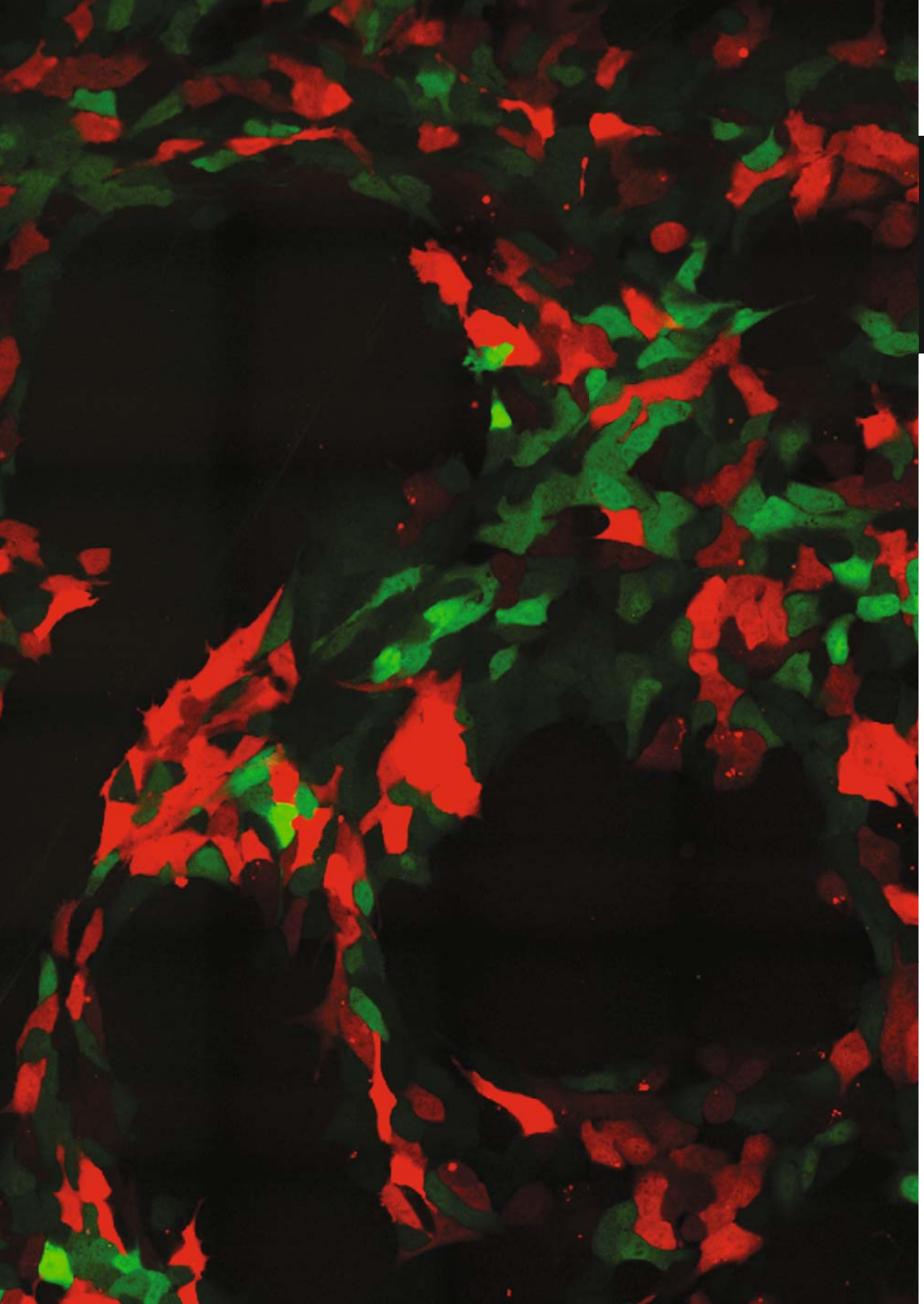
Advanced synthetic biology tools for genetic modification of human stem cells and their applications

Blanch-Asensio, A.



