivered into the desired starting cells. Over time, as cells begin to differentiate, only induced
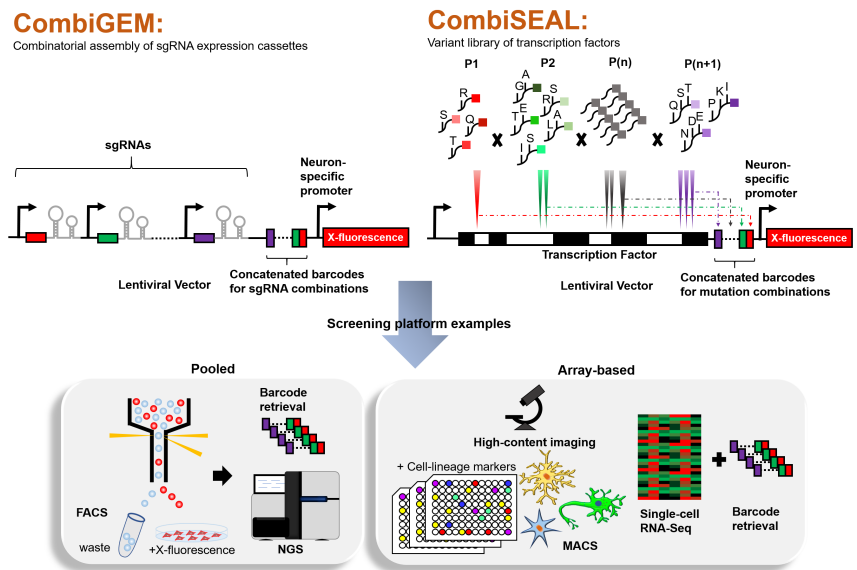
Figure 1 | **CombiGEM and CombiSEAL methods to screen for effective genetic drivers of differentiation.**
An example of the CombiGEM method is the assembly of multiple sgRNAs expressed from the same construct, and the sgRNA combinations are represented by their concatenated barcodes. The CombiSEAL method comprises mutations introduced to various parts of a protein, such as transcription factors, indicated as P1, P2, P(n), and P(n+1). And the combinations of mutations are reflected by their corresponding concatenated barcodes. The user can apply the setups to other DNA delivery systems, and in this representation lentiviral vectors are used. Screening strategies can be conducted by a pooled format, and in this example, identifying cell populations that differentiate to a specific neuron lineage is detected by activation of the promoter for fluorescent protein expression, labeled as X-fluorescence in the diagram. Next-generation sequencing (NGS) of the barcodes from the cell pool will decipher the DNA targets or transcriptional variant(s). Stringent analysis of factors that drive differentiation of different neuron types could use array-based methods such as the probing of cells with cell lineage markers followed by magnetic-activated cell sorting (MACS) into multi-well plates, which can be followed up with high-content imaging. Alternatively, single-cell RNA sequencing (RNA-Seq) allows the user to profile the gene expressions of cells with the desired phenotypes and would be able to retrieve the barcodes to identify sgRNA combinations or mutation combinations of the key variant(s). FACS: Fluorescence-activated cell sorting.

Perspective

neuron-like cells will express fluorescence, and these cells can be isolated using fluorescence-activated cell sorting for retrieving their harboring TF combinations via barcode sequencing. To gain a better understanding of the combinations that drive specific cell lineages, these cells can also be probed for known cell-lineage reporters or markers based on the preference of differentiation stages the user chooses to study. Magnetic-activated cell sorting could be used to separate cells of interest depending on the cell surface marker expressions. Single-cell RNA sequencing could be coupled to the combinatorial CRISPR screening as a readout to profile the type and levels of gene up or downregulation and the combination of targeted TFs can be determined via the barcode reads (Replogle et al., 2020). High-content imaging could be used to monitor the cell morphology changes from early to later stages of neuron development. The physiological relevance or functionality of the induced neuron-like cells can be determined by immunocytochemical labeling specific for neuron maturation, such as NeuN, TUJ1, and MAP2, and electrophysiological measures.

