Chapter 6

Discussion

6.1 Causes of complex lesions

Induction of DSB by Cas9 and subsequent repair are clearly the necessary factors involved in creation of Cas9-induced complex lesions. However, mechanistic details are unknown. Here, I speculate on three mutually non-exclusive factors contributing to complex lesions and briefly review tentative support for them in my data. I also discuss how these factors can be studied further.

First, generation of some complex lesions may primarily involve double-strand resection. This would explain deletions as a simple resection-and-ligation reaction (presumably mediated in part by components responsible for NHEJ). Furthermore, a double-stand resection could expose repeat sequences distal to the cut site. This in turn could lead to SSA repair between exposed direct repeats (resulting in more deletions) and to non-allelic homologous recombination (NAHR) between repeats on different chromatids. NAHR is the proposed mechanism for many large-scale rearrangements in cancer (deletions, duplications, translocations and inversions).

Second, complex lesions could result from HR repair being subverted by simultaneous breakage of both sister chromatids at the same locus. Failing to find an intact repair template, HR may either mediate NAHR or the whole process may revert to MMEJ. Normally, MMEJ involves limited single-strand resection, but during an abortive attempt at HR the DNA may be resected much more extensively. This could expose distal homologies that are not normally available to MMEJ. Synapsis of these homologies could cause large deletions and inversions.

Third, complex lesions could be stimulated by Cas9-intrinsic properties, such as its exonuclease activity and interference with DDR machinery by staying bound to the DNA after making a cut. This may influence the repair process, making relatively non-mutagenic NHEJ and HR less likely, both in favor of more mutagenic processes, such as MMEJ or NAHR.

Data presented in this thesis does not clearly exclude any of the three above mentioned mechanisms. Preliminary analysis has revealed that the amount of exact homology at large deletion breakpoints varies from zero to 12 nt (data not shown), which indicates at least some deletions may be a product of double-strand resection. A more detailed comparison with the expected distribution of homology lengths will be necessary to tell, whether MMEJ is likely to be involved. Moreover, if recombination between repeats was a major driver of large deletions, I would expect them to emerge in "deletion fingerprinting" experiments as reproducible, highly enriched deletion bands. While these were not observed, it may be either due to low sensitivity of the assay or due to low repeat content of the studied *PigA* locus. Further analysis to explore inexact homologies (i.e. stretches of homology with infrequent mismatches, also called "homeologies") at deletion breakpoints could also help decide, if repeat recombination is involved. Finally, many small, locally templated insertions and small, non-contiguous lesions implicate single-strand resection and MMEJ.

Dissection of requirements for complex lesions by means of genetic screens based on the assay described in chapter 4 is warranted. Further-

more, translocations could potentially be enriched for and studied separately by targeting very long introns and selecting for loss of gene expression (although such selection will also enrich for very large deletions and inversions). Systematic study of different loci in the genome in this way would allow better understanding of targets of translocations and thus proper risk estimation in the context of gene therapy. A recent report found that distal insertions induced by Cas9 (i.e. mapping to other loci in the genome than the edited locus) are enriched for sequences close to the cut site in 3D space (Leenay et al., 2018). I predict this would also be the case for translocations. Finally, careful comparison between different precision nucleases may shed some light on whether Cas9-intrinsic properties are contributing to creation of complex lesions. However, such comparisons are often difficult due to many confounding factors, such as differential levels of protein expression and nuclease acitivity.

6.2 Ways to avoid complex lesions

Complex lesions are generally an undesirable outcome of gene editing, both in gene therapy (where they are potentially pathogenic, as discussed in chapter 5) and in basic research (where they may be confounding and difficult to genotype). Based on the discussed causes of complex lesions, I propose four broad ways of tackling this issue - complete avoidance (or more controlled induction) of the DSB, avoidance of simultaneous chromatid breakage, manipulation of DDR and re-engineering of the wild-type Cas9.

First, editing without DSB creation is possible in principle by using base editors. In practice, indels consistent with DSB creation are observed upon base editing at a rate of around 1% (Gaudelli et al., 2017; Komor et al., 2016). Further engineering of base editing tools (e.g. using deactivated instead of nickase Cas9, as in the first generation of base editing Cas9) may help abolish DSBs completely or at least reduce them to marginal levels. Another potential solution to the DSB problem is creation of the break in a controlled fashion, as for example during V(D)J recombination. These reactions do still lead to carcinogenic translocations, but at a very low rate (Alt et al., 2013). Discovery or engineering of a programmable recombinase could achieve this goal. Initial foray in this area has been made by fusing Cas9 and Gin recombinase, but this enzyme only operates on specific recombinase recognition sites (Chaikind et al., 2016). Finally, precise modification of the DNA has been demonstrated in yeast transfected with ssDNA complementary to the lagging strand (Barbieri et al., 2017). This approach could potentially be adapted to human cells. However, no systematic safety assessment of this method has been conducted yet.