CHAPTER 1

General introduction

Introduction

Human pluripotent stem cells (hPSCs) possess the unique ability to self-renew and differentiate into all cell types derived from the three embryonic germ layers: endoderm, mesoderm, and ectoderm. This remarkable potential enables hPSCs to form any cell type in the human body (**Figure 1**), making them invaluable for studying developmental biology and advancing therapeutic applications for tissue repair and regeneration (Engle & Puppala, 2013; Zhu & Huangfu, 2013). There are two primary types of hPSCs: human embryonic stem cells (hESCs) and human induced pluripotent stem cells (hiPSCs). The fundamental distinction between these types lies in their origin: hESCs are derived from the inner cell mass of human blastocyst-stage embryos, whereas hiPSCs are reprogrammed somatic cells, typically sourced from adult tissues. The reprogramming process generally involves the overexpression of four specific transcription factors associated with pluripotency: OCT3/4, SOX2, c-MYC, and KLF4 (Takahashi et al., 2007). In 2010, synthetic modified mRNA (modRNA) enabled the reprogramming of human somatic cells through a non-integrating approach (Warren et al., 2010). By 2013, mouse somatic cells were successfully reprogrammed using chemically defined conditions (Hou et al., 2013). More recently, this chemically defined reprogramming method has been extended to human somatic cells (Guan et al., 2022; Liuyang et al., 2023).

hiPSCs and hiPSC-derived cell types

The discovery of hiPSCs by Shinya Yamanaka marked a transformative milestone in stem cell biology. Initially, Yamanaka and colleagues successfully reprogrammed adult mouse fibroblasts into PSCs (Takahashi & Yamanaka, 2006). Shortly thereafter, human dermal fibroblasts were reprogrammed using retroviral vectors to deliver the transcription factors OCT3/4, SOX2, c-MYC, and KLF4. The hiPSCs demonstrated characteristics nearly identical to hESCs, including similar morphology, proliferation rates, and gene expression profiles. Additionally, hiPSCs could differentiate into cell types from the three germ layers and form teratomas in mice (Takahashi et al., 2007).

This breakthrough not only expanded the possibilities for research but also provided an ethically favorable alternative to hESCs. Since hiPSCs can be generated from any somatic cell type, the need for human embryos as a source of pluripotent stem cells was eliminated. This addressed critical ethical issues

surrounding the destruction of embryos for research. Moreover, hiPSCs have revolutionized disease modeling by enabling the derivation of patient-specific cell lines. These cells carry the same genetic mutations as the patients from whom they were derived, providing a powerful platform for studying disease mechanisms (Soldner et al., 2011). Additionally, advances in gene editing technologies, such as CRISPR/Cas9, have further enhanced the utility of hiPSCs. For instance, isogenic control lines can be generated by correcting mutations in patient-derived hiPSCs or introducing specific disease-causing mutations into wild-type hiPSCs (Bassett, 2017). These isogenic pairs minimize genetic variability and improve the accuracy of genotype-phenotype correlations, offering a robust framework for studying disease mechanisms (Brandão et al., 2020; Soldner et al., 2011).

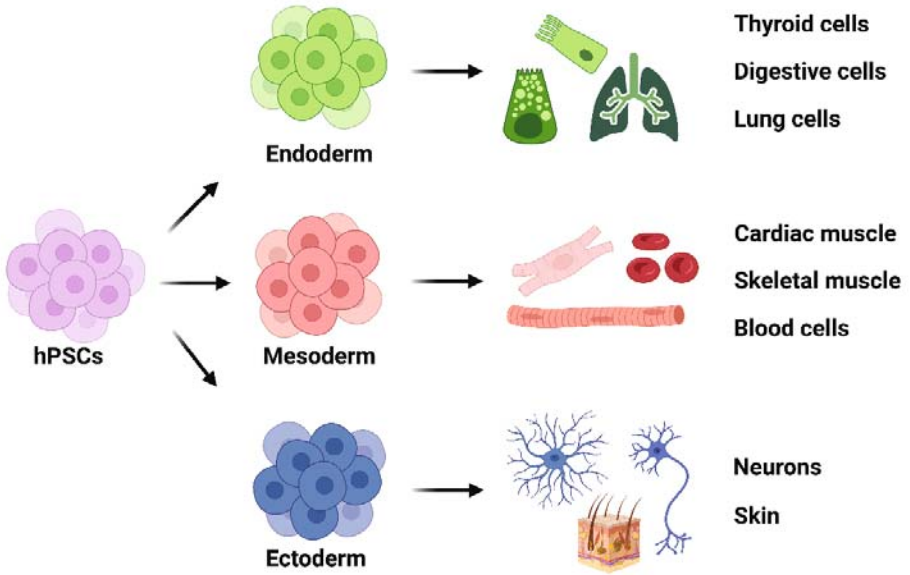


Figure 1. hPSCs differentiation potential.

