





Article

Insights from Initial Variant Detection by Sequencing Single Sperm in Cattle

Liu Yang ^{1,2,3,†}, Yahui Gao ^{1,4,†}, Clarissa Boschiero ¹ , Li Li ² , Hongping Zhang ² , Li Ma ^{4,*} 
and George E. Liu ^{1,*}

- ¹ Animal Genomics and Improvement Laboratory, Henry A. Wallace Beltsville Agricultural Research Center, Agricultural Research Service, USDA, Beltsville, MD 20705, USA; yangqism@gmail.com (L.Y.); gyhalvin@gmail.com (Y.G.); Clarissa.Boschiero@usda.gov (C.B.)
- ² College of Animal Science and Technology, Sichuan Agricultural University, Chengdu 611130, China; lily@sicau.edu.cn (L.L.); zhp@sicau.edu.cn (H.Z.)
- ³ Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China
- ⁴ Department of Animal and Avian Sciences, University of Maryland, College Park, MD 20742, USA
- * Correspondence: lima@umd.edu (L.M.); George.Liu@usda.gov (G.E.L.); Tel.: +1-301-405-1389 (L.M.); +1-301-504-9843 (G.E.L.); Fax: +1-301-504-8414 (G.E.L.)
- † Equal contribution.

Abstract: Meiotic de novo mutation (DNM) is one of the important phenomena contributing to gamete genome diversity. However, except for humans and a few model organisms, they are not well studied in livestock, including cattle. Moreover, bulk sperm samples have been routinely utilized in experiments, which include millions of single sperm cells and only report high-frequency variants. In this study, we isolated and sequenced 143 single sperms from two Holstein bulls and identified hundreds of candidate DNM events in ten sperms with deep sequencing coverage. We estimated DNM rates ranging from 1.08×10^{-8} to 3.78×10^{-8} per nucleotide per generation. We further validated 12 out of 14 selected DNM events using Sanger sequencing. To our knowledge, this is the first single sperm whole-genome sequencing effort in livestock, which provided useful information for future studies of point mutations and male fertility. Our preliminary results pointed out future research directions and highlighted the importance of uniform whole genome amplification, deep sequence coverage, and dedicated software pipelines for genetic variant detection using single-cell sequencing data.

Keywords: cattle; single sperm sequencing; de novo mutation; variant detection



Citation: Yang, L.; Gao, Y.; Boschiero, C.; Li, L.; Zhang, H.; Ma, L.; Liu, G.E. Insights from Initial Variant Detection by Sequencing Single Sperm in Cattle. *Dairy* **2021**, *2*, 649–657. <https://doi.org/10.3390/dairy2040050>

Received: 17 August 2021

Accepted: 18 October 2021

Published: 15 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Background

Recent breakthroughs in the development and application of single-cell sequencing technologies provide an avenue for dissecting population lineages and heterogeneity and understanding cell identity, differentiation, and function [1–5]. Single-cell DNA-seq (scDNA-seq) technologies produce data, which allow the detection of single nucleotide variants (SNVs) and short insertion and deletion (INDEL) variants, together with structural variations or abnormal chromosome numbers (aneuploidy) on the single-cell level [6–8].

Although small, sperm is one of the most important cells because it delivers the entire paternal genetic materials to the offspring. As novel mutations can occur during gametogenesis and postzygotically, studying mutations in sperm is particularly important for male fertility. Traditionally, bulk sperm sequencing can only detect high-frequency variants in the presence of millions of individual sperm cells. Recent genome and exome sequencing studies of parent–offspring trios have provided the first insights into the number and distribution of the de novo mutations (DNMs) [9]. In humans, DNMs have been shown to be a major cause of severe early-onset genetic disorders such as intellectual

disability and autism spectrum disorder [10]. Recent studies have also shown that DNMs are predominantly of paternal origin and that their number increases with advanced paternal age [11]. By setting up a single sperm sequencing approach, Wang et al. reported 25 to 36 DNMs per sperm in humans [1].

As in other mammals, reproductive performance in cattle is also affected by paternal fertility. Although the fertility of artificial insemination (AI) bulls is monitored routinely by major organizations using microscopic examination of sperm count, motility, abnormality, and other laboratory tests, an understanding of the mutation mechanisms involved in sperm production and their impacts on male fertility is equally important and essential for the dairy industry and other livestock industries. In fact, the occurrence of novel mutations in each generation explains why these reproductively lethal disorders continue to occur in mammalian species [9].

