RESEARCH   **Open Access**

A large-scale whole-genome sequencing analysis reveals highly specific genome editing by both Cas9 and Cpf1 (Cas12a) nucleases in rice

Xu Tang[1†], Guanqing Liu[2,3†], Jianping Zhou[1†], Qiurong Ren[1], Qi You[2,3], Li Tian[1], Xuhui Xin[1], Zhaohui Zhong[1], Binglin Liu[1], Xuelian Zheng[1], Dengwei Zhang[1], Aimee Malzahn[4], Zhiyun Gong[2], Yiping Qi[4,5*], Tao Zhang[2,3*] and Yong Zhang[1*]

Abstract

Background: Targeting specificity has been a barrier to applying genome editing systems in functional genomics, precise medicine and plant breeding. In plants, only limited studies have used whole-genome sequencing (WGS) to test off-target effects of Cas9. The cause of numerous discovered mutations is still controversial. Furthermore, WGS-based off-target analysis of Cpf1 (Cas12a) has not been reported in any higher organism to date.

Results: We conduct a WGS analysis of 34 plants edited by Cas9 and 15 plants edited by Cpf1 in T0 and T1 generations along with 20 diverse control plants in rice. The sequencing depths range from 45× to 105× with read mapping rates above 96%. Our results clearly show that most mutations in edited plants are created by the tissue culture process, which causes approximately 102 to 148 single nucleotide variations (SNVs) and approximately 32 to 83 insertions/deletions (indels) per plant. Among 12 Cas9 single guide RNAs (sgRNAs) and three Cpf1 CRISPR RNAs (crRNAs) assessed by WGS, only one Cas9 sgRNA resulted in off-target mutations in T0 lines at sites predicted by computer programs. Moreover, we cannot find evidence for bona fide off-target mutations due to continued expression of Cas9 or Cpf1 with guide RNAs in T1 generation.

Conclusions: Our comprehensive and rigorous analysis of WGS data across multiple sample types suggests both Cas9 and Cpf1 nucleases are very specific in generating targeted DNA modifications and off-targeting can be avoided by designing guide RNAs with high specificity.

* Correspondence: Yiping@umd.edu; zhangtao@yzu.edu.cn;
zhangyong916@uestc.edu.cn
†Xu Tang, Guanqing Liu and Jianping Zhou contributed equally to this work.
⁴Department of Plant Science and Landscape Architecture, University of
Maryland, College Park, MD 20742, USA
²Jiangsu Key Laboratory of Crop Genetics and Physiology, Co-Innovation
Center for Modern Production Technology of Grain Crops, Key Laboratory of
Plant Functional Genomics of the Ministry of Education, Yangzhou University,
Yangzhou 225009, China
¹Department of Biotechnology, School of Life Sciences and Technology,
Center for Informational Biology, University of Electronic Science and
Technology of China, Room 216, Main Building, No. 4, Section 2, North
Jianshe Road, Chengdu 610054, People's Republic of China
Full list of author information is available at the end of the article

Tang *et al. Genome Biology* (2018) 19:84

Page 2 of 13

Background

Bacterial type II CRISPR-Cas9 systems can effectively induce RNA-guided DNA double strand breaks (DSBs) [1], making them popular tools for genome editing in bacteria [2], animal cells [3], mammalian systems [4–7], and plants [8–11]. The most widely used *Streptococcus pyogenes* Cas9 (SpCas9) uses ~ 20 nucleotides (nt) of a single guide RNA (sgRNA) to recognize a complementary target DNA site along with an NGG protospacer adjacent motif (PAM) [1, 12]. More recently, type V CRISPR-Cpf1 (CRISPR-Cas12a) was shown to mediate efficient genome editing in human cells [13] and plants [14–16]. Cpf1(Cas12a) uses ~ 23 nt of an RNA guide to target DNA with a TTTV PAM [13]. RNA-guided nucleases (RGNs) such as Cas9 and Cpf1 represent versatile genome editing tools that promise to advance basic science, enable personalized medicine, and accelerate crop breeding. However, Cas9 may cause undesired off-target mutations due to sgRNAs recognizing DNA sequences with one to a few nucleotide mismatches, albeit with reduced nuclease binding and cleavage activity [1, 6, 17, 18]. Although similar rules apply to Cpf1, recent studies in human cells [19, 20] have shown Cpf1 is generally more specific than Cas9.

Understanding the scope of off-target mutations in Cas9- or Cpf1-edited crops is critical for research and regulation. Previously, whole-genome sequencing (WGS) was applied for detecting off-target mutations by Cas9 in *Arabidopsis* [21], rice [22], and tomato [23]. Unfortunately, these studies either only looked at potential off-target sites predicted by computer programs or fell short of full analysis of all the mutations identified by WGS in edited plants. Without inclusion of enough necessary controls, such WGS studies had limited power for isolating off-target mutations in edited plants because they were unable to fully assess the levels of preexisting mutations, spontaneous mutations, and mutations caused by tissue culture- and *Agrobacterium*-mediated transformation. Genome-wide identification of off-target mutations by Cas9 or Cpf1 will be empowered only if all background mutations can be isolated. Furthermore, WGS-based off-target analysis of Cpf1 has not been reported in any higher organism. In recent years, WGS studies on Cas9-edited mice have generated contrasting results; one study found few off-target mutations [24] while another found many [25]. This controversy raised the urgency for comprehensive and rigorous analyses of off-target mutations using WGS in edited animals and plants. We reasoned a large-scale and well-designed study is required for comprehensive assessment of off-target effects in crops by Cas9 and Cpf1, two leading CRISPR genome editing systems. Here, we describe a large-scale WGS study to assess off-target effects of Cas9 and Cpf1 in rice, an important food crop.

Our results suggest off-target mutations of Cas9 and Cpf1 are largely negligible when compared to spontaneous mutations or mutations caused by tissue culture and *Agrobacterium* infection in edited plants. The resulting knowledge is likely to serve as an important reference for plant researchers and regulatory agencies.

Results

Detection of off-target, spontaneous, and background mutations

To comprehensively evaluate potential off-target effects of Cas9 in rice, we generated ten T-DNA constructs to target seven genes with 12 sgRNAs, including two dual-sgRNA constructs for editing two circular RNA loci (Additional file 1: Figure S1 and Additional file 2: Table S1). All ten CRISPR-Cas9 nuclease expression constructs were active at target sites and resulted in editing frequencies ranging from 15 to 100% in T0 lines (Fig. 1a, b and Additional file 2: Table S1). For each Cas9 construct, two independent T0 plants carrying non-mosaic mutations (Additional file 1: Figure S1) were chosen for WGS. To assess off-target effects of Cpf1, we followed three previously published *Lachnospiraceae bacterium* ND2006 Cpf1 (LbCpf1) targeting constructs that resulted in 100% editing efficiency in T0 lines [14] (Additional file 1: Figure S2 and Additional file 2: Table S1). Two Cpf1 T0 plants per construct carrying non-mosaic on-target mutations were chosen for WGS (Additional file 1: Figure S2). For four T0 lines edited by four different Cas9 sgRNAs and two T0 lines edited by two different Cpf1 crRNAs, we selected two to five plants from each T0 line in the T1 generation for WGS (Fig. 1a, b). In addition, four wild-type (WT) plants each from three consecutive generations were also included for WGS to survey spontaneous mutations (Fig. 1b). To ensure high confidence on base calling, all 69 individual plants were sequenced at 45× to 105× in depth (Additional file 2: Table S2). A stringent mutation mapping and calling pipeline was developed for WGS analysis (Fig. 1c). Single-nucleotide variants (SNVs) and small insertions and deletions (indels) were each identified with three variant-calling software programs, with high-confident variants shared by all software being further analyzed for mutation identification (Fig. 1c). Based on our criteria, mutations with frequencies below 10% may not be called out, as such low frequency mutations may have resulted from sequencing errors.