This figure highlights that hPSCs have the potential to develop into any cell type in the human body. They can differentiate into cells from all three germ layers: endoderm, mesoderm, and ectoderm, which together contain all types of cells in the body.

The early protocols developed for the *in vitro* differentiation of hESCs were readily adapted for hiPSCs, enabling the efficient generation of various somatic cell types. Over the past two decades, differentiation techniques have been refined using developmental biology insights. By providing precise signaling cues,

hiPSCs can now be directed to differentiate into highly specialized cells, including pacemaker cells and motor neurons (Fernandopulle et al., 2018; Schweizer et al., 2017). Importantly, when hiPSCs harbor disease-associated mutations that are expressed in specific cell types, the differentiated cells also exhibit these mutations. This enables the modeling of disease phenotypes observed in patients. For instance, hiPSC-derived cardiomyocytes have been instrumental in studying cardiac diseases (Bellin et al., 2013). These cells have been shown to recapitulate electrophysiological abnormalities observed in patients, such as the prolonged repolarization phase caused by *KCNH2* mutations associated with long QT syndrome type 2 (Bellin et al., 2013; Brandão et al., 2020).

Similarly, hiPSCs provide unprecedented opportunities for neurodegenerative disease modeling. Disease variants linked to conditions such as Alzheimer's disease or amyotrophic lateral sclerosis (ALS) can be introduced into wild-type hiPSCs or corrected in patient-derived hiPSCs. These cells can then be differentiated into neurons, enabling the study of disease pathophysiology and drug discovery *in vitro* (Kondo et al., 2013; Workman et al., 2023).

Despite their transformative potential, generating patient-derived hiPSCs and creating isogenic controls remains a time-intensive and technically demanding process. Introducing or correcting genetic mutations in the human genome can require several months, as it involves laborious steps such as single-cell deposition and extensive screening to identify correctly edited clones (Brandão et al., 2022). This highlights the urgent need for high-throughput methodologies with improved efficiency and shorter timelines to accelerate the generation of genetically-modified hiPSC lines.

Genome engineering tools

Introducing genetic modifications into mammalian organisms and cell lines began in the 1970s. However, over the past decade, the techniques used to facilitate these modifications have advanced significantly. These improvements have made the process safer, reducing the likelihood of disrupting the endogenous genome, while also increasing efficiency and precision. While random integration methods have provided fundamental insights into gene modification, their limitations necessitate more precise approaches. This has led to the development of targeted genome editing techniques that ensure greater accuracy and safety.

Targeted approach is a more precise alternative. In this strategy, mutations or DNA cargo are inserted into specific regions of the human genome. Key tools enabling this precision are engineered endonucleases, which are essential components of the genome engineering toolbox. These include Zinc Finger Nucleases (ZFNs) and Transcription Activator-Like Effector Nucleases (TALENs). A groundbreaking advancement occurred in 2012 with the discovery of a bacterial adaptive immune system, CRISPR-Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated protein 9), and adaptation for mammalian genome editing in 2013 (Cong et al., 2013; Jinek et al., 2012; Mali et al., 2013). To tailor CRISPR-Cas9 for use in higher organisms, modifications were made. Spacer and repeat sequences were combined into a single-guide RNA (sgRNA) to streamline targeting, while the Cas9 protein was codon-optimized for enhanced expression in eukaryotic systems and equipped with a nuclear localization signal (Cong et al., 2013; Jinek et al., 2012). The sgRNA sequence typically contains around 20 nucleotides complementary to a chosen target sequence in the genome. This sequence is next to a protospacer-adjacent motif (PAM) sequence. For the commonly used *Streptococcus pyogenes* **Cas9**, the PAM sequence is **5'-NGG-3'**, where "N" can be any nucleotide. The Cas9 protein first binds to the DNA by recognizing the PAM sequence via its **PAM-interacting domain**. After PAM recognition, Cas9 induces local unwinding of the DNA upstream of the PAM sequence and the sgRNA then base-pairs with the complementary sequence in the target DNA strand. If the sgRNA-DNA hybrid forms correctly, the two nuclease domains of the Cas9 are activated which results in a precise double-strand break (DSB) at the target site.

