We accomplished the following:

In the first task we aim to understand the working mechanism of Cas9 and Cas12a nucleases including its specificity and target-selectivity. The results of these works also led to the generation of new variants which in turn help to reveal further features of these nucleases. Our results concerning both cas12a and Cas9 variants are explained in the next two paragraphs:

The widespread use of Cas12a (formerly Cpf1) nucleases for genome engineering is limited by their requirement for a rather long TTTV protospacer adjacent motif (PAM) sequence. Here we have aimed to loosen these PAM constraints and have generated new PAM mutant variants of the four Cas12a orthologs that are active in mammalian and plant cells, by combining the mutations of their corresponding RR and RVR variants with altered PAM specificities. LbCas12a-RVRR showing the highest activity was selected for an in-depth characterization of its PAM preferences in mammalian cells, using a plasmid-based assay. The consensus PAM sequence of LbCas12a-RVRR resembles a TNTN motif, but also includes TACV, TTCV CTCV and CCCV. Together with its enhanced alternative their target specificity complements the recently developed enAsCas12a in several applications. Disruption assays, NGS analysis, transcriptome modulation and base editing assay foreshadow the widespread use of this variant for genome editing applications and transcriptome analysis.

Increased fidelity mutants of the SpCas9 nuclease constitute the most promising approach to mitigating its off-target effects. However, these variants are effective only on a restricted target space, and many of them are reported to work less efficiently when applied in clinically relevant, pre-assembled, ribonucleoprotein forms. The low tolerance to 5'-extended 21G-sgRNAs contributes to a great extent to their decreased performance. To accommodate a 5'-extension, we generated/selected "Blackjack" mutations that increase the fidelity of WT-SpCas9, while keeping it effective with 21G-sgRNAs. When combined, Blackjack mutations cause the same effect to the other increased fidelity variants. We developed two "Blackjack" variants, eSpCas9-plus and SpCas9-HF1-plus that are superior variants of eSpCas9 and SpCas9-HF1, respectively, possessing matching on-target activity and fidelity but retaining 20G-level activity with 21G-sgRNAs. They facilitate the use of the existing pooled sgRNA libraries with higher specificity and show similar activities when delivered either as plasmids or as pre-assembled ribonucleoproteins.

In the second task we aim to understand a sequence specificity of some of the Cas9 variants. We characterized the sequence specificity of WT SPCas9 and its high-fidelity variant, SpCas9-HF1. At first, we used a bacterial selection system to characterize their activity as described in the following paragraph:

Detailed target-selectivity information and experiment-based efficacy prediction tools of Cas9s are available only for Streptococcus pyogenes Cas9 (SpCas9) and three of its orthologs and mutant variants. One obstacle in developing such tools is the rarity of accurate data. Here, we developed a method termed "Self-targeting gRNA Library Screen" (SLS) for assaying the activity of Cas9 nucleases in bacteria by random target/gRNA libraries of self-targeting gRNAs. Exploiting more than a million different sequences, we demonstrated the use of the method on SpCas9-HF1 variant to analyze its activity and reveal motifs that influence its target-selectivity. We have also developed an algorithm for predicting the activity of SpCas9-HF1 with an accuracy matching those of existing tools. SLS is a facile alternative to the existing much more expensive and laborious approaches and has the capability to deliver sufficient amount of data for most of the orthologs and variants of SpCas9.

During the current reporting period we worked on understanding the factors determining the efficiency and fidelity of target cleavage of the wild type and of the mutant variant SpCas9 nucleases (Task1.) We generated a fluorescent reporter system (Base Editor Activity Reporter; BEAR) to monitor the activity of SpCas9-based base editors in a semi-high throughput manner as well for increasing the efficiency and decrease the editing specificity of ABE and CBE. Specifically, using our approach we managed to increase the editing outcome of Adenine Base Editor (ABE) and Cytosine Base Editor (CBE) using the inactive SpCas9 variants to the outcome detectable when ABE and CBE applied with the nickase variant of SpCas9. The advantage of this result is that we could drastically reduce the unwanted indel background that accompanies the use of nickase base editors. Beside using BEAR to enrich the successfully base-edited cells in the population, using the increased-fidelity variants of Spcas9 we managed to reveal some interesting feature of ABE and CBE. Our results suggest that ABE needs a more full-length separation of the DNA strands of the targets while CBE is active on targets where the separation of the two strands of the DNA is less effective. This has practical consequence for their usage. CBE is active on off-target sequences where ABE is not, thus ABE shows higher specificity editing compared to CBE. In addition, increased-fidelity nucleases are more effective to increase the specificity of ABE than to increase the specificity of CBE. The mechanistic inside we gained helps to develop more specific CBE variants.

We also reach progress in deciphering factors governing the specificities of Cas nucleases in general (Task 2). We systematically examined the effect of self-complementarity either between the spacer sequence and the so called scaffold parts of the guide-RNA or within the spacer sequence. We have identified sequences among other features of the motives that are detrimental to the activity of the nucleases.

To facilitate efficient multiplex genome editing, we generated systems based on either NHEJ-mediated or HDR-mediated integration of fluorescent marker proteins. We have optimized the integration process by testing and selecting appropriate guides which exhibit appropriately high integration. We also incorporated an IRFP protein to each plasmid which are directed to the nucleus for easy segmenting of the cells.

Our research in this interim reporting period has focused on 3 sub-tasks. Firstly, we have recognized that the effect of mutations in increased fidelity variants of the SpCas9 nuclease generates a kind of monotonic order of increasing precision and correspondingly decreasing activity. On the other hand, we recognized that, although the ability of SpCas9 targets to be cut by nuclease variants is very different, it depends on the position of the nucleases in the fidelity ranking. Thus, the most accurate nucleases in the fidelity ranking that are still active enough to cut the target of interest will be the optimal variants that can cut the target of interest with the highest specificity, i.e., the least off-target effect. We have developed a method to facilitate the determination of the optimal nuclease variant for a given target, which has allowed us to greatly increase the specificity of gene editing.
On the other hand, we recognized from the analysis of the one million sequence library that the spacer-dependent preference of SpCas9 for sequence motifs is predominantly due to the effect of scaffold-complementary motifs in the spacer.
Finally, a reporter assay was developed to monitor prime editor activity in cells by fluorescence and flow cytometry. This will allow enrichment of successfully edited cells in the edited population and to investigate prime editing sequence dependencies by systematically varying the target sequence.

In the last reporting period we finalized the work we did in the previous period and published an article in eLIFE. For the other two studies our manuscripts are in revision phase at NAR and Nature Communication. Furthermore, we characterized the SuperFi SpCas9 variant and developed new base editor with distinct efficiency and specificity profile. This work has been published in Nature Communication.