To survey pre-existing mutations in the WT population and estimate the level of spontaneous mutations across generations, we analyzed the WGS data from 12 WT plants across three consecutive generations (Fig. 1b and Additional file 1: Figure S3). After filtering shared pre-existing mutations, we estimated an average of 23 SNVs and 18 indels as spontaneous mutations from
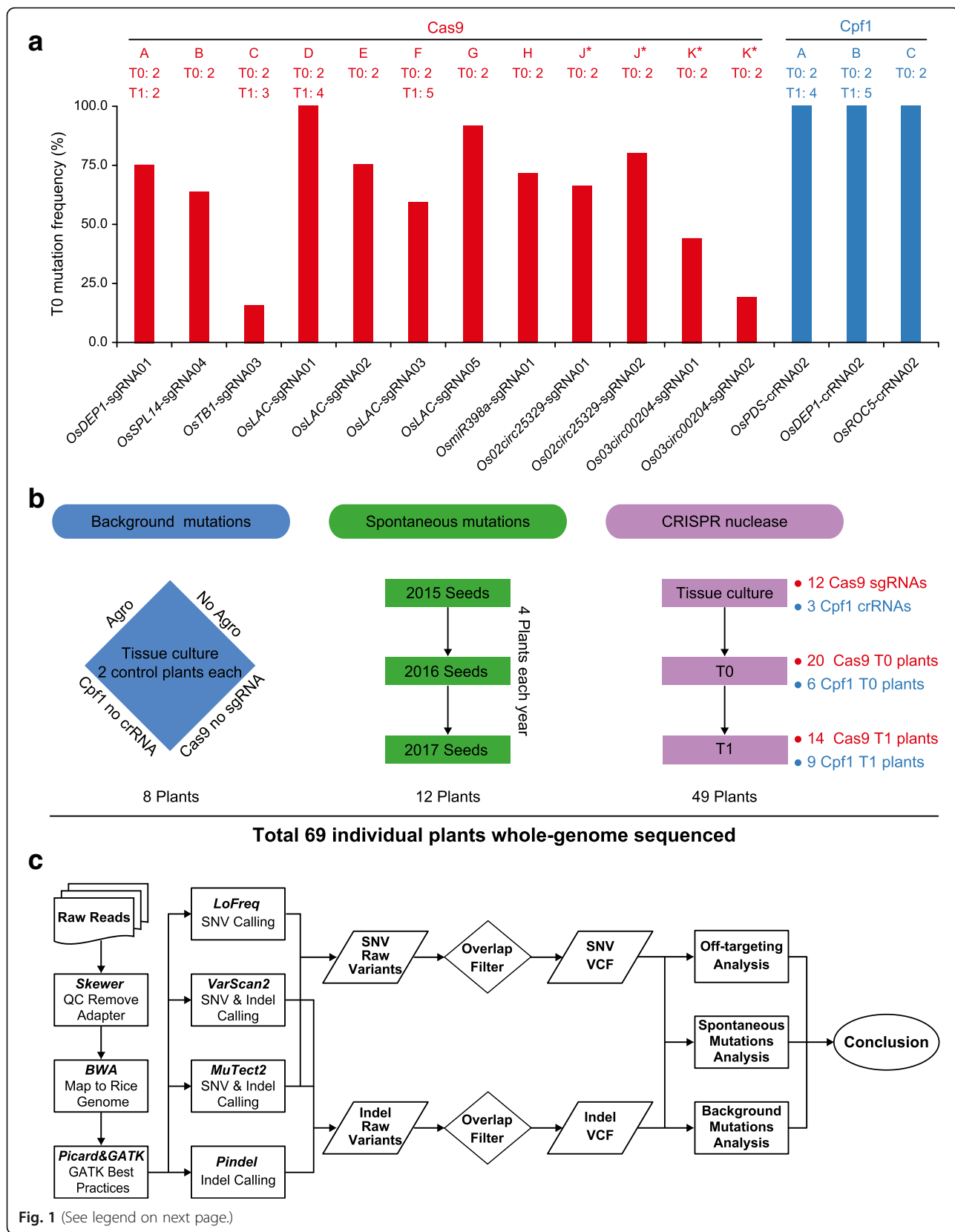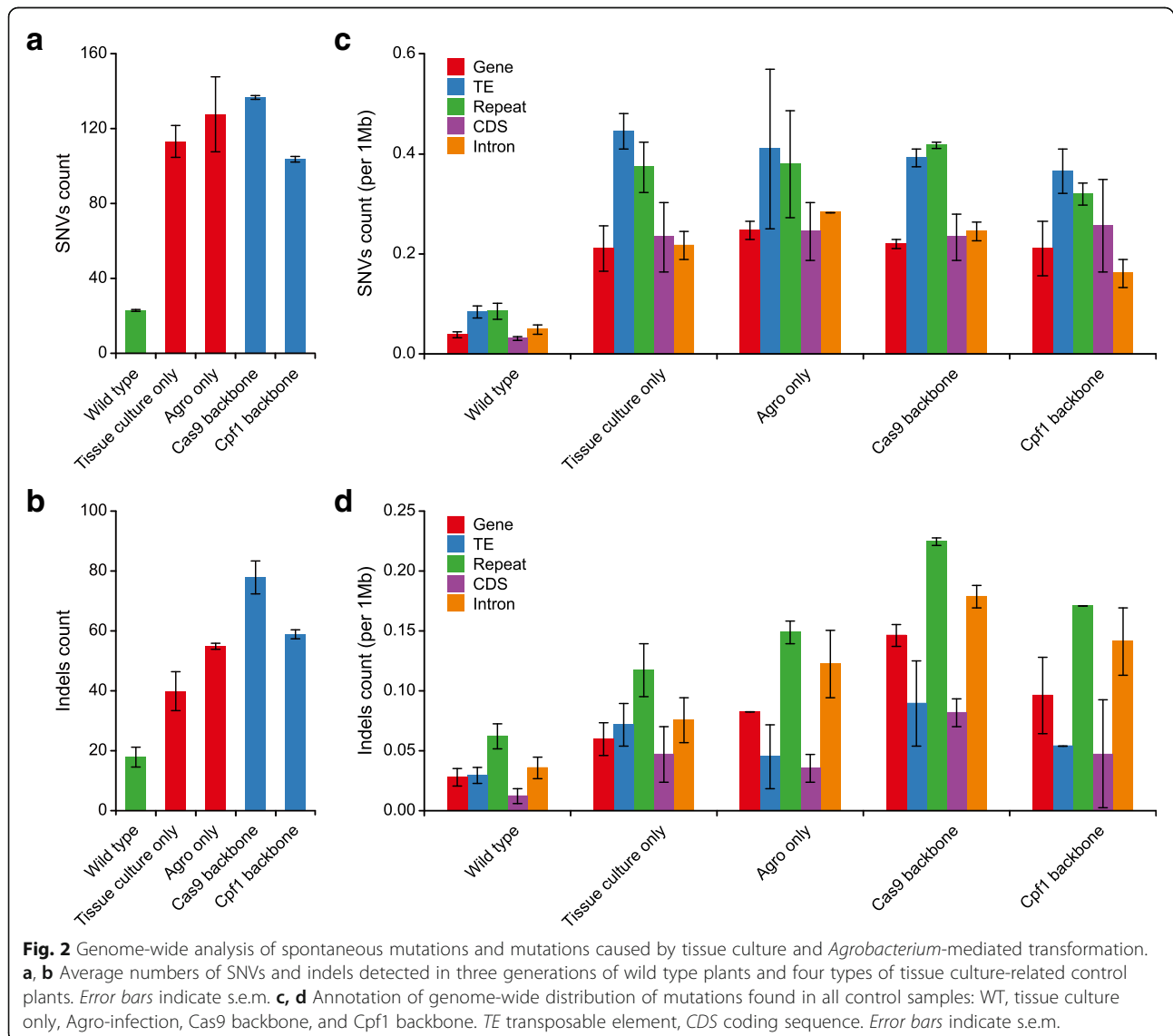
Tang *et al. Genome Biology* (2018) 19:84

Page 3 of 13

Fig. 1 (See legend on next page.)

Tang *et al. Genome Biology* (2018) 19:84

Page 4 of 13

(See figure on previous page.)
Fig. 1 Experimental design and work flow. **a** Genome editing efficiency at selected 12 Cas9 and three Cpf1 target sites in T0 rice plants. The x-axis shows the names of sgRNAs and crRNAs which are denoted as Cas9-A to Cas9-K and Cpf1-A to Cpf1-C. The numbers of T0 and/or T1 lines that are subjected to whole-genome sequencing (WGS) are indicated. The y-axis shows genome editing frequencies calculated based on genotyping data in T0 generation. Cas9-J* and Cas9-K* samples each express a dual-sgRNA construct, targeting two genes simultaneously. **b** Selection of plants for WGS. *Left*: four groups of controls are included for assessing different background mutations. *Middle*: three generations of wild-type plants are included for assessing parent–progeny spontaneous mutations. *Right*: multiple T0 and T1 lines edited by Cas9 and Cpf1 are chosen for assessing off-targeting by WGS. **c** Workflow of whole-genome detection of SNV and indel mutations. SNV analysis involves using three computer programs: LoFreq, VarScan2, and MuTect2. Indel analysis also involves using three programs: VarScan2, MuTect2, and Pindel

parents to progeny in rice (Fig. 2a, b). We calculated the spontaneous mutation rate at $\sim 5.4 \times 10^{-8}$ per site per diploid genome per generation, which is in line with the rates previously reported in maize (2.2–3.9×10^{-8}) [26] but higher than the rate in *Arabidopsis* (7–7.4×10^{-9}) [27, 28].

