Safety evaluation of myostatin-edited Meishan pigs by whole genome resequencing analyses

SHANSHAN XIE, LILI QIAN, CHUNBO CAI, SHENGWANG JIANG, GAOJUN XIAO, TING GAO, XIANG LI, WENTAO CUI*

Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, Beijing, P.R. China
*Corresponding author: cuiwentao@caas.cn


Abstract: Genome editing technology can make specifically target genomic modifications, resulting in site specific DNA insertion, deletion or replacement in the genome of an organism. We have recently produced genetically engineered (GE) Meishan pigs containing a ZFN-edited myostatin (MSTN) loss-of-function mutation that leads to a clear “double muscle” phenotype as observed for Belgian cattle. In this study, whole genome resequencing was used as an approach to evaluate the safety risk, if any, associated with the introduction of a ZFN-edited myostatin (MSTN) loss-of-function mutation in a local pig breed, the Meishan pigs. The results of resequencing analyses show that the effective data from pigs of wild-type group and MSTN-edited GE group is greater than 99%. The 1× coverage rate is > 98%, and the 4× coverage rate is > 96%. The genetic variation on each chromosome is close to 1. From this whole genome resequencing study, our results demonstrated that 99.7% of single nucleotide polymorphisms (SNPs) are the same in the same genetic variation from both wild-type group and MSTN-edited GE group, implying genomic sequence variations are highly similar between the two groups of pigs.

Keywords: MSTN; gene editing; genome resequencing; genetic variation; comparative analysis

Gene editing is a technique that enables precise site-directed mutagenesis in the genome of an organism. Accurate gene editing at the genomic level can alter specific traits or introduce specific traits in plants or animals that are of importance to agricultural or medical research. Genome editing technology has been widely used in animal breeding, plant cultivation, and other fields (Carroll 2011; Luo et al. 2014). Since no marker genes are used, gene editing methods are now widely used to generate genetically engineered plants and animals in the agricultural industry.

Whole genome sequencing is a fast and efficient technique for studying human genetic diseases and screening for high-quality traits in plants and animals. It can also be used to assess species evolution and determine phylogenetic relationships among species (Choi et al. 2015). Fan et al. (2013) reported that a large number of single nucleotide polymorphisms (SNPs) and InDel in a local chicken breed in Taiwan were found in the genes that are involved in metabolic regulation and growth. There were no off-target effects being observed in hornless cattle generated by TALEN-edited PCPOLLED following whole genome sequencing analyses (Carlson et al. 2016) and only one off-target mutation was detected from a total 119 off-target sites in rats where genome editing was performed for multiple genes by the CRISPR/Cas9 system. Myostatin (MSTN) is a transcriptional growth factor, also known as growth/differentiation fac-
tor 8 (GDF8). The main function of *MSTN* is its negative regulation of skeletal muscle growth and development. Muscle hypertrophy is observed in *MSTN* knockout mice, with a significant increase in the number of skeletal fibres and in the size of myofibres. It has been reported that natural mutations in MSTN gene led to a significant increase in skeletal muscle mass in cattle, sheep, dogs, and humans (Grobet et al. 1997; Schuelke et al. 2004; Mosher et al. 2007). Our lab has recently generated MSTN loss-of-function mutant Meishan pigs by using zinc finger nuclease (ZFN) technology (Qian et al. 2015). These ZFN-edited MSTN–/– (KO) Meishan pigs show the apparent double muscle phenotype as reported in the Belgian cattle containing naturally occurring loss-of-function MSTN mutations, with the total fat content being significantly lower than in wild type (WT) pigs. In this study, we performed the whole genome depth (20×) resequencing analysis to measure genomic variations for wild-type and MSTN–/– (KO) pigs with a goal of providing evidence supporting the safety of these MSTN–/– (KO) pigs at a genomic level.

**MATERIAL AND METHODS**

**Materials collection.** All Meishan pigs are kept at Qingdao Animal Husbandry Experimental Station and were fed the same standard diet and raised under the same conditions. There were 4 pigs in each group. One sample was collected from one randomly selected pig in each group for whole genome re-sequencing.

Ear tissues were collected from one WT pig and one MSTN-edited Meishan pig at 6 months of age (with an average body weight of about 50 kg for each pig) and immediately placed in liquid nitrogen and then transferred to −80°C freezer for long-term storage (Qian et al. 2015).