In this study, we manually isolated, amplified, sequenced, and analyzed 143 single sperm genomes from two Holstein bulls. We identified hundreds of candidate DNM events in ten sperms with deep sequencing coverage by comparing them to the somatic genome. After validating selected events using Sanger sequencing, we estimated rates for DNMs. To the best of our knowledge, this is the first large-scale single sperm whole-genome sequencing report in livestock, which could facilitate future studies of point mutations and male fertility.

2. Methods

2.1. Sample Collection and Whole Genome Amplification and Sequencing

We randomly chose two Holstein bulls: Sample1 has a DPR (daughter pregnancy rate) PTA value of 0.0, reliability of 0.99, estimated from 6528 daughters. In contrast, Sample2 has a DPR PTA value of −3.2, reliability of 0.99, estimated from 15,314 daughters. Somatic tissue (ear punch) samples of Holstein Sample1, together with its parent somatic tissues, were donated by Select Sires, Inc. (Plain City, OH, USA). Semen samples were freshly collected by Select Sires, Inc. in its routine artificial insemination semen straw production. After receiving them under liquid nitrogen in USDA-ARS Animal Genomics and Improvement Laboratory (AGIL), we manually isolated a total of 156 sperm cells from two Holstein bulls (Sample1 with 73 sperm cells and Sample2 with 83 sperm cells). Briefly, isolated sperms were thawed in 37 °C water for 30–45 s and treated with 0.25% Trypsin-EDTA, followed by dilution with PBS + 1% BSA and washing twice. The sperms were further diluted to a proper resolution using PBS + 1% BSA on a Petri dish, and active single sperms were picked up manually by pipetting into a reaction tube under a micromanipulator as described previously [12]. Whole-genome amplification was performed on single cells according to the manufacturer's protocol, using the Single-Cell Whole-Genome Amplification Kit (Yikon Genomics, Shanghai, China) developed from the Multiple Annealing and Looping-Based Amplification Cycles (MALBAC) method [6]. In brief, a single sperm was initially analyzed and pre-amplified by primers supplied in the kit with 8 cycles with multiple annealing steps. PCR generated fragments with variable lengths at random starting positions for next-generation sequencing. We also sequenced the somatic diploid genomes of the trio, including Sample1 (Sample1-diploid) and its parents (Sample1-sire and Sample1-dam). Using their somatic ear punch tissues, we isolated their diploid genomes using a QIAGEN DNA extraction kit. DNA samples extracted from the donor and his parents' ear skin samples were then used to prepare sequencing libraries using standard Illumina protocol and sequenced on Illumina HiSeq 2000/NextSeq 500 sequencing platforms.

2.2. Genotype Calling

Paired-end sequencing reads for single sperm and diploid samples were quality controlled by FastQC v0.11.9. (Available online: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed on 3 June 2020)) and trimmed by Trimmomatic v0.39 [13] (Available online: <http://www.usadellab.org/cms/?page=trimmomatic> (accessed on 3 June 2020)). BWA v0.7.17 [14] mem (Available online: <http://bio-bwa.sourceforge.net/> (accessed on

3 June 2020)) was used with default parameters to align clean reads against the latest bovine reference genome ARS-UCD1.2 (Available online: ftp://ftp.ensembl.org/pub/release-99/fasta/bos_taurus/dna/Bos_taurus.ARS-UCD1.2.dna.toplevel.fa.gz (accessed on 3 June 2020)). To avoid potential PCR or sequencing optical artifacts, we marked duplicated reads that were mapped to the same location by MarkDuplicates function in GATK v4.0.8.1 [15]. FixMateInformation was also employed to ensure all mate-pair information is in sync between each read and its mate-pair. For detecting systematic errors made by the sequencing machine, Base Quality Score Recalibration (BQSR) was called for each BAM by BaseRecalibrator and ApplyBQSR with the known SNP file from 1000 bull projects (Available online: <http://www.1000bullgenomes.com/> (accessed on 26 June 2020)). HaplotypeCaller in GATK was used to call variants, and the parameter -ERC GVCF in CombineGVCFs was set for data combining and then performed by GenotypeGVCFs. We separated SNPs and INDELs in a combined VCF file by using the function SelectVariants.

