Soft genome editing based on CRISPR nickases: it takes one break to tango

Wang, Q.

Genome editing technologies permit introducing specific genetic changes within the vast genomes of living eukaryotic cells *in vitro* and *in vivo*. As such, these technologies are having an ever-increasing impact on both basic and applied science. In the framework of human health, the ultimate goal is that of translating these techniques into therapeutically relevant applications, including those directed at (i) permanently correcting *ex vivo* or *in vivo* mutations associated with hereditary diseases, (ii) interfering with the replication cycle of infectious agents, and (iii) improving the efficacy and safety of cancer immunotherapies that make use of genetically engineered T lymphocytes or natural killer cells as armed "living drugs". Notwithstanding the remarkable progress observed during the past two decades on the development of genome editing tools and strategies, trend fostered after the inception in 2013 of RNA-guide nucleases (RGNs), several longstanding bottlenecks limit the application of these technologies as effective and safe gene and cell therapies. These bottlenecks include large-scale and small-scale mutagenic events (stochastic or otherwise), off-target activities, activation of DNA damage responses and ineffective cellular delivery of the large and multiple reagents required to effectuate the intended chromosomal modification(s) in the proper cell types, tissue or organs.

Typically, genome editing protocols comprise the delivery of sequence-tailored designer nucleases (e.g., CRISPR-Cas9-based RGNs) that, upon targeted double-stranded DNA break (DSB) formation and ensuing activation of endogenous DNA repair pathways, yield specific chromosomal DNA modifications. For the purpose of site-specific chromosomal addition of exogenous genetic information (gene targeting or knock-in), delivery of designer nucleases is combined with that of surrogate donor DNA-repairing templates whose sharing of homology to genomic target sequences, makes them prone to precise homology-directed DNA repair (HDR) processes. Yet, instead of HDR-mediated genome editing, designer nuclease-induced DSBs are more often engaged by competing error-prone DNA repair mechanisms, e.g., non-homologous end joining (NHEJ) and microhomology-mediated end joining (MMEJ). Although numerous studies have shown that HDR-mediated genome editing can be favored through the inhibition of NHEJ and MMEJ factors, it is, in principle, preferable to improve the ectopic HDR process itself rather than interfering with the activity of DNA repair factors whose consequences for genomic stability are mostly unpredictable. Another critical bottleneck concerns the need for effective and safe delivery of the large and multicomponent elements underpinning genome editing procedures. Hence, **Chapter 1**, besides reviewing classical and more recent genome editing tools and strategies, it also covers the use of adenoviral vectors (AdVs) as delivery agents for targeted genetic manipulation of human stem cells, progenitor cells, and their differentiated progenies, focusing on *in vitro* and *ex vivo* protocols. In this context, high-capacity adenoviral vectors (HC-AdVs) deleted of all viral genes constitute particularly valuable vehicles for ferrying large genome editing reagents owing to their low cytotoxicity profile and amenability to cell tropism modifications. Indeed, in this thesis, HC-AdVs displaying CD46-specific capsid fibers from adenovirus serotype-50 instead of coxsackievirus and adenovirus receptor (CAR)-binding fibers from prototypic serotype-5, permitted effective testing of emerging genome editing principles in scientifically and therapeutically relevant CAR-negative human cell types, e.g., mesenchymal stem cells (hMSCs) and muscle progenitor cells (myoblasts) as well as CD46- and CAR-positive induced pluripotent stem cell (iPSC)-derived cardiomyocytes.

Motivated by the aforementioned limitations of commonly used genome editing procedures, strongly associated with their dependency on designer nucleases, the experimental chapters presented in this thesis focus on investigating genome editing principles based on the use of sequence- and strand-specific nucleases ("nickases"). In this regard, **Chapter 2** demonstrates that *in trans* paired nicking (ITPN), comprising simultaneous single-stranded DNA break (SSB) formation at genomic target sites and donor DNA constructs by *Streptococcus pyogenes* CRISPR-Cas9-derived nickases, triggers seamless and scarless HDR-mediated gene knock-ins at endogenous loci whose products are essential for regular cell function, in particular, alleles fundamental for DNA damage responses (i.e., *H2AX* and *PARP1*) and for the maintenance of pluripotency in *bona fide* pluripotent stem cells and iPSCs (i.e., *OCT4*). Importantly, the low mutagenic character of ITPN was shown to preserve target protein dosages and to prevent phenotypic and fitness losses in gene-edited cell populations. In addition, through a collaborative effort, **Chapter 2** introduces the orthogonal high-throughput genome-wide translocation sequencing (oHTGTS) technique for unbiased identification and characterization of off-target sites and effects, respectively, resulting from cleaving versus nicking RGNs. Using oHTGTS it was established that nicking RGNs greatly reduce the frequency of large-scale chromosomal rearrangements and translocations when compared with their DNA cleaving counterparts. Nicking RGNs could nonetheless