**Ethics statement.** All experimental protocols related to animal work described in this study were reviewed and approved by the Institutional Animal Care and Use Committee (IACUC) at the Institute of Animal Sciences, Chinese Academy of Agricultural Sciences. All experiments were performed in accordance with the approved guidelines for animal care and management of research projects.

**Library construction and sequencing.** DNA was extracted from the ear tissue in EDTA using a Wizard Genomic DNA kit (Promega, USA), and DNA concentration was measured by Qubit dsDNA HS Assay (Invitrogen, USA). 1.5 µg of each DNA sample was used to build the DNA library. DNA samples were randomly broken into 350 bp fragments by using a Covaris S2 system (Covaris, USA). After DNA terminal repair, addition of Poly-A tail and sequencing linkers/adaptors, and PCR amplification, a TruSeq Library Construction Kit (Illumina, USA) was used to construct libraries. The constructed library was sequenced by Illumina HiSeq 2000. After the library is constructed, Qubit2.0 is used for initial quantification, dilution of the library to 1 ng/µl, followed by detection of the insert size of the library using the Agilent 2100 Bioanalyzer (Agilent Technologies, USA), insert size in line with the expected concentration of the library using q-PCR Accurate quantification (effective concentration of the library > 2nM), to ensure the quality of the library. Library inspection qualified, according to the effective concentration of the library and data output requirements for Illumina HiSeq 2000 sequencing. To ensure the quality of the information analysis, it is necessary to filter the raw reads and get clean reads. The obtained valid sequencing data were compared by reference to the reference genome by BWA software (Li and Durbin 2009) (parameters: mem -t 4 -k 32 -M), and the results were removed by SAMtools (Li et al. 2009) (parameter: rmdup).

**SNP detection and annotation.** We used SAMtools (Wang et al. 2010) (-q 1-C -m 2 -F 0.002-d 1000) for individual SNP detection. ANNOVAR (Wang et al. 2010) is an efficient software tool that utilizes the latest information to annotate functional variations detected in multiple genomes. As long as the genetic variations are identified in specific chromosomes, start sites, termination sites, and reference nucleotides and mutant nucleotides are known, ANNOVAR can be used for gene-based annotation, region-based annotations, filter-based annotation, and other functionalities.

**InDel detection and annotation.** We used SAMtools (-q 1-C -m 2 -F 0.002-d 1000) to detect the insertion and deletion of small fragments shorter than 50 bp (InDel).
RESULTS

Whole genome resequencing. In this study, there were 4 pigs in each group, but only one sample was collected from one randomly selected pig in each group for whole genome resequencing. The ration for using one sample from one pig is that the depth of the genome resequencing is 20×, which is really deep enough for data accuracy and sufficiency for subsequent analysis. Under normal conditions, 10× depth of the genome resequencing can detect SNP and InDel, with the data Q30 being close to 89%, and the data being sufficient for subsequent analysis. For example, Stothard et al. (2011) performed whole genome resequencing for a Black Angus bull and a Holstein bull at 22-fold and 19-fold coverage, respectively.

The raw data from the two samples collected from one WT and one MSTN-edited Meishan pig is between 55.39 G and 70.32 G (55.39~70.32 billion bp). For WT Meishan pigs, the raw data average is 55.39 G, and the clean data average is 55.18 G, which represents a data efficiency of 99.62%. For MSTN-edited Meishan pigs, the raw data average is 64.18 G, and the clean data average is 60.03 G, which represents a data efficiency of 99.78%. The Q20 value is greater than 95%, and the Q30 value is greater than 89%. The total GC content is greater than 42%.

The mapping rate for both samples is 92%. The 1× coverage for samples in both WT and MSTN-edited (KO) groups is greater than 98%, while the depth is 18.31× and 21.27× for MSTN-edited (KO) Meishan pigs and wild-type Meishan pigs, respectively. The data obtained from our current deep sequencing is highly reliable and thus can be used for subsequent data analysis. The reference genome size is 2.81 × 10^9 bp, and the ratio of the two samples is 92.06–92.54%. And the 1× coverage rate is greater than 98.52% and the 4× coverage rate is greater than 96% (Table 1). The average coverage depth of the reference genome (excluding the N region) is 18.31–22.94×. All results are normal and can be used for subsequent analysis.