To assess mutations generated by tissue culture and *Agrobacterium* infection, we produced and sequenced four types of control plants: tissue culture only, tissue culture with *Agrobacterium*, tissue culture with *Agrobacterium* transformation of Cas9 without sgRNA, and tissue culture with *Agrobacterium* transformation of Cpf1

Fig. 2 Genome-wide analysis of spontaneous mutations and mutations caused by tissue culture and *Agrobacterium*-mediated transformation. **a**, **b** Average numbers of SNVs and indels detected in three generations of wild type plants and four types of tissue culture-related control plants. *Error bars* indicate s.e.m. **c**, **d** Annotation of genome-wide distribution of mutations found in all control samples: WT, tissue culture only, Agro-infection, Cas9 backbone, and Cpf1 backbone. *TE* transposable element, *CDS* coding sequence. *Error bars* indicate s.e.m.

Tang *et al. Genome Biology* (2018) 19:84

Page 5 of 13

without crRNA (Fig. 1b and Additional file 1: Figure S4). Tissue culture is known to be mutagenic, causing somaclonal variations [29]. Indeed, the two tissue culture-only samples contained an average of 114 SNVs and 36 indels (Fig. 2a, b), resulting in a background mutation rate of 1.86×10^{-7}, which is similar to the rates (1.7–3.3×10^{-7}) previously published [30]. Importantly, similar numbers of SNVs were observed from *Agrobacterium*-infected or Cas9/Cpf1 backbone-transformed plants (Fig. 2a). These three controls generated ~ 15 to 41 more indels compared to tissue culture-only samples (Fig. 2b), suggesting *Agrobacterium* infection is mutagenic with a preference for introducing indels. This warrants further investigation as these three controls show large variations on indel counts. We mapped all identified mutations from these four control types to the rice genome across 12 chromosomes (Additional file 1: Figure S5). Further analysis of the genome-wide distribution of these background mutations revealed high enrichment of SNVs in transposable elements (TEs) and repeats (Fig. 2c), as well as high enrichment of indels in repeats (Fig. 2d).

SNVs and indels identified in edited T0 plants are largely background mutations

WGS of 20 Cas9 and six Cpf1-edited T0 lines confirmed all target site mutations that were initially identified with Sanger sequencing (Additional file 1: Figures S1 and S2 and Additional file 2: Table S1). We identified SNVs and indels in these Cas9 T0 lines (Additional file 1: Figure S6) and Cpf1 T0 lines (Additional file 1: Figure S7) and mapped these mutations to the rice genome (Additional file 1: Figure S5). We found their numbers are close to those in Cas9 or Cpf1 backbone controls, with about twice as many SNVs as indels (Fig. 3a, b). This mutation pattern is not consistent with Cas9- or Cpf1-generated mutations in rice which are largely indels [9, 14]. For example, all target site mutations in these selected 26 T0 lines are indels (Additional file 1: Figures S1 and S2 and Additional file 2: Table S1). The SNV and indel mutations in Cas9- and Cpf1-edited T0 samples share similar genome-wide distribution with the tissue culture-related controls (Additional file 1: Figure S5). We identified a total of 31 T-DNA insertion events in 26 T0 lines and found T-DNA copy numbers ranging from 1 to 3; most T0 lines had only one T-DNA insertion (Additional file 1: Figure S8). No significant difference was found for the numbers of SNVs and indels among T0 lines with different T-DNA copy numbers (Fig. 3c, d). Cas9-J and Cas9-K T0 lines each expressed a dual-sgRNA construct for simultaneous expression of two sgRNAs, targeting two putative circle RNA genes (Fig. 1a). No significant difference was found for the numbers of SNVs and indels in these four dual-sgRNA lines and the other 22 single sgRNA lines (Fig. 3e, f). Moreover, there is no correlation between the

numbers of SNVs or indels and the on-target editing efficiency by Cas9 or Cpf1 in these T0 plants (Fig. 3g, h). All these analyses strongly suggest mutations in these genome-edited T0 lines are mostly background mutations caused during tissue culture and *Agrobacterium*-mediated transformation.

Identification of true off-target mutations in T0 plants

To identify true off-target mutations in the T0 plants, we first evaluated the specificity of 12 sgRNAs of Cas9 and three crRNAs of Cpf1 with CRISPOR [31] and Cas-OFFinder [32]. With a stringent criterion allowing only a 1-nt mismatch in the protospacer, three Cas9 sgRNAs (Cas9-D, Cas9-E, and Cas9-J-sgRNA01; Fig. 1a) had predicted off-target sites (Fig. 4a and Additional file 2: Table S3). When we mapped all identified mutations to these potential off-target sites by allowing up to 10-nt mismatches to the protospacers of Cas9 (Additional file 1: Figure S9) and Cpf1 (Additional file 3: Figure S10), only Cas9-J-sgRNA01 showed evidence of true off-targeting. It is worth noting that these off-target sites showed high sequence homology to the Cas9-J-sgRNA01 target site and could be accurately predicted by software such as CRISPOR and Cas-OFFinder (Additional file 2: Table S3). We reasoned true off-target mutations are likely to occur separately in independent T0 lines. Indeed, among 12 off-target sites identified for Cas9-J-sgRNA01, seven sites were overlapped between two T0 lines while the remaining five sites were only validated from one T0 line (Fig. 4b, c). All 12 off-target sites show very high sequence homology with the target site (Fig. 4c). Among them, one site at Chr1:22043904 is technically an on-target site because it has the same 20-nt protospacer with 1-nt silent mismatch in the PAM (CGG vs TGG). For the remaining 11 true off-target sites, eight sites carry one mismatch mutation in the 20-nt protospacer. For the additional three sites with two or three mismatch mutations, only one mutation is present in the 1–18-nt sequence from the PAM (Fig. 4c). Further analysis of these 12 off-target sites found four have silent mutations in NGG PAM and one has a non-canonical CAG PAM, which was reported as an alternative PAM (NAG) for SpCas9 nuclease [33] and recently shown to mediate Cas9 activity in rice [34]. All mutations at these 12 sites were indels, and, importantly, the two Cas9-J T0 lines carried distinct alleles at these sites (Fig. 4d and Additional file 3: Figure S11); validating these mutations were truly caused by Cas9.