ZFNs, TALENs, and CRISPR-Cas9 serve as molecular scissors, enabling the precise cutting of specific DNA sequences. These DSBs are repaired in mammalian cells through two primary pathways: non-homologous end joining (NHEJ), which is commonly used for gene knockouts; and homology-directed repair (HDR), used for introducing specific mutations or performing gene knock-ins (Shrivastav et al., 2008). HDR requires an additional double-stranded DNA donor template to guide recombination. Among these tools, CRISPR-Cas9 has revolutionized the field of gene editing with the breakthrough earning Jennifer Doudna and Emmanuelle Charpentier the 2020 Nobel Prize in Chemistry. When combined with hiPSC technology, CRISPR-Cas9 provides an unparalleled platform for disease modeling (Bassett, 2017), gene therapy, drug discovery, and precision medicine (Ortiz-Vitali & Darabi, 2019).

Site-specific recombinases (SSRs) are another essential set of tools in genome engineering, with decades of application not only in mammalian cell lines but also in whole organisms, such as mice (Hosur et al., 2022; Zhang et al., 2023). Among SSRs, Cre recombinase has been widely favored by scientists due to its exceptional efficiency and specificity. It recognizes specific recombination sites, such as *loxP*, and facilitates precise DNA modifications, including integration, excision, and inversion of sequences flanked by or containing these sites (Turan et al., 2013). Bxb1 recombinase has become a preferred tool for mediating the integration of large DNA sequences into mammalian genomes, offering high accuracy and efficiency for such applications (Brown et al., 2011; Xu et al., 2013).

Mammalian synthetic biology

Synthetic biology is a multidisciplinary scientific field that integrates principles from biology, engineering, computer science, and chemistry to design, construct, and optimize biological systems. It involves the engineering of genomes in living organisms, including mammalian cells, to endow them with novel properties or capabilities. This field encompasses the design and construction of biological modules, systems, and machines, as well as the re-engineering of existing biological systems to serve innovative and practical purposes. At its core, synthetic biology often relies on the introduction of exogenous genetic material or synthetic genetic circuits into cells, enabling precise programming of their behavior and functions (Peterman et al., 2024).

Within the last decade, mammalian synthetic biology has emerged as a transformative approach, offering new strategies for gene- and cell-based therapies. These therapies have demonstrated significant potential in diverse fields, including precision medicine and the treatment of complex diseases. For instance, synthetic biology has enabled advancements in cancer therapies, such as the engineering of chimeric antigen receptor (CAR) T cells that target specific tumor antigens (Yan et al., 2023). Similarly, it has contributed to the development of treatments for metabolic disorders through the creation of engineered cells that regulate homeostasis or compensate for genetic deficiencies (Ye & Fussenegger, 2014). Additionally, synthetic biology has facilitated innovations in immune therapies, providing tools to modulate immune responses for autoimmune diseases or to enhance immune recognition of pathogens and cancer cells (Yan et al., 2023).

As the field progresses, mammalian synthetic biology continues to drive advancements in therapeutic applications, offering the potential to address previously intractable medical challenges. By enabling precise control and reprogramming of cellular functions, synthetic biology represents a critical step forward in the development of next-generation treatments tailored to individual patients and specific disease contexts.

Aim and scope of this thesis

In the rapidly evolving fields of genome engineering and mammalian synthetic biology, there remains a pressing need for fast, precise, and efficient methods to integrate large DNA constructs into the genome of mammalian cells. This thesis aims to address this challenge by developing a novel tool for DNA integration, followed by iterative improvements to enhance its performance. We also demonstrate the applicability of this novel and advanced platform to the fields of drug screening, disease modeling and forward programing.