Table 1. Comparative sequencing results from wild type (WT) (MSTN<sup>+/+</sup>) pig and KO (MSTN<sup>−/−</sup>) pig

<table>
<thead>
<tr>
<th>Sample</th>
<th>Mapped reads</th>
<th>Total reads</th>
<th>Mapping rate (%)</th>
<th>Average depth (×)</th>
<th>Coverage at least 1× (%)</th>
<th>Coverage at least 4× (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KO</td>
<td>395 058 468</td>
<td>426 898 700</td>
<td>92.54</td>
<td>21.27</td>
<td>98.52</td>
<td>96.83</td>
</tr>
<tr>
<td>WT</td>
<td>338 670 264</td>
<td>367 880 670</td>
<td>92.06</td>
<td>18.31</td>
<td>98.56</td>
<td>96.49</td>
</tr>
</tbody>
</table>

By comparing with the reference genome, we found that the genome coverage, sequencing depth and total GC content are basically the same between the MSTN-edited Meishan pigs and the WT Meishan pigs, indicating that the trend of the total genomic variations is the same.

SNP detection and annotation. From the SNP support reads number and neighbouring SNP distance distribution (Figure 1), it is believed that the SNP in each sample is highly reliable.

We integrated the information on genetic variations detected in samples from both WT and MSTN-edited (KO) groups and compared this information at the same locus. The results indicate that 7.66 × 10^6 SNPs are found in the same location in both MSTN-edited (KO) and WT groups. Among these SNPs, 99.7% (7.64 × 10^6) of SNPs contain the same nucleotide changes, and the genetic variations in both WT and MSTN-edited (KO) groups are highly consistent, indicating that there was hardly any genetic diversity between MSTN-edited (KO) and WT groups. Additionally, we also found that 22.7% (1.74 × 10^6) of SNPs are located in the protein coding region, accounting for a relatively small proportion, and 75.9% (5.8 × 10^6) of SNPs are located between genes, with similar distribution in the genome.

We further analysed the SNPs in both MSTN-edited Meishan and WT pigs and noticed that there was a large number of SNPs on chromosomes 1, 2, 6, 8, 9 and 13 (Figure 2A), and the SNP distribution density in both MSTN-edited and WT pigs is very similar (Figure 3A, B). There were approximately 6.28 × 10^4 and 6.65 × 10^4 SNPs in the exon regions of MSTN-edited Meishan pigs and WT Meishan pigs, respectively. There are 35% (2.18 × 10^4) and 35% (2.30 × 10^4) of nonsynonymous mutations in the exon regions of MSTN-edited Meishan and WT pigs, respectively. There are 169 and 189 mutations that lead to termination codons in MSTN-edited (KO) Meishan and WT pigs, respectively. There are 30 and 35 mutations that lead to a loss of the stop codons in MSTN-edited Meishan and WT pigs, respectively. There are 64.8% (4.07 × 10^4)
and 65% (4.32 × 10^4) of synonymous mutations in MSTN-edited Meishan (KO) and WT pigs, respectively. The intron regions contain 2.29 × 10^6 and 2.39 × 10^6 SNPs in MSTN-edited Meishan and WT pigs, respectively. The regions between genes contain 7.82 × 10^6 and 7.61 × 10^6 SNPs in MSTN-edited Meishan (KO) and WT pigs, respectively. The genetic variations in exon regions are 0.61% and 0.62% for MSTN-edited (KO) and WT groups, respectively. Of these genetic variations, 0.21% belong to non-sense variations in both MSTN-edited (KO) and WT groups. These genetic variations are very small relative to the reference genome, and the variations are highly consistent in both MSTN-edited (KO) and WT groups.