Cas9-E sgRNA was predicted by CRISPOR and Cas-OFFinder to contain six off-target sites when up to a 3-nt mismatch was allowed (Fig. 4a and Additional file 2: Table S3). However, no off-target mutations were found at these predicted sites. Although the two Cas9-E T0 lines shared seven SNVs and three indels (Fig. 4b), these ten shared mutations had very poor sequence homology to
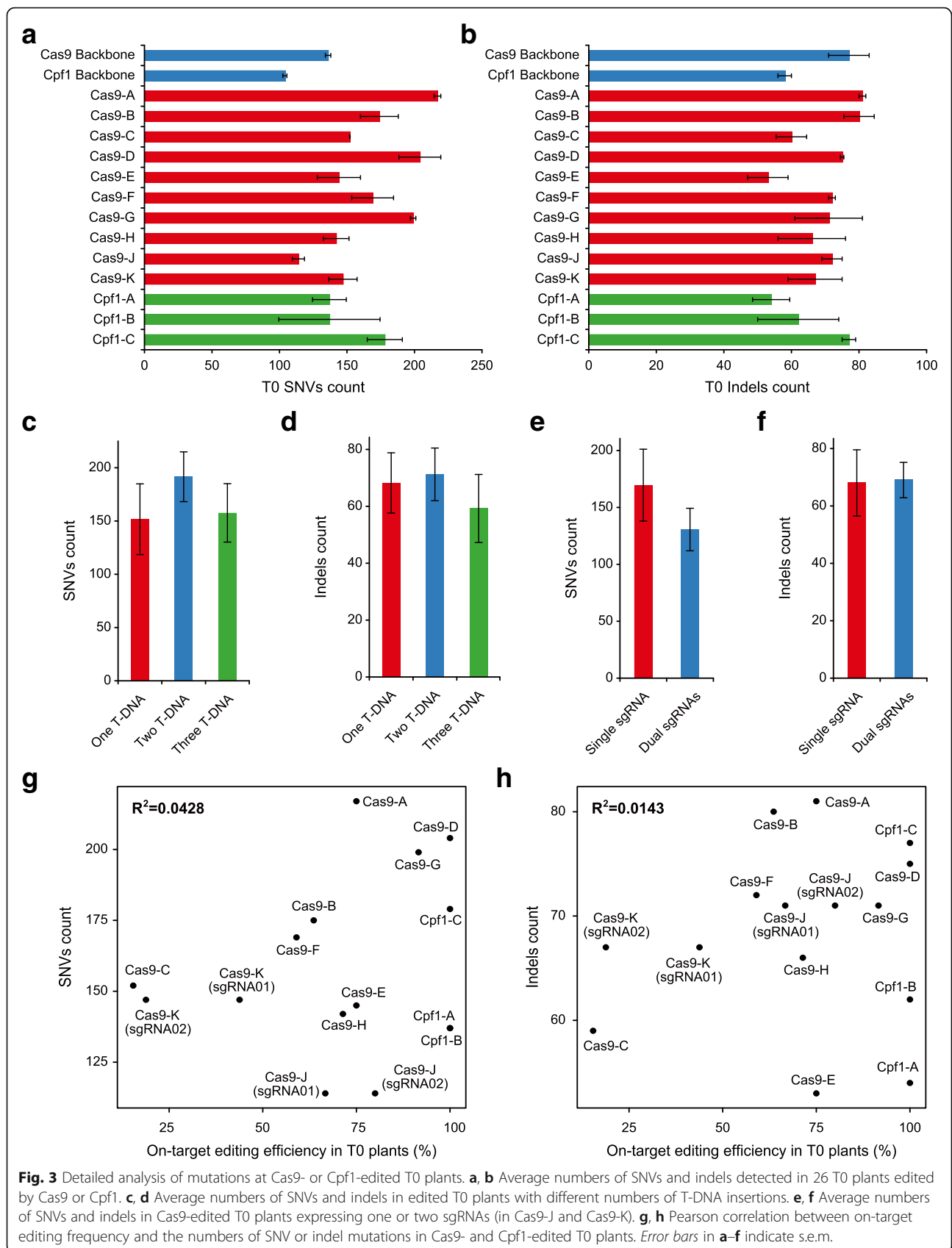
Tang *et al. Genome Biology* (2018) 19:84

Page 6 of 13

Fig. 3 Detailed analysis of mutations at Cas9- or Cpf1-edited T0 plants. **a**, **b** Average numbers of SNVs and indels detected in 26 T0 plants edited by Cas9 or Cpf1. **c**, **d** Average numbers of SNVs and indels in edited T0 plants with different numbers of T-DNA insertions. **e**, **f** Average numbers of SNVs and indels in Cas9-edited T0 plants expressing one or two sgRNAs (in Cas9-J and Cas9-K). **g**, **h** Pearson correlation between on-target editing frequency and the numbers of SNV or indel mutations in Cas9- and Cpf1-edited T0 plants. *Error bars* in **a**–**f** indicate s.e.m.
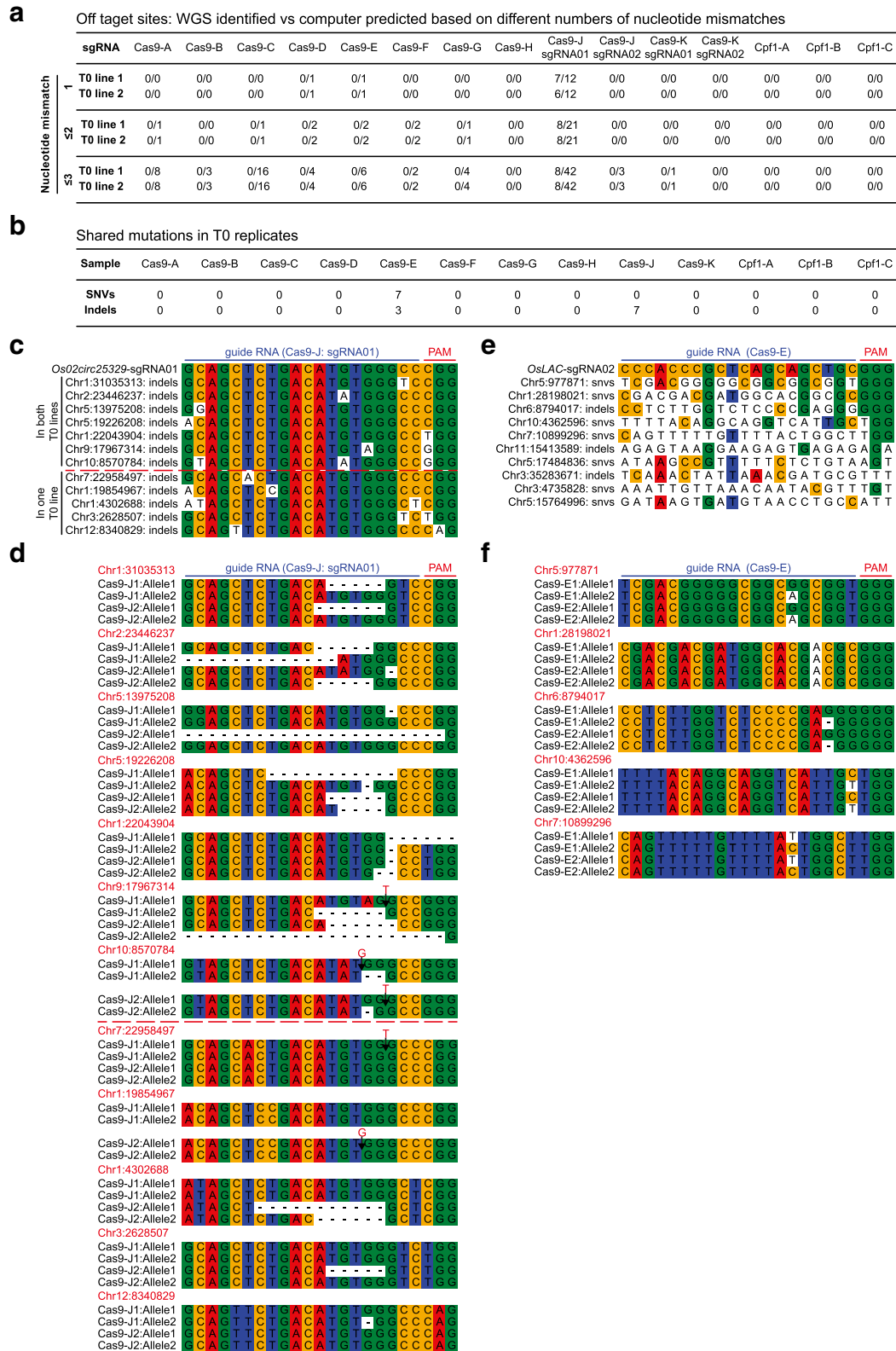
Tang *et al. Genome Biology* (2018) 19:84

Page 7 of 13

a Off taget sites: WGS identified vs computer predicted based on different numbers of nucleotide mismatches

		sgRNA	Cas9-A	Cas9-B	Cas9-C	Cas9-D	Cas9-E	Cas9-F	Cas9-G	Cas9-H	Cas9-J sgRNA01	Cas9-J sgRNA02	Cas9-K sgRNA01	Cas9-K sgRNA02	Cpf1-A	Cpf1-B	Cpf1-C
Nucleotide mismatch	1	T0 line 1	0/0	0/0	0/0	0/1	0/1	0/0	0/0	0/0	7/12	0/0	0/0	0/0	0/0	0/0	0/0
		T0 line 2	0/0	0/0	0/0	0/1	0/1	0/0	0/0	0/0	6/12	0/0	0/0	0/0	0/0	0/0	0/0
	≤2	T0 line 1	0/1	0/0	0/1	0/2	0/2	0/2	0/1	0/0	8/21	0/0	0/0	0/0	0/0	0/0	0/0
		T0 line 2	0/1	0/0	0/1	0/2	0/2	0/2	0/1	0/0	8/21	0/0	0/0	0/0	0/0	0/0	0/0
	≤3	T0 line 1	0/8	0/3	0/16	0/4	0/6	0/2	0/4	0/0	8/42	0/3	0/1	0/0	0/0	0/0	0/0
		T0 line 2	0/8	0/3	0/16	0/4	0/6	0/2	0/4	0/0	8/42	0/3	0/1	0/0	0/0	0/0	0/0

b Shared mutations in T0 replicates

Sample	Cas9-A	Cas9-B	Cas9-C	Cas9-D	Cas9-E	Cas9-F	Cas9-G	Cas9-H	Cas9-J	Cas9-K	Cpf1-A	Cpf1-B	Cpf1-C
SNVs	0	0	0	0	7	0	0	0	0	0	0	0	0
Indels	0	0	0	0	3	0	0	0	7	0	0	0	0

Fig. 4 (See legend on next page.)