Chapter 2 provides an overview of recent advancements in DNA integration platforms, with a focus on methodologies developed in the past five years. The chapter identifies key challenges and areas requiring further refinement and concludes with a discussion on potential applications that could benefit from these platforms.

Chapter 3 explores the use of genetically encoded calcium and voltage indicators delivered as mRNA in hiPSC-derived cardiomyocytes. This approach enables repeatable recordings of action potentials and calcium transients, offering valuable insights into the electrophysiological properties of these differentiated cells.

Chapter 4 introduces STRAIGHT-IN, a rapid and efficient platform for integrating large DNA payloads into the genome of hiPSCs. By combining the strengths of CRISPR/Cas9 and SSRs, this platform enables the generation of multi-reporter hiPSC lines expressing genetically encoded calcium and voltage indicators. These lines facilitate simultaneous recording of voltage, intracellular calcium, and contraction in hiPSC-derived cardiomyocytes. Additionally, STRAIGHT-IN is used to create a panel of hiPSC lines carrying disease-associated *KCNH2* variants that exhibit the expected LQT2 phenotype observed in patients with the same mutations.

Chapter 5 describes the development of STRAIGHT-IN version 2 (v2), which integrates a novel landing pad at either the *AAVS1* or *CLYBL* safe harbor loci. The modifications to the landing pad allows for more rapid and efficient generation of genetically modified hiPSC lines, further streamlining the process.

Chapter 6 provides a comprehensive, step-by-step protocol for the STRAIGHT-IN platform, covering from the culture of the hiPSCs to how to characterize the resulting genetically-modified hiPSC lines.

Chapter 7 presents STRAIGHT-IN Dual, a platform for the targeted integration of two DNA payloads into hiPSCs. By inserting a second landing pad into the free allele of the *CLYBL* locus, this platform achieves simultaneous integration of two DNA constructs with high efficiency within a single week.

Chapter 8 details the generation of hiPSC lines expressing doxycycline-inducible CRISPR/Cas variants. These lines have applications in gene editing and lineage tracing, offering powerful tools for stem cell research.

Chapter 9 concludes the thesis by discussing the work and findings presented, and exploring the future prospects of genome engineering and mammalian synthetic biology in the field of stem cell biology.

This body of work represents an important contribution to genome engineering, offering a novel tool that addresses critical challenges in the integration of large DNA constructs. By enabling rapid and efficient generation of genetically modified hiPSC lines, the STRAIGHT-IN platform paves the way for new applications in disease modeling, drug screening and forward programming.

References

Bassett, A. R. (2017). Editing the genome of hiPSC with CRISPR/Cas9: disease models. *Mammalian Genome*, *28*(7–8), 348–364. https://doi.org/10.1007/s00335-017-9684-9

Bellin, M., Casini, S., Davis, R. P., D'Aniello, C., Haas, J., Ward-Van Oostwaard, D., Tertoolen, L. G. J., Jung, C. B., Elliott, D. A., Welling, A., Laugwitz, K. L., Moretti, A., & Mummery, C. L. (2013). Isogenic human pluripotent stem cell pairs reveal the role of a KCNH2 mutation in long-QT syndrome. *EMBO Journal*, *32*(24), 3161–3175. https://doi.org/10.1038/emboj.2013.240

Brandão, K. O., Brink, L. van den, Miller, D. C., Grandela, C., van Meer, B. J., Mol, M. P. H., de Korte, T., Tertoolen, L. G. J., Mummery, C. L., Sala, L., Verkerk, A. O., & Davis, R. P. (2020). Isogenic Sets of hiPSC-CMs Harboring Distinct KCNH2 Mutations Differ Functionally and in Susceptibility to Drug-Induced Arrhythmias. *Stem Cell Reports*, *15*(5), 1127–1139. https://doi.org/10.1016/j.stemcr.2020.10.005

Brandão, K. O., Grandela, C., Yiangou, L., Mummery, C. L., & Davis, R. P. (2022). CRISPR/Cas9-Mediated Introduction of Specific Heterozygous Mutations in Human Induced Pluripotent Stem Cells. In *Methods in Molecular Biology* (Vol. 2454, pp. 531–557). Humana Press Inc. https://doi.org/10.1007/7651_2021_368