2.3. Filtration of SNPs, INDELs, and Samples

To improve the genotyping accuracy for single sperms, we applied a stringent cut-off on the raw genotyping quality score to call genotypes. We removed low-quality variants with quality by depth (QD) < 2, Fisher strand (FS) > 30, strand odds ratio (SOR) > 3, root mean square of the mapping quality (MQ) < 40, and quality score (QUAL) < 40. Using the VariantFiltration function in GATK, we defined the window size as 35 bp to evaluate clustered SNPs and allowed three SNPs to make up a cluster. For sperm data, we kept variants with at least two allele support reads and removed heterozygous (0/1) SNPs or INDELs because it was potentially caused by sequencing errors or sperm chromosome-scale genomic anomalies [8]. As a result, 12 sperm samples were removed as their read depth was lower than $0.5 \times (10 \text{ sperms})$ or genome coverage rate was lower than 10% (2 sperms). In addition, for diploid data, we filtered those variants with allele support reads less than $1/2$ genome-wide depth.

2.4. De Novo Mutation Detection

The genotypes called by GATK from ten sperms of Sample1 were used to identify DNMs. To minimize the artifacts or sequencing errors, we required a genotyping quality score (QUAL) ≥ 100 and the number of supporting reads to be more than $3/4$ genome coverage of each sample. We defined a candidate DNM in sperm when a distinct sperm genotype existed from the somatic homozygous genotype. We excluded signals from repetitive regions or with low alignment confidence. The DNM rate was calculated as the alternate allele calling frequency in genome sequence length.

2.5. Gene Annotation

We mapped variations to the bovine reference gene annotation of the ARS-UCD1.2 genome from ENSEMBL using BEDtools v2.26.0 [16] (Available online: <https://bedtools.readthedocs.io/en/latest/> (accessed on 29 June 2020)). The gene features included transcripts, exons, CDS, 3'-UTR, 5'-UTR, start codons, and stop codons.

2.6. Amplification and Sequencing of Cattle Mutations

We designed the PCR and sequencing primers using Primer-BLAST [17] based on the bovine ARS-UCD1.2 genome. All the primers used in the present study are listed in Table S8. The PCR amplification was performed with 25 μ L reaction volume according to Taq DNA polymerase manufacturer's protocol (Taq PCR Master Mix Kit, Qiagen, Hilden, Germany), and the genomic DNA was amplified on a BioRad MyIQ thermocycler. The PCR cycle was as follows: initial denaturation at 95 °C for 5 min; followed by 25 cycles of 94 °C for 60 s, annealing at 56 °C for 60 s; primer extension at 72 °C for 60 s; and final extension at 72 °C for 10 min. All the amplified products were run in 1.5% agarose gel. After purification, DNA was sequenced using PCR primers at GENEWIZ (South Plainfield, NJ, USA).

3. Results

3.1. Sequencing and Genotyping of Haploid Sperms

Sequencing of sperms: We amplified and sequenced a total of 156 single sperm cells manually picked from two Holstein bulls' semen using the MALBAC method [12]. After quality control filtering and mapping with BWA, 143 sperm data (71 for Sample1 and 72 for Sample2) were kept for downstream analyses. The sequenced sperms had an average of $1.79 \times$ genome coverage, and 16 of them were at $\sim 4 \times$ genome coverage, achieving an overall coverage of $\sim 11.40\%$ to $\sim 41.35\%$ of the genome, respectively (Table S1). On average, 98.18% of sequencing reads from single sperms were mapped on the bovine ARS-UCD1.2 genome.

Genotyping of sperms: We used GATK to call the raw genotypes for single nucleotide polymorphisms (SNPs) and INDELs. Each sperm yielded raw calls for 15.5–43.0 million SNPs and 2.4–7.2 million INDELs (Table S2). As sperms are haploid cells, extensive heterozygous genotype calls are considered anomalies and thus removed. We only detected a small fraction of heterozygous raw calls with an average frequency of 2.46% (ranging from 1.03% to 7.39%) for SNPs and 2.97% (1.03% to 9.16%) for INDELs, respectively. These indicated that most of the sperms were isolated successfully with low contamination before sequencing. After QC, 0.42 to 2.68 million for SNPs and 0.23 to 1.04 million for INDELs were kept.