Tang et al. Genome Biology (2018) 19:84

Page 8 of 13

(See figure on previous page.)

Fig. 4 Analysis and identification of potential off-target sites in T0 lines. **a** Number of off-target sites identified in replicated T0 plants vs the number of all off-target sites that are predicted by CRISPOR and Cas-OFFinder with allowing up to 3-nt mismatch for all 15 Cas9 or Cpf1 target sites. **b** Identification of shared SNVs and indels between replicated T0 plants. **c** Potential off-target sites identified in both Cas9-J T0 samples (above the *red dashed line*) and only in one T0 sample (below the *red dashed line*). **d** Off-target mutations identified by WGS at off-target sites in both Cas9-J T0 samples (above the *red dashed line*) and only in one T0 sample (below the *red dashed line*). **e** Potential off-target sites identified in Cas9-E samples based on shared mutations in two T0 plants. **f** Sequence analysis of the shared mutations in Cas9-E samples

the target site (Fig. 4e). Only five sites contained the NGG PAM. Among them, the site sharing highest sequence homology with the target site still contained a 10-nt mismatch, making it unlikely to be a true off-target site. Unlike indels found in Cas9-J samples, these putative off-target mutations are mostly SNVs (Fig. 4b, f). Furthermore, both independent T0 lines always carried the same mutant alleles (Fig. 4f and Additional file 3: Figure S11). These observations suggest that the ten shared mutations of two Cas9-E T0 lines were not caused by Cas9, but were pre-existing mutations from a parental line.

Cas9 was previously shown to induce off-target mutations at sites with missing or extra nucleotides when compared to the target site, which forms bulges when targeted by guide RNAs [35]. To detect such off-target mutations, we extracted all T0 mutation site flanking sequences (25 bp upstream and downstream) and aligned them to corresponding sgRNA/crRNA sequences using BLAST. Only Cas9-J1 and Cas9-J2 samples had alignments to the Cas9-J-sgRNA01 target (15 in Cas9-J1 and ten in Cas9-J2); other samples had no hit. None of the detected mutations were caused by bulge-forming DNA–sgRNA recognition. We also investigated whether DNA translocation events were induced by Cas9 or Cpf1 by searching for structural variants (SVs) and gene fusion events in the whole rice genome. We did not detect any translocation event in all T0 lines. Given the level of nuclease-induced DNA translocation can be used for assessing targeting specificity [36], absence of detectable translocation events in all T0 samples here indicates these Cas9 and Cpf1 reagents are indeed very specific, limiting cleavage activity almost exclusively to the target sites.
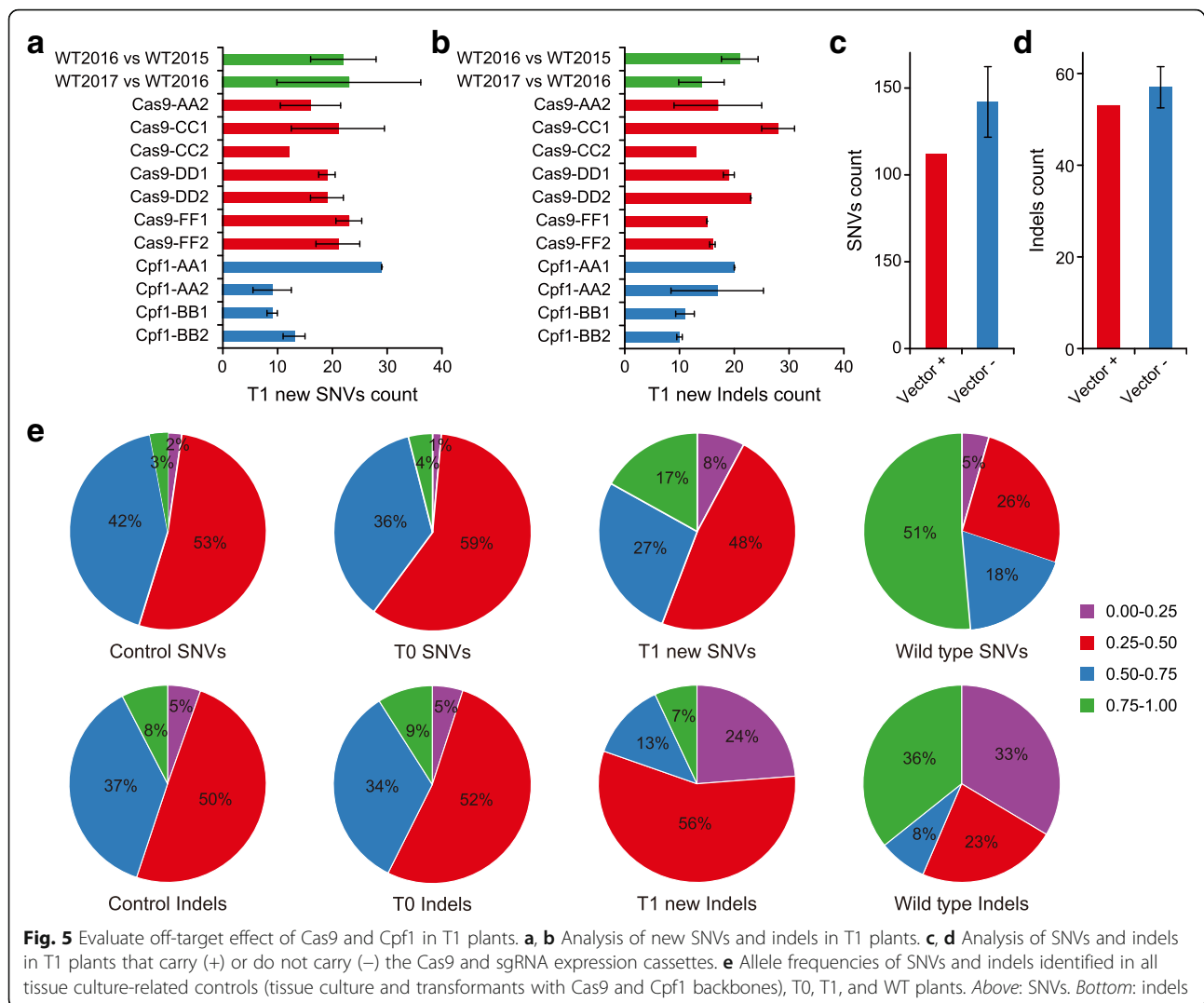
No evidence of off-target mutations in T1 plants

Our analysis of T0 plants suggested 11 out of 12 Cas9 sgRNAs and all three Cpf1 crRNAs are very specific as no off-target mutations were detected. However, lack of off-target mutations might be attributed to low expression or activity of Cas9 or Cpf1. It is also important to determine whether continued expression of the RGNs into the next generation will result in de novo off-target mutations. Therefore, we decided to sequence 14 T1 plants from Cas9 T0 lines with diverse levels of on-target editing efficiency (15, 60, 75, and 100%) at four

target sites and nine T1 plants from Cpf1 T0 lines at two target sites (Fig. 1a, b and Additional file 1: Figure S1). Germline-transmitted on-target mutations in 14 Cas9-edited or nine Cpf1-edited T1 lines were validated by Sanger sequencing (Additional file 3: Figures S12 and S13). With WGS analysis, we identified all SNVs and indels in Cas9 T1 lines (Additional file 3: Figure S14) and Cpf1 T1 lines (Additional file 3: Figure S15). The WGS results confirmed the germline-transmitted on-target mutations (Additional file 3: Figures S12 and S13 and Additional file 2: Table S1). Among all other SNVs and indels, most of them were identified in the corresponding T0 lines, suggesting they have been fixed (Additional file 3: Figure S16). For the other new mutations identified in T1 lines, the average number of SNVs ranged from 9 to 29 (Fig. 5a), while the average of indels ranged from 10 to 28 (Fig. 5b). Such spontaneous mutation rates are consistent with the spontaneous mutation rates we found earlier in WT samples (Fig. 2a, b), which are also in line with a previous study [28].