induce detectable translocations involving on-target and off-target sites. Presumably such events can arise when, for example, an advancing replication fork collapses after hitting a nickase-induced SSB product. Moreover, given the fact that eukaryotic genomes contain multiple repetitive elements whose individual units share full or high sequence identity with units scattered elsewhere throughout the genome (e.g., retroelements, amplified gene clusters, gene paralogs and pseudogenes), there is a pressing need to identify high-specificity nickases permitting a judicious access to specific chromosomal sequences while averting similar off-target sites. Therefore, in **Chapter 3**, a representative panel of RuvC-disabled *S. pyogenes* Cas9 nickases (SpCas9^{D10A}) was assembled on the basis of the respective high-specificity nucleases, i.e., SpCas9-KA^{D10A}, SpCas9-KARA^{D10A}, eSpCas9(1.1)^{D10A}, Sniper-Cas9^{D10A}, xCas9-3.7^{D10A}, evoCas9^{D10A} and SpCas9-HF1^{D10A}. Subsequent benchmarking experiments and functional screens described in **Chapter 3** identify high-specificity SpCas9^{D10A} variants that can outperform their regular counterparts at the levels of discriminating on-target from off-target sequences and minimizing genome-wide translocations as determined through functional screens and oHTGTS analysis, respectively. Moreover, high-specificity SpCas9^{D10A} nickases operating as dual nicking RGNs also outperformed their conventional counterparts in terms of yielding highly specific gene knockouts and, together with matched donor constructs, achieve specific gene knock-ins by minimizing off-target insertions at similar pseudogene elements. Following from these findings, **Chapter 4** illustrates that high-specificity SpCas9^{D10A} nickases are capable of eliciting ITPN genome editing to the same or higher extents than those triggered by the parental SpCas9^{D10A} protein, including at “safe harbor” loci (e.g., *AAVS1* and *CCR5*) whose HDR-mediated DNA targeting allows for long-term and homogenous transgene expression in engineered cell populations. Critically, **Chapter 4** further shows that, in contrast to regular and high-specificity SpCas9 nucleases, neither regular nor high-specificity SpCas9^{D10A} nickases activate the canonical P53-dependent DNA damage response signaling pathway in human iPSCs, further stressing the potentially higher safety profile of nickases over nucleases for the genomic engineering of cells with high sensitivity to DNA damage, e.g., pluripotent and tissue-specific stem cells. Indeed, these data indicate that SpCas9^{D10A} nickases might offer a heightened safety profile to engineered cell products derived from stem cells as, in addition to cell-cycle arrest and apoptosis, DSB-induced signaling pathways have been associated with the selection of cells bearing mutations in cancer-associated genes, e.g., *TP53* itself and *KRAS*.

As aforementioned, a critical bottleneck regarding the application of genome editing technologies concerns the need for introducing, in an effective and non-cytotoxic manner, the required large and multicomponent reagents into cells, tissues or organs of interest. **Chapter 5** demonstrates that HC-AdVs, in particular CD46-targeting HC-AdVs, are a suitable option for all-in-one delivery of full-length prime editing reagents, in the form of prime editors and prime editing gRNAs (pegRNAs), into human cells regardless of their transformation and replication statuses. Indeed, up to 90% prime editing efficiencies are achievable without overt cytotoxicity in transduced cells. Additionally, a direct correlation between the replication status of target cells and prime editing activities was found by using this cell cycle-independent viral vector delivery platform.