As mentioned earlier, studies have combined TF and small molecules to manipulate cell fate. CombiGEM can also be used to identify small molecule combinations that can maintain self-renewal of stem cells, induce cell-lineage differentiation, or facilitate reprogramming by increasing the efficiency and potentially replacing the genetic factors. CombiGEM can also be applied by first determining the small molecule targets of signaling pathways, epigenetic, or cellular process factors. Then via designing a library of sgRNA to activate or inactivate the druggable targets, one can then identify the effective combinations of small molecules to enhance the differentiation or reprogramming. This concept is well reflected by the studies we have earlier conducted to discover drug target combinations against cancer and Parkinson's disease (Zhou et al., 2020).

The use of CombiGEM is not only limited to CRISPR-based perturbations but could also be applied to other DNA or RNA regulatory factors to study the loss-of-function screens with RNA interference, as well as gain-of-function ones with the expression of microRNAs (Wong et al., 2015) and sequence-verified human open reading frames such as those reported in the human TFome library (Ng et al., 2020).

CombiSEAL method for high-throughput engineering of transcription factors: Finding the ideal combination of naturally occurring genetic factors can still pose challenges, such as being limited to characterized TFs of the genome with prior knowledge about their expression and roles in differentiation. Studies have demonstrated that engineering TFs or generating artificial transcription factors opens up a new avenue of options to speed up the rate of differentiation through alternate pathways, surpassing the dependence of prerequisite expression of other endogenous co-factors within the gene regulatory network, or altogether replacing natural TFs with more efficient ones (Jauch, 2018). This enables researchers to tailor the required factors according to their experimental needs.

We propose that if one aims to conduct high-throughput mutagenesis on multiple sites to a TF that comprises of different domains, CombiSEAL could prove to be a useful platform (Choi et al., 2019). The CombiSEAL method assembles amino acid-coding DNA fragments that are combinatorially tagged with barcodes (**Figure 1**). CombiSEAL begins with first sectioning parts of the protein where mutagenesis is desired. The

user can generate any number of variants for each section, then tagging them with barcodes to identify the position and combination of mutations. Each pool of sections is then assembled sequentially by inserting them into a destination vector harboring the wild-type protein sequence modified with flanking type IIS enzymes at the region of intended mutagenesis. This scarless-fusion scheme of linking multiple parts of a protein is important to avoid adding unwanted amino acids to the protein. The CombiSEAL method allows determining the amino acid sequences of the selected pool of variants of interest more efficiently and cost-effectively, avoiding the need for long-read sequencing over the entire protein to identify the types of mutations and where they are situated across the protein. This methodical setup allows easier analysis by conducting a high-throughput sequencing of the pool of short concatenated barcodes that infers the combination of types and positions of amino acid mutations, or other desired modifications such as domain swapping, insertion, and deletion, that had been initially designed and installed on the TF.

Conclusion: In comparison to trial-and-error methods, advances have been made in employing more systematic approaches to determine the essential factors or combinations required to drive the conversion of one cell type to a neuronal-lineage. However, the ability to identify higher-order combinations has been lacking, and here we propose that the CombiGEM method may be able to address such limitations. CombiGEM uses a barcoded, one-pot ligation system to piece together higher-order combinations of DNA binding factors to target DNA in one-go, circumventing the process of multiple rounds of screening to narrow down the number of hits to a manageable size for downstream validations. Furthermore, less efficient natural TFs can be substituted with artificial or engineered TFs to better regulate gene expression. We describe a protein mutagenesis method, CombiSEAL, that will enable users to create large pools of transcription factor variants by straightforwardly assembling multiple site mutations tagged with barcodes within the protein and their different domains. This will speed up and lower the cost of screening procedures to retrieve the information on the types of mutations of the variants of interest. We hope that the proposed methods can assist with further understanding and broadening the possibilities of promoting neural regeneration and can be widely applicable to other areas of research.

Dr. Alan S. L. Wong would like to declare that patent applications have been filed based on our published work on the presented combinatorial genetics platforms.

Dawn G. L. Thean, Alan S. L. Wong*

Laboratory of Combinatorial Genetics and Synthetic Biology, School of Biomedical Sciences, The University of Hong Kong, Pokfulam, Hong Kong Special Administrative Region, China (Thean DGL, Wong ASL)
Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam, Hong Kong Special Administrative Region, China (Wong ASL)
*Correspondence to: Alan S. L. Wong, PhD, aslw@hku.hk.
https://orcid.org/0000-0003-1790-3233 (Alan S. L. Wong)

References

Choi GCG, Zhou P, Yuen CTL, Chan BKC, Xu F, Bao S, Chu HY, Thean D, Tan K, Wong KH, Zheng Z, Wong ASL (2019) Combinatorial mutagenesis en masse optimizes the genome editing activities of SpCas9. Nat Methods 16:722-730.