These new mutations were mapped to the rice genome together with new mutations that were discovered in WT plants across two generations (Additional file 3: Figure S17). The genome distribution of these new mutations in T1 lines also showed enrichment in repeats (Additional file 3: Figure S16), consistent with the spontaneous mutations discovered in the WT (Fig. 2c, d). Detailed analysis of SNVs among all sample types revealed T1 lines have higher rates of G:C > A:T transitions than T0 lines (Additional file 3: Figure S18), consistent with the observation on spontaneous mutations in *Arabidopsis* [27]. Further analysis of T1 lines either with or without the Cas9 transgene did not reveal any difference in the numbers of new SNVs and indels among these two subpopulations (Fig. 5c, d). By applying similar methods from the analysis of T0 plants, we were unable to identify any off-target mutations by Cas9 or Cpf1 in T1 lines. Given most T1 lines analyzed still carry the RGN constructs, our results suggest continued expression of Cas9 or Cpf1 constructs did not cause de novo off-target mutations in T1 lines.

To further assess the new mutations found in T1 lines, we calculated and compared the allele frequency of SNVs and indels among four groups: tissue culture controls, T0 plants, T1 plants, and WT (Fig. 5e). The tissue

Tang *et al. Genome Biology* (2018) 19:84

Page 9 of 13

Fig. 5 Evaluate off-target effect of Cas9 and Cpf1 in T1 plants. **a**, **b** Analysis of new SNVs and indels in T1 plants. **c**, **d** Analysis of SNVs and indels in T1 plants that carry (+) or do not carry (−) the Cas9 and sgRNA expression cassettes. **e** Allele frequencies of SNVs and indels identified in all tissue culture-related controls (tissue culture and transformants with Cas9 and Cpf1 backbones), T0, T1, and WT plants. *Above*: SNVs. *Bottom*: indels

culture controls and Cas9/Cpf1 T0 lines share strikingly similar (mostly heterozygous-like) allele frequency distribution. This reiterates our earlier conclusion that all mutations in T0 samples (except a few found in Cas9-J samples) are background mutations. By contrast, T1 plants show more homozygous-like SNVs (0.75 to 1.0 in allele frequency) and somatic-like indels (0 to 0.25 in allele frequency). This trend of rapidly fixing SNVs and the increase of somatic indels in T1 is interesting and relatively in line with the observation in WT plants.

Discussion

More attention has been given to the specificity of CRISPR-Cas RGN systems in humans than in animals or plants due to medicinal applications of RGNs. Earlier WGS studies in human cells found low incidence of off-target mutations by Cas9 [37, 38]. Recently, two WGS off-target studies in mice showed conflicting results [24, 25]. However, the study that claimed unexpected large-scale off-target

effects by Cas9 may be flawed due to limitations in its experimental design and WGS data analysis [39]. Given the wide adoption of CRISPR-Cas systems in agriculture, with genome-edited crop products reaching market in record time [40], it becomes urgent to conduct large-scale and exhaustive WGS analysis of off-target effects by Cas9 and Cpf1 (Cas12a), two leading RGN systems, in agriculturally important crops. Such studies will help assess the safety of Cas9 and Cpf1 in precise crop breeding as well as provide valuable information to scientists, breeders, regulators, and consumers.

In this study, we conducted a large-scale WGS analysis for detecting potential off-target mutations caused by 12 Cas9 sgRNAs and three Cpf1 crRNAs in rice, an important food crop. We confirmed WGS-identified mutations by Sanger sequencing at randomly selected sites with a 100% success rate (Additional file 2: Table S4), which is consistent with the high quality of our WGS data. Our experimental design took into account background

Tang *et al. Genome Biology* (2018) 19:84

Page 10 of 13

mutations caused by tissue culture and *Agrobacterium*-mediated transformation, pre-existing mutations in parents, and spontaneous mutations that arise from seed propagation. Through sequencing 20 control plants of different types and 49 Cas9 or Cpf1-edited T0 and T1 plants, we only found true off-target mutations in two T0 lines expressing Cas9 protein with Cas9-J-sgRNA01. Importantly, these empirically validated off-target sites can be readily predicted computationally. Our examination of T1 plants that continue to carry Cas9-sgRNA or Cpf1-crRNA did not reveal off-target mutations, suggesting continued presence of the RGN reagents with varying activity in plants does not cause off-target mutations if the guide RNAs are well-designed for specificity. This observation is also highly significant because it encourages the use of Cas9 and Cpf1 in certain breeding applications that may require expression of RGNs across several generations. For example, a RGN cassette may be introduced from a transgenic line into a transformation-recalcitrant variety of the same plant species for genome editing with simple genetic crossing.

Our study also provided insights into avoiding off-target effects of Cas9 and Cpf1 in edited crops. To minimize off-target effects, many systems have been developed, including paired Cas9 nickases [41], high fidelity Cas9 proteins [42–44], FokI-dCas9 fusions [45, 46], truncated sgRNAs [47], and ribonucleotide protein (RNP) delivery [48]. To assess and identify off-target sites, in vivo [18, 36] and in vitro [49–51] tools have also been developed in human cells, which may be applied in plants. Our WGS analysis with WT SpCas9 and LbCpf1 proteins did not find off-target mutations for 14 out of 15 guide RNAs tested in T0 and T1 plants, suggesting utilization of a high-fidelity enzyme, which are typically of lower activity, may be unnecessary in crop applications. When a mismatch up to 3 nt of the protospacer is allowed, Cas9-OFFinder programs predicted a total of 37 off-target sites for 7 out of 11 Cas9 sgRNAs. Yet, we could not detect any mutations at these putative off-target sites. Alternatively, Cas9-OFFinder predicted all the off-target sites that we identified for Cas9-J-sgRNA01; many of the sites have just 1-nt mismatch to the protospacer of the target site. Therefore, we can deduce a simple rule to alleviate off-target effects: make sure even the highest scored potential off-target sites will have at least a 2-nt mismatch to the seed sequence of the protospacer. We note this may not always be possible if the target sequence shares many homologous sequences in the genome. For example, maize has a very repetitive genome and wheat has A, B, D sub-genomes that share high similarity. In these cases, targeted amplicon sequencing using next-generation sequencing technologies may be an appropriate and cost-effective method to look for off-target mutations [52].

Finally, we hope our data can be a valuable reference for regulatory agencies and other entities. It is reasonable and necessary to scrutinize any new technology for its efficacy and safety. Cas9 and Cpf1, as new crop-breeding technologies, are no exception. Although Cas9-based off-target effects have been studied by WGS in plants [21–23], our study differs from previous studies significantly with regard to scale, depth, and comprehensiveness. Our research also represents the first report of using WGS to assess off-targeting by Cpf1 in any edited higher eukaryotic organism. We could not find any off-target mutations in 47 out of 49 rice plants edited by 11 Cas9-sgRNA and three Cpf1-crRNA constructs. This precise level of genome modification is in stunning contrast to many conventional breeding technologies. For example, we found that even the safest breeding approach, harvesting seeds from parental lines, introduces ~ 30 to 50 spontaneous mutations into the next generation in rice. We also observed ~ 200 tissue culture-introduced somaclonal variations per rice plant, even though few are affecting coding sequences. In conclusion, our data support a recent call to "Regulate genome-edited products, not genome editing itself" [53].

Methods

Plant material and growth conditions
This study used the rice variety Nipponbare (*Oryza sativa* L. ssp. Japonica cv. Nipponbare). All plants were grown in growth chambers under controlled environmental conditions with a 16/8 h light/dark regime at 28 °C and 60% relative humidity.

Vector construction
Plasmids encoding Cas9 and a single sgRNA were generated by ligating annealed oligos with a 4-bp overhang into a *Bsa*I digested backbone (either pZHY988 or pTX172) [54–56]. Plasmids with two sgRNAs were created by ligating pZHY988 with a 485-bp fragment, after digestion with *Bsa*I. This 485-bp fragment contains two sgRNAs generated by overlap extension PCR [55]. All CRISPR-Cpf1 (CRISPR-Cas12a) nuclease expression vectors were reported in our previous study [14]. The sequences of all primers used to construct vectors are shown in Additional file 2: Table S5.