Second, ensuring that only one sister chromatid is broken at any given time could be achieved e.g. by reducing the effective concentration of the nuclease, choosing low efficiency guides or reducing the activity of the Cas9-gRNA complex by engineering of its components. This does not necessarily require having one Cas9 molecule per cell, but only that on the average the activity should be low enough to induce one cut every few hours, to allow HR to finish the repair process. While initially the editing efficiency would drop, it may be possible to maintain a level of nuclease activity that is low enough to avoid breaking both chromatids, but high enough to keep the reaction going despite creation of new wild-type alleles due to DNA replication. Alternatively, the cutting could be confined to non-replicative stages of the cell cycles, e.g. by coupling Cas9 to cell stage specific degrons (Huang et al., 2017) or timing the delivery in synchronized cells (for analogous approaches trying to increase HR repair see Gutschner et al., 2016; Lin et al., 2014; Yang et al., 2016a).

Third, blocking excessively mutagenic DSB repair pathways or promoting 'safe' ones could be a promising way of avoiding complex lesions, providing these mechanisms are well understood. In basic research, such modulation could be achieved by overexpression and silencing of specific DDR proteins. However, this could have unexpected consequences at other loci than the edited one.

Coupling Cas9 to proteins involved in DNA repair or to the repair template (if templated repair is the goal, e.g. Savic et al., 2018; Shou et al., 2018) is a more "topical" solution, which may be compatible with gene therapy.

Fourth, re-engineering of Cas9 to abolish exonuclease activity or make it release DNA after cut could potentially improve its risk profile. A molecular evolution approach could perhaps be employed, if selective conditions against exonuclease activity / extended binding and for high endonucleotic activity can be obtained. However, there is a risk that these properties are inextricably linked. In particular, while a simple DSB would predominantly be repaired with no change to DNA sequence, a more complex, Cas9-blocked DSB may elicit more vigorous, mutagenic repair mechanisms.

6.3 Ways to exploit complex lesions

Cas9-induced cross-overs have already been used in yeast to enable genetic mapping at higher resolutions than allowed by natural recombination rate (Sadhu et al., 2016). "Distal" translocations, which cause a more dramatic reshufling of the genome, could be used to investigate the effect of putting different loci in a linkage disequilibrium and the significance of the particular chromosomal setup of a given organism. Other forms of genome engineering are already being used in this field, for example in a recent work in which all of yeast chromosomes were combined into one (or two) units (Luo et al., 2018a; Shao et al., 2018).

Single gRNA-induced deletions could be a useful tool for studying non-coding elements (enhancers, long non-coding RNAs etc.). In contrast to coding genes, which can be inactivated by introduction of a small, frameshifting indel, non-coding elements usually require more extensive mutagenesis. Such approach could complement currently available tools, which use CRISPRi, saturation mutagenesis or paired gRNA-induced deletions (Aparicio-Prat et al., 2015; Canver et al., 2015; Gasperini et al., 2017, 2018; Korkmaz et al., 2016; Zhu et al., 2016). Due to its simplicity, it could potentially match the throughput of the CRISPRi approach, if multiple gRNAs are multiplex per cell (despite the confounding risk of translocations between multiple cut sites). This would make it possible to study the effect of tens of thousands deletions in one experiment. Such deletions could also be isolated and studied in detail, as opposed to less well defined chromatin silencing induced by CRISPRi.

Single gRNA-induced deletions of varying size could be exploited to create genomic "deletion series". Normally, exonic deletions are created through Cre recombination between lox sites flanking the exon. This approach offers a high degree of precision. However, that also means potential confounding factor may not be discovered, if e.g. a regulatory element is consistently co-deleted with the exon. An exonic deletion series could easily be created by single gRNA in exons of moderate size (100-5000 bp). By introducing slight variability in resulting genotypes, this approach could make the experiment more robust. Another application of this technique could be to investigate the extent and the function of different protein domains. Normally, this is done by cloning of the cDNA of interest ("DNA complementary to the mRNA", containing only the coding part of the gene without introns) into a plasmid and deleting various elements in vitro. The modified product is then usually transiently expressed from the plasmid in the cells knocked-out for the endogenous gene. Using single gRNAs, a deletion series could be created either directly in the wild-type gene (if studying large exons) or at a locus in which the endogenous gene was replaced with its cDNA copy. The advantage of this approach over the current solution is that it could be done at a larger scale and in the endogenous regulatory context.