DeWeirdt PC, Sanson KR, Sangree AK, Hegde M, Hanna RE, Feeley MN, Griffith AL, Teng T, Borys SM, Strand C, Joung JK, Kleinstiver BP, Pan X, Huang A, Doench JG (2020) Optimization of AsCas12a for combinatorial genetic screens in human cells. Nat Biotechnol 39:94-104.

Han K, Jeng EE, Hess GT, Morgens DW, Li A, Bassik MC (2017) Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. Nat Biotechnol 35:463-474.

Jauch R (2018) Cell fate reprogramming through engineering of native transcription factors. Curr Opin Genet Dev 52:109-116.

Najm FJ, Strand C, Donovan KF, Hegde M, Sanson KR, Vaimberg EW, Sullender ME, Hartenian E, Kalani Z, Fusi N, Listgarten J, Younger ST, Bernstein BE, Root DE, Doench JG (2018) Orthologous CRISPR-Cas9 enzymes for combinatorial genetic screens. Nat Biotechnol 36:179-189.

Ng AHM, Khoshakhlagh P, Rojo Arias JE, Pasquini G, Wang K, Swiersy A, Shipman SL, Appleton E, Kiaee K, Kohman RE, Vernet A, Dysart M, Leeper K, Saylor W, Huang JY, Graveline A, Taipale J, Hill DE, Vidal M, Melero-Martin JM, et al. (2020) A comprehensive library of human transcription factors for cell fate engineering. Nat Biotechnol doi: 10.1038/s41587-020-0742-6.

Replogle JM, Norman TM, Xu A, Hussmann JA, Chen J, Cogan JZ, Meer EJ, Terry JM, Riordan DP, Srinivas N, Fiddes IT, Arthur JG, Alvarado LJ, Pfeiffer KA, Mikkelsen TS, Weissman JS, Adamson B (2020) Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. Nat Biotechnol 38:954-961.

Shen JP, Zhao D, Sasik R, Luebeck J, Birmingham A, Bojorquez-Gomez A, Licon K, Klepper K, Pekin D, Beckett AN, Sanchez KS, Thomas A, Kuo CC, Du D, Roguev A, Lewis NE, Chang AN, Kreisberg JF, Krogan N, Qi L, et al. (2017) Combinatorial CRISPR-Cas9 screens for de novo mapping of genetic interactions. Nat Methods 14:573-576.

Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126:663-676.

Tian R, Gachechiladze MA, Ludwig CH, Laurie MT, Hong JY, Nathaniel D, Prabhu AV, Fernandopulle MS, Patel R, Abshari M, Ward ME, Kampmann M (2019) CRISPR interference-based platform for multimodal genetic screens in human iPSC-derived neurons. Neuron 104:239-255.

Truong VA, Hsu MN, Kieu Nguyen NT, Lin MW, Shen CC, Lin CY, Hu YC (2019) CRISPRai for simultaneous gene activation and inhibition to promote stem cell chondrogenesis and calvarial bone regeneration. Nucleic Acids Res 47:e74.

Wong AS, Choi GC, Cheng AA, Purcell O, Lu TK (2015) Massively parallel high-order combinatorial genetics in human cells. Nat Biotechnol 33:952-961.

Wong AS, Choi GC, Cui CH, Pregernig G, Milani P, Adam M, Perli SD, Kazer SW, Gaillard A, Hermann M, Shalek AK, Fraenkel E, Lu TK (2016) Multiplexed barcoded CRISPR-Cas9 screening enabled by CombiGEM. Proc Natl Acad Sci U S A 113:2544-2549.

Zhou P, Chan BKC, Wan YK, Yuen CTL, Choi GCG, Li X, Tong CSW, Zhong SSW, Sun J, Bao Y, Mak SYL, Chow MZY, Khaw JV, Leung SY, Zheng Z, Cheung LWT, Tan K, Wong KH, Chan HYE, Wong ASL (2020) A three-way combinatorial CRISPR screen for analyzing interactions among druggable targets. Cell Rep 32:108020.