Rice stable transformation
Agrobacterium-mediated rice transformation was performed as described in published protocols [57] with slight modification [16, 56]. The binary vectors were introduced into *Agrobacterium tumefaciens* strain EHA105 by the freeze-thaw method [58]. For rice transformation, dehusked seeds were sterilized with 70% ethanol for 1 min. Afterwards, seeds were washed five times with sterile water, then further sterilized for 15 min with a 2.5% sodium

Tang et al. Genome Biology (2018) 19:84

Page 11 of 13

hypochlorite solution containing a drop of Tween 20. The washing and sterilization step were repeated, this time without addition of Tween. Seeds were then rinsed an additional five times before being dried on sterilized filter paper and cultured on solid medium at 28 ℃ in a dark growth chamber for 2–3 weeks. Actively growing calli were collected for subculture at 28 ℃ in the dark for 1–2 weeks. *Agrobacterium* cultures were collected and resuspended in liquid medium (OD600 = 0.06–0.1) containing 100 μM acetosyringone. Rice calli were immersed in the *Agrobacterium* suspension for 30 min, then dried on sterilized filter paper and co-cultured for 3 days on solid medium at 25 ℃ in a dark growth chamber. The infected calli were moved to a sterile plastic bottle and washed five times with sterile water to remove excessive *Agrobacterium*. After being dried on sterilized filter paper, these calli were transferred onto screening medium at 28 ℃ in a dark growth chamber for 5 weeks. During the screening stage, infected calli were transferred to fresh screening medium every 2 weeks. After the screening stage, actively growing calli were moved onto regenerative medium for regeneration at 28 ℃ with a 16 h light/8 h dark cycle. After 3–4 weeks, transgenic seedlings were transferred to sterile plastic containers containing fresh solid medium and grown for 2–3 weeks before being transferred into soil. Transgenic rice plants were grown in a growth chamber at 28 ℃ with a 16 h light/8 h dark cycle.

Mutagenesis analysis at target sites

Genomic DNA was extracted from transgenic plants using the CTAB method [59]. The genomic region flanking the CRISPR target site for each gene was amplified and sequenced. Samples with heterozygous and biallelic mutations were decoded using CRISP-ID [60].

WGS and data analysis

For each sample, about 1 g of fresh leaves were collected from seedlings between 5 and 6 weeks old. DNA samples were extracted using the Plant Genome DNA Kit (Tiangen) as described by the manufacturer. All 69 samples were sequenced by Bionova (Beijing, China) using the Illumina X10 platform. Adapters were trimmed using SKEWER (v. 0.2.2) [61] and the Illumina TruSeq adapter. Cleaned reads were mapped to rice reference sequence TIGR7 (http://rice.plantbiology.msu.edu/) [62] with BWA (v. 0.7.15) software [63]. The Genome Analysis Toolkit (GATK) [64] was used to realign reads near indels and recalibrate base quality scores by following GATK best practices [65]. A known SNPs and indels database for GATK best practices was downloaded from Rice SNP-Seek Database (http://snp-seek.irri.org/) [66]. Whole genome SNVs were detected by LoFreq [67], MuTect2 [68], and VarScan2 [69]. Whole genome indels were identified using MuTect2 [68], VarScan2, and Pindel [70]. Bedtools [71] and BCFtools [72] were used to process overlapping SNVs/indels. Off-target sites were predicted with CRISPOR [31] online and Cas-OFFinder software [32] by allowing up to 10-nt mismatch. A genome-wide map of mutations was plotted with Circos software [73]. Structural variants and translocation events were analyzed using TopHat2 [74] with the '—fusion-search' parameter, DELLY [75] with default parameters, and manually checking with IGV software [76]. NCBI BLAST+ with the '-task blastn-short' parameter for off-target mutation site analysis, which includes mismatch, deletion, and insertion. Read-mapping screenshots were from Golden Helix GenomeBrowse ® visualization tool v2.1. Data processing and analyses were completed using R and Python. One T1 sample of Cas9-CC2 was excluded from analyses due to contamination of fungal DNA.

Additional files

> **Additional file 1: Figures S1–S9.** Supplemental Figures- part I. (PPTX 12111 kb)
>
> **Additional file 2: Tables S1–S5.** Supplemental Tables. (XLSX 43 kb)
>
> **Additional file 3: Figures S10–S18.** Supplemental Figures-part II. (PPTX 13267 kb)

Availability of data and materials
The WGS raw data reported in this article are listed in Additional file 2: Table S2 and have been deposited in the Sequence Read Archive in National Center for Biotechnology Information (NCBI) under the accession number PRJNA420933 (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA420933) [77] and the Genome Sequence Archive in Beijing Institute of Genomics (BIG) under the accession number PRJCA000656 (http://bigd.big.ac.cn/bioproject/browse/PRJCA000656) [78].

Authors' contributions
YZ proposed the project. YQ, TZ, and YZ designed the experiments. XT, JZ, DZ, and AM generated CRISPR-Cas nuclease vectors. XT, JZ, LT, XX, DZ, and XZ generated stable transgenic rice and identified the rice mutants. XT, JZ, QR, ZZ, BL, and XZ prepared samples for WGS and confirmed WGS results with Sanger sequencing. GL, QY, ZG, and TZ performed WGS data analysis. YQ, TZ, and YZ analyzed the data and wrote the manuscript. All authors participated in discussion and revision of the manuscript. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Tang *et al. Genome Biology* (2018) 19:84

Page 12 of 13

Author details

[1]Department of Biotechnology, School of Life Sciences and Technology, Center for Informational Biology, University of Electronic Science and Technology of China, Room 216, Main Building, No. 4, Section 2, North Jianshe Road, Chengdu 610054, People's Republic of China. [2]Jiangsu Key Laboratory of Crop Genetics and Physiology, Co-Innovation Center for Modern Production Technology of Grain Crops, Key Laboratory of Plant Functional Genomics of the Ministry of Education, Yangzhou University, Yangzhou 225009, China. [3]Joint International Research Laboratory of Agriculture and Agri-Product Safety of Ministry of Education of China, Yangzhou University, Yangzhou 225009, China. [4]Department of Plant Science and Landscape Architecture, University of Maryland, College Park, MD 20742, USA. [5]Institute for Bioscience and Biotechnology Research, University of Maryland, Rockville, MD 20850, USA.