3.2. Sequencing and Genotyping of Sample1 Diploid Somatic Genome

For Sample1 somatic diploid genomes, we sequenced bulk DNA samples extracted from ear punches of Sample1 to approximately $40 \times$ genome coverage, with over 99% genome mapping rate and covering 96% genome sequence (Table S3). After QC filtering, mapping, and genotyping, ~ 5.61 million (62.89%) SNPs and ~ 0.72 million (65.26%) INDELs of Sample1 were obtained. Of them, 44.45% and 46.48% high-quality SNPs and INDELs were heterozygous, respectively (Table S4).

3.3. De Novo Mutations Detected in Single Sperms

We estimated the DNM rates from bulls to sperms based on genotyped SNPs. From ten Sample1 sperms with deep sequencing coverage, we detected 955 candidate DNM sites by comparing them to their somatic genome. After removing 501 sites in repeat/segmental duplication regions and 74 sites in 100 bp clusters (39 overlapped with repeat/segmental duplication region), we found 419 DNM sites. Their statistics and potential effects are summarized in Table 1. On average, each chromosome contained 1.54 (0.5 to 2.5) mutations (Figure 1). The most frequent mutations were A-G, C-T, G-A, and T-C with the count of 7.5, 7, 6.5, and 5.5 on the median, respectively (Figure 2A). Within a chromosome, DNM distribution was generally even (Figure 2B). These 419 DNM sites led to an average mutation rate of 1.68×10^{-8} per nucleotide per generation in each sperm, with a range of 1.08 to 3.78×10^{-8} per nucleotide per generation (Figure 3A, Table 1, Tables S5 and S6). Mutations were also enriched in the AT cluster region (Figure 3B). Interestingly, we detected 15 DNMs in more than one sperm. Three sites at chr7:66596324, chr18:12015775, and chr29:439948 occurred in 7, 5, and 4 sperms. Of those 419 high-confidence DNMs, 190 sites overlapped 188 genes and 432 transcripts, with 15 sites mapped to 15 exons, six sites mapped to ten 3'-UTRs for different transcripts, and eight sites mapped to 16 coding sequence (CDS) regions for different proteins (Table S7).

Table 1. Statistics of de novo mutations in 10 sperms with deep sequencing coverage.

Sperm ID	12	32	33	34	36	43	71	76	88	92
Physical coverage (Gbp)	11.34	12.39	29.30	24.05	20.56	22.00	19.46	18.32	18.64	12.25
Mutation counts	34	39	38	35	27	28	49	57	94	45
Mutation rates ($\times 10^{-8}$)	1.37	1.57	1.53	1.41	1.08	1.12	1.97	2.29	3.78	1.81
Transition/Transversion ratio	2.78	1.60	2.80	1.50	1.08	1.33	2.50	2.00	3.27	3.09
CpG	0.00	0.00	0.03	0.06	0.00	0.00	0.00	0.04	0.04	0.02
Coding-missense	0	0	1	0	0	0	1	2	2	2
UTR	0	1	0	1	0	3	2	0	2	1
Noncoding genes	0	0	0	0	1	0	1	0	1	0
Pseudo genes	0	0	0	0	0	0	0	1	0	1
Protein coding genes	12	12	14	19	11	19	16	26	44	19
Exonic	0	1	1	1	0	1	2	2	3	4
CDS	0	0	1	0	0	0	1	2	2	2
Intronic	12	10	13	17	11	15	12	24	39	14
Intergenic	22	27	24	16	15	9	32	30	49	25