Brown, W. R. A., Lee, N. C. O., Xu, Z., & Smith, M. C. M. (2011). Serine recombinases as tools for genome engineering. *Methods*, *53*(4), 372–379. https://doi.org/10.1016/j.ymeth.2010.12.031

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science*, *339*(6121), 819–823. https://doi.org/10.1126/science.1231143

Engle, S. J., & Puppala, D. (2013). Integrating human pluripotent stem cells into drug development. *Cell Stem Cell*, *12*(6), 669–677. https://doi.org/10.1016/j.stem.2013.05.011

Fernandopulle, M. S., Prestil, R., Grunseich, C., Wang, C., Gan, L., & Ward, M. E. (2018). Transcription Factor–Mediated Differentiation of Human iPSCs into Neurons. *Current Protocols in Cell Biology*, *79*(1). https://doi.org/10.1002/cpcb.51

Guan, J., Wang, G., Wang, J., Zhang, Z., Fu, Y., Cheng, L., Meng, G., Lyu, Y., Zhu, J., Li, Y., Wang, Y., Liuyang, S., Liu, B., Yang, Z., He, H., Zhong, X., Chen, Q., Zhang, X., Sun, S., … Deng, H. (2022). Chemical reprogramming of human somatic cells to pluripotent stem cells. *Nature*, *605*(7909), 325–331. https://doi.org/10.1038/s41586-022-04593-5

Hosur, V., Low, B. E., & Wiles, M. V. (2022). Programmable RNA-Guided Large DNA Transgenesis by CRISPR/Cas9 and Site-Specific Integrase Bxb1. *Frontiers in Bioengineering and Biotechnology*, *10*, 910151. https://doi.org/10.3389/fbioe.2022.910151

Hou, P., Li, Y., Zhang, X., Liu, C., Guan, J., Li, H., Zhao, T., Ye, J., Yang, W., Liu, K., Ge, J., Xu, J., Zhang, Q., Zhao, Y., & Deng, H. (2013). Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. *Science*, *341*(6146), 651–654. https://doi.org/10.1126/science.1239278

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* , *337*, 816–821. https://www.science.org

Kondo, T., Asai, M., Tsukita, K., Kutoku, Y., Ohsawa, Y., Sunada, Y., Imamura, K., Egawa, N., Yahata, N., Okita, K., Takahashi, K., Asaka, I., Aoi, T., Watanabe, A., Watanabe, K., Kadoya, C., Nakano, R., Watanabe, D., Maruyama, K., … Inoue, H. (2013). Modeling Alzheimer's disease with iPSCs reveals stress phenotypes associated with intracellular Aβ and differential drug responsiveness. *Cell Stem Cell*, *12*(4), 487–496. https://doi.org/10.1016/j.stem.2013.01.009

Liuyang, S., Wang, G., Wang, Y., He, H., Lyu, Y., Cheng, L., Yang, Z., Guan, J., Fu, Y., Zhu, J., Zhong, X., Sun, S., Li, C., Wang, J., & Deng, H. (2023). Highly efficient and rapid generation of human pluripotent stem cells by chemical reprogramming. *Cell Stem Cell*, *30*(4), 450-459. e9. https://doi.org/10.1016/j.stem.2023.02.008

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., & Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science* , *339*(6121), 823–826. https://doi.org/10.1126/science.1232033

Ortiz-Vitali, J. L., & Darabi, R. (2019). iPSCs as a platform for disease modeling, drug screening, and personalized therapy in muscular dystrophies. *Cells*, *8*(1). https://doi.org/10.3390/cells8010020

Peterman, E. L., Ploessl, D. S., & Galloway, K. E. (2024). Accelerating Diverse Cell-Based Therapies Through Scalable Design. *Annual Review of Chemical and Biomolecular Engineering*, *15*, 267–292. https://doi.org/10.1146/annurev-chembioeng

Schweizer, P. A., Darche, F. F., Ullrich, N. D., Geschwill, P., Greber, B., Rivinius, R., Seyler, C., Müller-Decker, K., Draguhn, A., Utikal, J., Koenen, M., Katus, H. A., & Thomas, D. (2017). Subtype-specific differentiation of cardiac pacemaker cell clusters from human induced pluripotent stem cells. *Stem Cell Research and Therapy*, *8*(1). https://doi.org/10.1186/s13287-017-0681-4

