Determinants of genome editing outcomes: the impact of target and donor DNA structures

Chen, X.

General introduction

The past few years have witnessed the explosive enthusiasm on genome editing. Scientists are better equipped with molecular tools to answer hard-to-answer questions, to cure incurable diseases, and to redefine pre-defined genetic code. Typically, genome editing is based on generating targeted chromosomal double-stranded DNA breaks (DSBs) in living cells. In response to such targeted DSBs, cells have evolved DNA repair mechanisms that can be exploited for genome editing purposes. The non-homologous end joining (NHEJ) pathway, which ligates broken chromosomal ends, can lead to gene knockouts after the incorporation of disruptive small insertions and deletions (indels); the homology-directed repair (HDR) pathway, in turn, copies the genetic information into the target sites from exogenous donor DNA templates that share homology with the target genomic sequence. Designer nucleases (a.k.a. programmable nucleases) are the molecular tools that generate the site-specific chromosomal DSBs necessary for genome editing. These tools include zinc-finger nucleases (ZFNs) and transcription activator-like effector (TALE) nucleases (TALENs), whose DNA recognition is mediated by protein-based zinc finger motifs and TALE repeats, respectively. Recently, designer nucleases based on prokaryotic CRISPR (clustered regularly interspaced short palindromic repeats)-Cas9 systems have been introduced. In contrast to their ZFN and TALEN predecessors, the CRISPR-Cas9-based platforms, consisting of RNA-guided nucleases (RGNs), do not require protein engineering efforts as their target DNA recognition is mediated by RNA-DNA hybridization, as opposed to protein-DNA interactions.

The adaptation of CRISPR-Cas9 nucleases to eukaryotic cells not only propels the development of genome editing, but also facilitates the engineering of a broad range of molecular tools for targeted transcriptional regulation, epigenetic remodeling and cellular imaging. Owing to their ease-of-use and cost-effectiveness, RNA-guided molecular tools based on CRISPR-Cas9 systems have, in a short time span, become the most widely used genome, transcriptome and epigenome manipulating platforms. Moreover, encouragingly, a number of clinical trials which make use of RGNs have recently been granted the green light by regulatory authorities. Examples include clinical trials aiming at the genetic engineering of chimeric antigen receptor T-cells and CD34+ cells from cancer patients and HIV-infected individuals, respectively. Nevertheless, one of the biggest concerns on the clinical translation of CRISPR-Cas9 nucleases is off-target activity and related potential adverse effects, e.g., deleterious or unpredictable mutations possibly compounded with oncogenic chromosomal translocations. Indeed, a number of studies have revealed that RGNs can cleave not only the chromosomal DNA at the intended target site, but also at off-target sequences bearing a varying number of mismatches. Therefore, whenever genome editing is meant to be applied in clinical settings, its specificity and accuracy

will be paramount and, as a result, potential deleterious effects caused by off-target activity should be thoroughly assessed. Although a great amount of efforts have been made to improve the safety profiles of RGNs, the fact that conventional genome editing strategies rely on DSB formation confers them a non-negligible risk. Of note, this risk is associated with unwanted mutations at not only, off-target, but also on-target sites.

In this thesis, **Chapter 1** reviews the fast-paced investigations aiming at improving the efficiency and specificity of genome editing procedures, covering the strategies based on protein engineering, gRNA design and donor DNA optimization. **Chapter 2** focuses on investigating a DSB-free genome editing principle that triggers seamless genetic modifications in human cells after the recruitment of the HDR pathway. This hereafter called in trans paired nicking strategy is based on inducing single-stranded DNA breaks (SSBs), or nicks, at genomic target sites and donor DNA templates. In this work, I present experimental evidence showing that, when compared to conventional DSB-dependent approaches, in trans paired nicking can improve the efficiency, specificity and accuracy of chromosomal DNA insertion at different human loci, including the dystrophin-encoding *DMD* gene, and the purported "safe harbor" loci *AAVS1* and *CCR5*. The extent of the site-specific genetic modifications varied from small restriction enzyme polymorphisms to whole transgene expression units. Taken together, these data *bode well* for the applicability of in trans paired nicking to settings in which non-disruptive and high-fidelity genetic manipulation of complex genomes is key.

Hitherto, it remains ill-defined how, and to what extent, the nuclear DNA architecture and the epigenetically regulated chromatin compaction affect genome editing endeavors based on designer nucleases. Hence, next to investigating DSB-free high-fidelity genome editing strategies, in this thesis, I have also probed the impact of alternate higher-order chromatin conformations on (i) the performance of TALENs and RGNs (**Chapter 3**); (ii) the performance of high-specificity designer nuclease constructions (**Chapter 4**); and (iii) the balance between NHEJ and HDR events during the gene editing interventions (**Chapter 5**). In this regard, I carried out genome editing experiments in complementary gain-of-function and loss-of-function cellular models in which the chromatin states of isogenic reporter loci are tightly regulated by small-molecule drug availability (i.e. doxycycline). The reporter genes in these cellular models are flanked by *TetO* sequences and are subjected to epigenetic regulation by the chimeric tTR-KRAB protein which, in the absence of doxycycline, binds to its cognate *TetO* recognition sites, inducing a local heterochromatic environment. This experimental setup has enabled the unbiased evaluation of the relative impact of "closed" heterochromatin and "open" euchromatin on both NHEJ- and HDR-mediated gene editing at isogenic target sequences. Gene editing experiments performed in these cellular models showed that NHEJ-mediated tar-

geted gene knockouts induced by TALENs and RGNs are significantly hampered at heterochromatic target sites (**Chapter 3**). **Chapter 4** further extends the investigations in **Chapter 3** by describing the differential impact of alternate higher-order chromatin conformations on different high-specificity RGN constructions, in particular those based on (i) offset RGN "nickase" pairs, (ii) improved Cas9 variants, (iii) truncated gRNAs and (iv) the ortholog *S. aureus* Cas9 protein. The experimental section of this thesis ends with **Chapter 5**, which investigates the impact of chromatin on genome editing outcomes resulting from the action of the HDR and NHEJ pathways with donor DNA templates of synthetic, non-viral and viral origins.

Another important aspect of genome editing is to ensure that the delivery of designer nucleases and donor DNA constructs is specific, robust and non-disruptive to the target cell's genome. This key "delivery" issue remains challenging especially when patient cells are to be modified either *ex vivo* or *in vivo*. Given their integration-defective character and well-established cell entry mechanisms, episomal viral vectors, i.e., engineered replication-defective viruses, might constitute ideal delivery vehicles for introducing in a targeted and efficient manner gene editing reagents into therapeutically relevant cells. In this regard, **Chapter 6** summarizes the principal characteristics, advantages and drawbacks of the main viral vector systems that are being adapted for genome editing purposes, namely, integration-defective lentiviral vectors (IDLVs), adeno-associated viral vectors (AAVs) and adenoviral vectors (AdVs). As a case-in-point, **Chapter 7** discusses the emerging role of viral vectors in repairing defective dystrophin-encoding genes in mouse models and human cells, whose mutations cause the lethal muscle-wasting X-linked disorder Duchenne muscular dystrophy (DMD).

In conclusion, by investigating the interaction between different types of nucleases (i.e. nicking versus cleaving), donor DNA structures and target chromatin environments, this thesis provides important insights into how to improve the three crucial parameters of genome editing: efficiency, specificity and fidelity. Hence the work presented in this thesis expand the range of possibilities for high-fidelity genetic manipulation of human cells.