**InDel identification and analysis.** The average total number of InDels is 1.80 × 10^6 for MSTN-edited (KO) Meishan pigs and 1.82 × 10^6 for WT Meishan pigs. By further analysing the InDels in MSTN-edited (KO) Meishan pigs and WT Meishan pigs, we found that there are a lot of InDels (Figure 2B) on chromosomes 1, 2, 4, 6, 8, 9, 13, and 15, with the InDel distribution density on each chromosome.
Figure 3. Density of single nucleotide polymorphisms (SNP) and InDel on each chromosome: (A) SNP density in wild type (WT) \((MSTN^+/+)\) pig and (B) SNP density in KO \((MSTN^{-/-})\) pig. (C) InDel density in WT \((MSTN^+/+)\) pig. (D) InDel density in KO \((MSTN^{-/-})\) pig. Axis \(x\): position Mb, axis \(y\): chromosome.
being similar (Figure 3C, D). There were approximately $1.95 \times 10^3$ and $1.94 \times 10^3$ InDels in the exon regions in MSTN-edited (KO) Meishan pigs and WT Meishan pigs, respectively. In exon regions, there are 16 and 17 mutations that result in the generation of the stop codons in MSTN-edited (KO) Meishan pigs and WT Meishan pigs, respectively. There are 4 and 3 mutations that result in a loss of the stop codons in MSTN-edited (KO) Meishan pigs and WT Meishan pigs, respectively. There are on average $1.32 \times 10^2$ and $1.33 \times 10^2$ insertions or deletions that result in translational frameshift in MSTN-edited (KO) Meishan pig and WT Meishan pig, respectively. There are 605 and 592 insertions or deletions that result in the insertion or deletion of 3 or integer multiples of 3 base pairs but do not change the translational frameshift. In the intron regions, there are on average $4.04 \times 10^5$ and $4.06 \times 10^5$ InDels in MSTN-edited (KO) Meishan pigs and WT Meishan pigs, respectively. In the regions between genes, there are $1.36 \times 10^6$ and $1.38 \times 10^6$ InDels in MSTN-edited (KO) Meishan pigs and WT Meishan pigs, respectively. The genomic heterozygosity was 0.16‰ and 0.22‰ for MSTN-edited (KO) Meishan pigs and WT Meishan pigs, respectively. The number of total genes containing frameshift mutations is $1.60 \times 10^3$ for both MSTN-edited (KO) Meishan pigs and WT Meishan pigs, with the ratio being 1.

We calculated the length distribution of InDels in the coding regions in the whole genome (Figure 4) for both MSTN-edited (KO) Meishan pigs and WT Meishan pigs. Gene coding regions contain important genetic information and thus they are very sensitive to mutations. The length of InDels has a very large effect on gene functions. We selected 1854 unique InDels from the coding regions of MSTN-edited (KO) Meishan pigs and performed length distribution analysis. The major length is 1 bp, and about 31% (567) of InDels contain a base number that equals the multiple of 3. To explore these InDel variants further, we also conducted a functional enrichment analysis. The results showed that those genes affected by InDels are mainly involved in basic cellular functions. These data indicate that the genetic variations are the same between MSTN-edited (KO) and WT groups.

**DISCUSSION**

One of the key goals in porcine breeding industry is to generate high-quality pork with high lean yield and low fat mass (Kanis et al. 2005). Meishan pigs are a locally famous breed in China, and are well known for their high prolificacy, early sexual maturity, and delicious meat (Legault 1985). However, the Meishan breed has a high percentage of carcass fat, slower growth, and poor feed efficiency. Therefore, there is a need to make genetic
Modified to improve pork quality produced by Meishan pigs. Our lab has recently generated ZFN-edited MSTN Meishan pigs containing the loss-of-function mutation. These MSTN^{−/−} pigs (KO) produce pork with higher percentage of lean yield and lower percentage of fat (Qian et al. 2015). At the same time, because the organ structure of pigs is similar to humans, the genome of pigs and the human genome are highly conserved and homologous, and pigs become an important evolutionary and disease research model (Rehfeldt et al. 2000; Fang et al. 2005; Gorodkin et al. 2007; Lunney 2007; Walters et al. 2012). Therefore, we can use MSTN-edited (KO) Meishan pigs as a model to gain more insight into how to improve pork quality and how to use pigs as an animal model to study human diseases.