References

1. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012;337:816–21.
2. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol. 2013;31:233–9.
3. Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol. 2013;31:227–9.
4. Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J. RNA-programmed genome editing in human cells. elife. 2013;2:e00471.
5. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-guided human genome engineering via Cas9. Science. 2013;339:823–6.
6. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013;339:819–23.
7. Cho SW, Kim S, Kim JM, Kim JS. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol. 2013;31:230–2.
8. Li JF, Norville JE, Aach J, McCormack M, Zhang D, Bush J, et al. Multiplex and homologous recombination-mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. Nat Biotechnol. 2013;31:688–91.
9. Shan Q, Wang Y, Li J, Zhang Y, Chen K, Liang Z, et al. Targeted genome modification of crop plants using a CRISPR-Cas system. Nat Biotechnol. 2013;31:686–8.
10. Nekrasov V, Staskawicz B, Weigel D, Jones JD, Kamoun S. Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. Nat Biotechnol. 2013;31:691–3.
11. Lowder LG, Zhang D, Baltes NJ, Paul JW, Tang X, Zheng X, et al. A CRISPR/Cas9 toolbox for multiplexed plant genome editing and transcriptional regulation. Plant Physiol. 2015;169:971–85.
12. Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Almendros C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology. 2009;155:733–40.
13. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell. 2015;163:759–71.
14. Tang X, Lowder LG, Zhang T, Malzahn AA, Zheng X, Voytas DF, et al. A CRISPR-Cpf1 system for efficient genome editing and transcriptional repression in plants. Nat Plants. 2017;3:17018.
15. Endo A, Masafumi M, Kaya H, Toki S. Efficient targeted mutagenesis of rice and tobacco genomes using Cpf1 from *Francisella novicida*. Sci Rep. 2016;6:38169.
16. Zhong Z, Zhang Y, You Q, Tang X, Ren Q, Liu S, et al. Plant genome editing using FnCpf1 and LbCpf1 nucleases at redefined and altered PAM sites. Mol Plant. 2018. https://doi.org/10.1016/j.molp.2018.03.008.
17. Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol. 2013;31:822–6.
18. Tsai SQ, Zheng Z, Nguyen NT, Liebers M, Topkar VV, Thapar V, et al. Guide-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat Biotechnol. 2015;33:187–97.
19. Kim D, Kim J, Hur JK, Been KW, Yoon SH, Kim JS. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. Nat Biotechnol. 2016;34:863–8.
20. Kleinstiver BP, Tsai SQ, Prew MS, Nguyen NT, Welch MM, Lopez JM, et al. Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. Nat Biotechnol. 2016;34:869–74.
21. Feng Z, Mao Y, Xu N, Zhang B, Wei P, Yang DL, et al. Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in *Arabidopsis*. Proc Natl Acad Sci U S A. 2014; 111:4632–7.
22. Zhang H, Zhang J, Wei P, Zhang B, Gou F, Feng Z, et al. The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. Plant Biotechnol J. 2014;12:797–807.
23. Nekrasov V, Wang C, Win J, Lanz C, Weigel D, Kamoun S. Rapid generation of a transgene-free powdery mildew resistant tomato by genome deletion. Sci Rep. 2017;7:482.
24. Iyer V, Shen B, Zhang W, Hodgkins A, Keane T, Huang X, et al. Off-target mutations are rare in Cas9-modified mice. Nat Methods. 2015;12:479.
25. Schaefer KA, Wu WH, Colgan DF, Tsang SH, Bassuk AG, Mahajan VB. Unexpected mutations after CRISPR-Cas9 editing *in vivo*. Nat Methods. 2017; 14:547–8.
26. Yang N, Xu XW, Wang RR, Peng WL, Cai L, Song JM, et al. Contributions of *Zea mays* subspecies mexicana haplotypes to modern maize. Nat Commun. 2017;8:1874.
27. Ossowski S, Schneeberger K, Lucas-Lledo JI, Warthmann N, Clark RM, Shaw RG, et al. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. Science. 2010;327:92–4.
28. Yang S, Wang L, Huang J, Zhang X, Yuan Y, Chen JQ, et al. Parent-progeny sequencing indicates higher mutation rates in heterozygotes. Nature. 2015; 523:463–7.
29. Evans DA. Somaclonal variation–genetic basis and breeding applications. Trends Genet. 1989;5:46–50.
30. Wei FJ, Kuang LY, Oung HM, Cheng SY, Wu HP, Huang LT, et al. Somaclonal variation does not preclude the use of rice transformants for genetic screening. Plant J. 2016;85:648–59.
31. Haeussler M, Schonig K, Eckert H, Eschstruth A, Mianne J, Renaud JB, et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biol. 2016;17:148.
32. Bae S, Park J, Kim JS. Cas-offinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics. 2014;30:1473–5.
33. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol. 2013;31:827–32.
34. Meng X, Hu X, Liu Q, Song X, Gao C, Li J, et al. Robust genome editing of CRISPR-Cas9 at NAG PAMs in rice. Sci China Life Sci. 2018;61:122–25.
35. Lin Y, Cradick TJ, Brown MT, Deshmukh H, Ranjan P, Sarode N, et al. CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. Nucleic Acids Res. 2014;42:7473–85.
36. Frock RL, Hu J, Meyers RM, Ho YJ, Kii E, Alt FW. Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. Nat Biotechnol. 2015;33:179–86.
37. Veres A, Gosis BS, Ding Q, Collins R, Ragavendran A, Brand H, et al. Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. Cell Stem Cell. 2014;15:27–30.
38. Smith C, Gore A, Yan W, Abalde-Atristain L, Li Z, He C, et al. Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. Cell Stem Cell. 2014;15:12–3.
39. Editorial. CRISPR off-targets: a reassessment. Nat Methods. 2018;15:229–30.
40. Waltz E. CRISPR-edited crops free to enter market, skip regulation. Nat Biotechnol. 2016;34:582.
41. Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE, et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell. 2013;154:1380–9.
42. Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z, et al. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. Nature. 2016;529:490–5.
43. Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F. Rationally engineered Cas9 nucleases with improved specificity. Science. 2016;351: 84–8.
44. Chen JS, Dagdas YS, Kleinstiver BP, Welch MM, Sousa AA, Harrington LB, et al. Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. Nature. 2017;550:407–10.

Tang *et al. Genome Biology* (2018) 19:84

Page 13 of 13

45. Tsai SQ, Wyvekens N, Khayter C, Foden JA, Thapar V, Reyon D, et al. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. Nat Biotechnol. 2014;32:569–76.

46. Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nat Biotechnol. 2014;32:577–82.

47. Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat Biotechnol. 2014;32:279–84.

48. Kim S, Kim D, Cho SW, Kim J, Kim JS. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome Res. 2014;24:1012–9.

49. Cameron P, Fuller CK, Donohoue PD, Jones BN, Thompson MS, Carter MM, et al. Mapping the genomic landscape of CRISPR-Cas9 cleavage. Nat Methods. 2017;14:600–6.

50. Kim D, Bae S, Park J, Kim E, Kim S, Yu HR, et al. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. Nat Methods. 2015;12:237–43.

51. Tsai SQ, Nguyen NT, Malagon-Lopez J, Topkar VV, Aryee MJ, Joung JK. Circle-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. Nat Methods. 2017;14:607–14.

52. You Q, Zhong Z, Ren Q, Hassan F, Zhang Y, Zhang T. CRISPRMatch: an automatic calculation and visualization tool for high-throughput CRISPR genome-editing data analysis. Int J Biol Sci. 14:858–62.

53. Carroll D, Van Eenennaam AL, Taylor JF, Seger J, Voytas DF. Regulate genome-edited products, not genome editing itself. Nat Biotechnol. 2016;34:477–9.

54. Tang X, Zheng X, Qi Y, Zhang D, Cheng Y, Tang A, et al. A single transcript CRISPR-Cas9 system for efficient genome editing in plants. Mol Plant. 2016;9:1088–91.

55. Zhou J, Deng K, Cheng Y, Zhong Z, Tian L, Tang X, et al. CRISPR-Cas9 based genome editing reveals new insights into microRNA function and regulation in rice. Front Plant Sci. 2017;8:1598.

56. Zheng X, Yang S, Zhang D, Zhong Z, Tang X, Deng K, et al. Effective screen of CRISPR/Cas9-induced mutants in rice by single-strand conformation polymorphism. Plant Cell Rep. 2016;35:1545–54.

57. Hiei Y, Ohta S, Komari T, Kumashiro T. Efficient transformation of rice (*Oryza sativa* l.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. Plant J. 1994;6:271–82.

58. Weigel D, Glazebrook J. Transformation of *Agrobacterium* using the freeze-thaw method. CSH Protoc. 2006. https://doi.org/10.1101/pdb.prot4666.

59. Stewart CN Jr, Via LE. A rapid ctab DNA isolation technique useful for rapd fingerprinting and other PCR applications. BioTechniques. 1993;14:748–50.

60. Dehairs J, Talebi A, Cherifi Y, Swinnen JV. CRISP-ID: decoding CRISPR mediated indels by Sanger sequencing. Sci Rep. 2016;6:28973.

61. Jiang H, Lei R, Ding SW, Zhu S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics. 2014;15:182.

62. Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, et al. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. Rice. 2013;6:4.

63. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. In: ArXiv e-prints. 2013;arXiv:1303.3997v2.

64. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303.

65. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;43:11.10.1–33.

66. Mansueto L, Fuentes RR, Borja FN, Detras J, Abriol-Santos JM, Chebotarov D, et al. Rice SNP-seek database update: new SNPs, indels, and queries. Nucleic Acids Res. 2017;45:D1075–D81.

67. Wilm A, Aw PP, Bertrand D, Yeo GH, Ong SH, Wong CH, et al. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Res. 2012;40:11189–201.

68. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013;31:213–9.

69. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012;22:568–76.

70. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics. 2009;25:2865–71.

71. Quinlan AR. BEDTools: the swiss-army tool for genome feature analysis. Curr Protoc Bioinformatics. 2014;47:1121–34.

72. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27:2987–93.

73. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19:1639–45.

74. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14:R36.

75. Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics. 2012;28:i333–i9.

76. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14:178–92.

77. Tang X, Liu G, Zhou J, Ren Q, You Q, Tian L, Xin X, Zhong Z, Liu B, Zheng X, Zhang D, Malzahn A, Gong Z, Qi Y, Zhang T, Zhang Y. A large-scale whole-genome sequencing analysis reveals highly specific genome editing by both Cas9 and Cpf1 (Cas12a) nucleases in rice. https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA420933.

78. Tang X, Liu G, Zhou J, Ren Q, You Q, Tian L, Xin X, Zhong Z, Liu B, Zheng X, Zhang D, Malzahn A, Gong Z, Qi Y, Zhang T, Zhang Y. A large-scale whole-genome sequencing analysis reveals highly specific genome editing by both Cas9 and Cpf1 (Cas12a) nucleases in rice http://bigd.big.ac.cn/bioproject/browse/PRJCA000656.