6.4 Probing protein isoform diversity using CRISPR/Cas9-based assays

I have detected a variety of discrete protein expression levels following Cas9 mutagenesis at the *Cd9* locus. Although I have not investigated them

in detail, it is likely they represent stable isoforms brought about by specific types of genomic damage (loss of exon, out-of-frame mutation, epitope modifying mutation). Systematic mutagenesis of the coding gene coupled to a continuous protein-related readout (e.g. abundance of protein measured by a flow cytometry, functional assay or depletion of specifically edited cells in case of an essential genes) could therefore be used as a tool to study the plasticity of protein isoforms. Abundance, and ideally the specific sequence, of mRNA could be measured in the same assay to provide additional layer of information and decouple transcriptional effects. In particular, such an assay would enable investigation of general splicing and folding rules, stability requirements, relationship between conservation and essentiality of specific protein domains as well as epitope malleability. A similar procedure investigating functionality of so-called variants of unknown significance in the essential *Brca1* gene has recently been described (Findlay et al., 2018; Starita et al., 2018).

6.5 New methods for genotyping of complex lesions

Reliable, unbiased, high throughput methods for genotyping of complex lesions are necessary to understand them in more detail, to monitor gene therapy applications and to guide the development of preventative measures.

Methods I used in this study suffer from a number of shortcomings. Single cell cloning and Sanger sequencing are low throughput. Flow cytometric assay can only be applied to a subset of transcriptionally active loci. PacBio sequencing of bulk DNA results in biased readouts. New methods need to be developed to enable more in-depth study and systematic monitoring of complex lesions at other loci, in other cell types or using other nucleases. Barcoding of PCR products in early cycles could improve PacBio readout by removing amplification and sequencing biases as well as allowing accurate genotyping of small indels. Oligos preventing amplification of wild-type and small indel products could be used to enrich for complex lesions without the need for selection based on a loss of gene expression. Depletion of wild-type products could also be performed post-amplification, e.g. by hybridization with complementary RNA and digestion using duplex-specific nuclease (Zhulidov et al., 2004) or using wild-type sequence-specific oligos conjugated to magnetic beads. Copy number profiling using qPCR would allow quick assessment of the extent of large deletions. RNA quantification by qPCR and full-length RNA isoform sequencing could also be used to approximately quantify the frequency of complex lesions (incl. translocations and small non-contiguous lesions) in genes whose activity cannot be assessed by flow cytometry.

6.6 Complex lesions and risk management

Direct consequences of complex lesions were discussed in chapter 5. Here, I discuss feasibility of Cas9 usage in basic research and gene therapy given these consequences.

As mentioned earlier, complex lesions can be considered a nuisance in most basic research applications. The presence of a large deletion, insertion, inversion or translocation can often be readily detected as failure to amplify any mutant genotype in the offspring of edited animals crossed to the wild-type. However, it does incurr a cost in time and money spent breeding and genotyping unsuitable lesions. Complex lesions can compromise interpretation of multiplexed editing, if animals are not sufficiently backcrossed to the wild-type, as lack of mutant detection in F1 animals cannot be interpreted as lack of editing. Non-contiguous lesion in F1 animals may also go unnoticed, unless a proper rescue experiment is performed. These problems compound when using single cell cloned cell lines, whose alleles cannot be separated by breeding. In practical terms, complex lesions highlight the necessity for well established experimental controls - careful assessment of expression and functional phenotypes, use of independently derived cell clones or edited animals and rescue experiments.

In terms of gene therapy, two factors seem crucial - whether editing is somatic or germline and whether the purpose of the gene editing experiment is curative or prophylactic. Whether the therapy is conducted ex vivo or in vivo is another important factor, since the former allows a higher degree of quality control and may reduce the risk of affecting long-lived proliferative stem cells. Currently, the discovery of complex lesions, our general understanding of off-target effects and dearth of studies on long-term consequences of gene editing in relevant animal models (such as monkeys) or in humans suggests extreme caution when considering gene therapeutic application beyond life-saving interventions. Therapies for terminal diseases (incl. most forms of cancer), where risk of cancer induced by Cas9 is outweighed by the imminent risk of death, fall squarely into the life-saving category. Consistently, therapies for cancer using ex vivo edited T cells are currently a major focus of gene therapeutic clinical trials. There is also a "grey zone" of applications that may not be life-saving, but are life-changing. One example may be the potential cure for patients with Hunter's or Hurler's syndrome, in which case a physician may elect together with the patient to accept the risk of carcinogenesis in exchange for a potential cure. Finally, prophylactic attempts aimed at reducing cholesterol levels (*Pcsk9* editing, not yet in clinical trials) or prevent (but not cure) HIV infection (*Ccr5* editing) and most forms of therapeutic embryo editing, which is bound to influence the germline (especially, when preimplantation diagnostics offers a viable alternative), may need to be deferred until our understanding of Cas9 mutagenesis improves. Finally, I want to acknowledge that the ultimate decision as to if and when to allow usage of Cas9 and other precision nucleases on patients should be primarily in the hands of people with extensive experience in healthcare related risk management - physicians running clinical trials and regulators specializing in drug safety.