Kerstens et al. (2009) found that there are nearly 100,000 SNPs in the porcine genome after comparative analysis of the swine genomic database. The number of SNPs we obtained in this study is consistent with the number of SNPs reported by Kerstens et al. (2009). The transition/transversion (ts/tv) ratio has been used as an indicator of potential sequencing errors. Recent studies have demonstrated that the ts/tv ratio is about 2.1 and 2.2 for humans and bovines, respectively. Previous studies (Altshuler et al. 2012; Choi et al. 2013, 2014) show that the ts/tv ratio in most pigs is approximately the same as that in humans and cattle. Our analysis showed that the ts/tv ratio is 2.36 for both MSTN-edited (KO) Meishan pigs and pure WT Meishan pigs, indicating that most of the SNPs in this study were reasonable and accurate. Statistical analysis indicates that the total number of SNPs/InDels in MSTN^{−/−} (KO) Meishan pigs and WT Meishan pigs is $1.03 \times 10^7/1.80 \times 10^6$ and $1.08 \times 10^7/1.82 \times 10^6$, respectively. Further analysis of these SNPs and InDels on each chromosome demonstrated that there is no difference in the distribution of SNPs and InDels in each chromosome between MSTN-edited (KO) Meishan pigs and WT Meishan pigs, and the variation ratio is close to 1. All these results indicate that both MSTN-edited Meishan pigs and WT Meishan pigs have the same trend of genetic variation with the whole genome being extremely similar to each other.

Special attention has been paid to the safety risk of GE organisms generated by genomic editing technology. Although there are not yet many GE animals being approved to produce food for human consumption (Laible et al. 2015), GE plants produced by genomic editing technology have entered a commercial stage. In 2012, the US government approved the field evaluation of the first phosphorus-efficient maize produced by genome-targeted modification (Jones 2015). Calyxt (Minneapolis, USA) has recently developed two kinds of soybean varieties by the use of TALEN gene editing technology to produce healthier oils containing the same level of monounsaturated fatty acid as found in olive oil and rapeseed oil. These soybeans have been planted in the United States (Ledford 2016). We observed no other significant differences except the lean yield and fat mass between MSTN^{−/−} (KO) Meishan pigs and WT pigs. For example, Qian et al. (2015) tested the off-target situation of ZFN. The primers were designed according to the predicted sequence information of the 11 sites most likely to be off-targeted by ZFN, and the DNA of the MSTN gene-edited (KO) cloned pigs and part of the offspring was used as a template for PCR amplification, and the PCR product was sequenced, and the target site was obtained. The normal sequence was subjected to an alignment analysis to detect whether or not ZFN was off-target. The results of the assay showed that there were no insertions or deletions of the nucleic acid sequence at the predicted 11 off-target sites, and it was thus determined that the ZFN did not have an off-target effect at these sites. Blood tests showed that there were no significant differences in other parameters between MSTN^{−/−} (KO) and WT pigs (Qian et al. 2015). Recently, we evaluated the safety of pork produced by MSTN-edited (KO) Meishan pigs (Xiao et al. 2016). The results from a 90-day sub-chronic toxicity study showed that the food consumption of MSTN-edited (KO) Meishan pork did not have any long-term adverse effect on the health in rats (Xiao et al. 2016). Thus, these MSTN^{−/−} (KO) pigs may have an obvious advantage and potential to produce improved quality pork for human consumption.

Our present study evaluated the safety of these MSTN^{−/−} (KO) Meishan pigs by comprehensive genome resequencing analyses and did not observe any obvious safety risk associated with MSTN editing by ZFN technology.

CONCLUSION

In this study, we conducted genomic resequencing of the wild type and MSTN^{−/−} (KO) Meishan pigs.
The analysis of resequencing data showed that genomes between the WT and MSTN KO Meishan pigs are highly similar to each other, with 99.7% of SNPs being identical. Additionally, InDels analysis indicated that the genomic variations between the WT and MSTN KO Meishan pigs are also similar to each other. Thus, it is concluded that there is no significant safety risk associated with the introduction of a ZFN-edited myostatin (MSTN) loss-of-function mutation in Meishan pigs.

Acknowledgement. We thank Dr. Jin-an Jiao for editing this manuscript; Mr. Ke Jiang and Mr. Hegang Li for care and management of pig farm; Drs. Shulin Yang, Yanfang Wang, Zhonglin Tang, Yulian Mu, and Hong Ao for their guidance and suggestions.

REFERENCES

sequencing of black Angus and Holstein cattle for SNP and CNV discovery. BMC Genomics, 12, 559.


Received: 2018–01–10
Accepted: 2019–01–02