The findings presented in **Chapter 5** are further expanded in **Chapter 6** by leveraging HC-AdVs for delivering advanced prime editing systems designed for installing precise *DMD* gene edits in human myogenic cells. In particular, in myoblasts and mesenchymal stem cells, with efficiencies of up to 80% and 64%, respectively, and in cardiomyocytes differentiated from iPSCs isolated from a Duchenne muscular dystrophy (DMD) patient, with efficiencies of up to 82%. Defective *DMD* alleles underlie DMD (OMIM #310200), a common and lethal X-linked muscle-wasting disorder that afflicts circa 1 in 4,700 boys whose treatment options are, currently, merely palliative. HC-AdV transduction experiments designed for defective *DMD* reading frame repair readily led to the detection of mRNA transcripts encoding proteins corresponding to shortened, yet partially functional, dystrophin variants (i.e., Becker-like dystrophins) in unselected muscle cell populations. Crucially, proximity ligation assays revealed that the resulting Becker-like dystrophin proteins were capable of connecting to β-dystroglycan, a key component of the dystrophin-associated glycoprotein complex located at the sarcolemma of normal muscle cells. Moreover, additional *DMD* reading frame restoration experiments demonstrate the feasibility of leveraging HC-AdV delivery for multiplexing prime editing based on the concerted action of pairs of prime editing complexes. Finally, the straightforward HC-AdV delivery process combined with the non-mutagenic character of prime editing can be exploited for the selective accumulation of precise chromosomal edits in target cell populations through consecutive transduction rounds. Taken together,

the research covered in **Chapter 5** and **Chapter 6** reveals that the integrated delivery of prime editing systems in single HC-AdV particles yields efficient and precise modification of target alleles in human stem/progenitor cells. As a consequence, HC-AdV-assisted prime editing warrants further research and testing, including for the modelling and repairing of genetic defects in *ex vivo* and *in vivo* settings.

The chromatin environment of genomic DNA sequences varies in different cells types and, often, is highly dynamic as a result of the spatiotemporal regulation of epigenetic mechanisms underlying organismal development and cellular differentiation. Interestingly, it is becoming evident that the performance of genome-editing reagents is dependent on a combination of genetic and epigenetic variables, i.e., the target nucleotide sequences *per se* and their epigenetically-regulated chromatin environment, respectively. Possibly, besides controlling to varying degrees the accessibility of target sequences to genome editing tools, the local epigenetic context may also influence DNA repair pathway choices and other DNA editing determinants whose combined effects ultimately define the observed genome modification endpoints. For instance, previous studies obtained in the hosting group have revealed that the activity of designer nucleases based on CRISPR-Cas9 systems and transcription activator-like effectors are significantly hampered by heterochromatic states and that the ratio between HDR and mutagenic NHEJ events can vary in a chromatin context-dependent manner.

Considering that base editors and prime editors are formed by fusing CRISPR nickases to secondary effector domains (i.e., deaminases and reverse transcriptases, respectively), in **Chapter 7**, it is investigated whether there are cause-effect associations between alternate chromatin conformations and genome editing endpoints when using these two powerful DSB-independent genome editing platforms. By implementing complementary loss-of-function and gain-of-function cellular systems, prime editing was found to be frequently hindered at heterochromatin impinged by the KRAB/KAP-1/HP1 axis alone or together with the DNA methyltransferases DNMT3A and DNMT3L. Moreover, the extended portions of gRNAs forming pegRNAs contribute to the underperformance of prime editors at heterochromatic sequences. Notably, in striking contrast with prime editors and designer nucleases (CRISPR-based or otherwise), the DNA editing activity of base editors at closed heterochromatic states ranges in a target site-dependent manner from lower to, often, significantly higher than that observed at open euchromatin. In addition, the chromatin environment of target sequences was also found to be capable of significantly influencing the fidelity and purity of base editing products in a gRNA-dependent manner. As a corollary, the performance and safety profiles of base editing and prime editing technologies necessitates in-depth scrutiny for guiding their selection, further development and application in specific contexts. On the basis of the data presented in **Chapter 7**, one can also submit that algorithms trained to predict the activities of base editing and prime editing reagents, besides inputs on target sequences, will profit from the processing of information regarding the epigenetic context of said sequences. Finally, these data can further guide the development of combinatorial strategies in which targeted epigenetic modulators and DSB-free genome editing tools act in concert for achieving a more efficient and/or more precise genetic modification of cellular (epi)genomes.

In summary, by predominantly investigating genome editing tools and strategies based on CRISPR-Cas9 nickases as such or on their prime editing and base editing derivatives, this thesis provides insights on how these tools and strategies operate in human cells opening up, in this process, new avenues for the seamless modification of cellular (epi)genomes. Moving ahead it is expected that by further developing and refining “soft” genome editing procedures that, besides the efficiency, take into account specificity and accuracy parameters, will allow for translating ‘genomic surgery’ interventions into effective and safe gene and cell therapies.