Shrivastav, M., De Haro, L. P., & Nickoloff, J. A. (2008). Regulation of DNA double-strand break repair pathway choice. *Cell Research*, *18*(1), 134–147. https://doi.org/10.1038/cr.2007.111

Soldner, F., Laganière, J., Cheng, A. W., Hockemeyer, D., Gao, Q., Alagappan, R., Khurana, V., Golbe, L. I., Myers, R. H., Lindquist, S., Zhang, L., Guschin, D., Fong, L. K., Vu, B. J., Meng, X., Urnov, F. D., Rebar, E. J., Gregory, P. D., Zhang, H. S., & Jaenisch, R. (2011). Generation of isogenic pluripotent stem cells differing exclusively at two early onset parkinson point mutations. *Cell*, *146*(2), 318–331. https://doi.org/10.1016/j.cell.2011.06.019

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., & Yamanaka, S. (2007). Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell*, *131*(5), 861–872. https://doi.org/10.1016/j.cell.2007.11.019

Takahashi, K., & Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*, *126*(4), 663–676. https://doi.org/10.1016/j.cell.2006.07.024

Turan, S., Zehe, C., Kuehle, J., Qiao, J., & Bode, J. (2013). Recombinase-mediated cassette exchange (RMCE) - A rapidly-expanding toolbox for targeted genomic modifications. *Gene*, *515*(1), 1–27. https://doi.org/10.1016/j.gene.2012.11.016

Warren, L., Manos, P. D., Ahfeldt, T., Loh, Y. H., Li, H., Lau, F., Ebina, W., Mandal, P. K., Smith, Z. D., Meissner, A., Daley, G. Q., Brack, A. S., Collins, J. J., Cowan, C., Schlaeger, T. M., & Rossi, D. J. (2010). Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell*, *7*(5), 618–630. https://doi.org/10.1016/j.stem.2010.08.012

Workman, M. J., Lim, R. G., Wu, J., Frank, A., Ornelas, L., Panther, L., Galvez, E., Perez, D., Meepe, I., Lei, S., Valencia, V., Gomez, E., Liu, C., Moran, R., Pinedo, L., Tsitkov, S., Ho, R., Kaye, J. A., Thompson, T., … Svendsen, C. N. (2023). Large-scale differentiation of iPSC-derived motor neurons from ALS and control subjects. *Neuron*, *111*(8), 1191-1204.e5. https://doi.org/10.1016/j.neuron.2023.01.010

Xu, Z., Thomas, L., Davies, B., Chalmers, R., Smith, M., & Brown, W. (2013). Accuracy and efficiency define Bxb1 integrase as the best of fifteen candidate serine recombinases for the integration of DNA into the human genome. *BMC Biotechnology*, *13*, 1–17. https://doi.org/10.1186/1472-6750-13-87

Yan, X., Liu, X., Zhao, C., & Chen, G. Q. (2023). Applications of synthetic biology in medical and pharmaceutical fields. In *Signal Transduction and Targeted Therapy* (Vol. 8, Issue 1). Springer Nature. https://doi.org/10.1038/s41392-023-01440-5

Ye, H., & Fussenegger, M. (2014). Synthetic therapeutic gene circuits in mammalian cells. *FEBS Letters*, *588*(15), 2537–2544. https://doi.org/10.1016/j.febslet.2014.05.003

Zhang, W., Golynker, I., Brosh, R., Fajardo, A., Zhu, Y., Wudzinska, A. M., Ordoñez, R., Ribeiro-dos-Santos, A. M., Carrau, L., Damani-Yokota, P., Yeung, S. T., Khairallah, C., Vela Gartner, A., Chalhoub, N., Huang, E., Ashe, H. J., Khanna, K. M., Maurano, M. T., Kim, S. Y., … Boeke, J. D. (2023). Mouse genome rewriting and tailoring of three important disease loci. *Nature*, *623*(7986), 423–431. https://doi.org/10.1038/s41586-023-06675-4

Zhu, Z., & Huangfu, D. (2013). Human pluripotent stem cells: An emerging model in developmental biology. *Development (Cambridge)*, *140*(4), 705–717. https://doi.org/10.1242/dev.086165