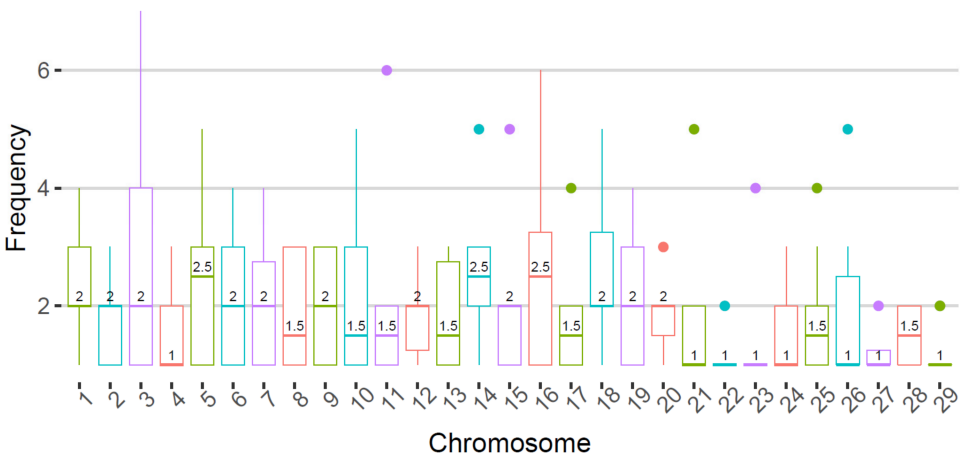


Figure 1. Distribution on chromosomes of de novo mutation sites from Sample1 to sperms. X-axis: chromosomes; Y-axis: frequency boxplots of de novo mutations with the number on each boxplot representing the median.

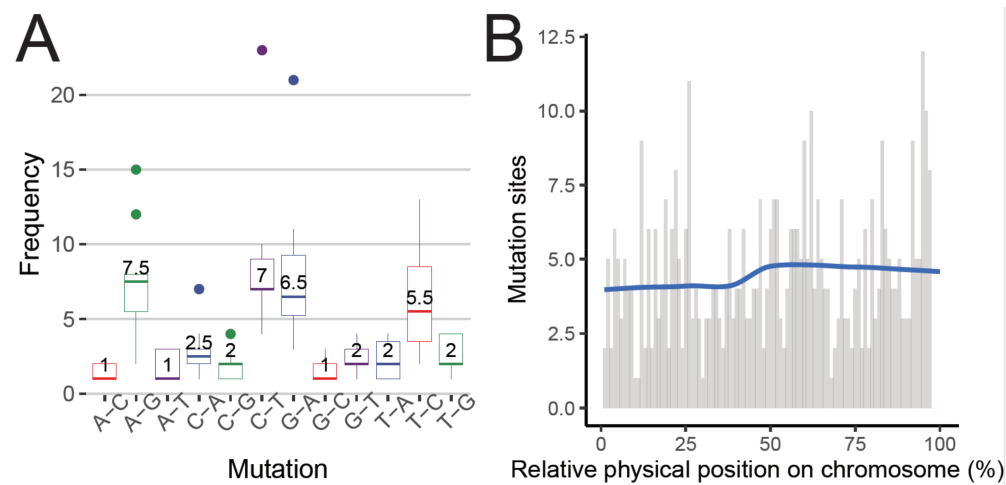


Figure 2. (A) Frequency of de novo mutations from Sample1 to sperms by mutation types. X-axis: mutation types; Y-axis: frequency boxplots of de novo mutations with the number on each boxplot representing the median. (B) Distribution of de novo mutations from Sample1 to sperms on over the relative chromosomal position.

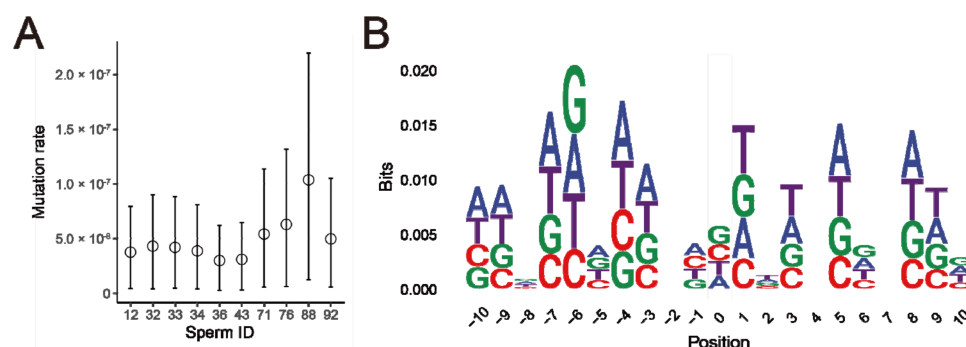


Figure 3. (A) Mutation rates from 10 high-coverage single sperms have overlapping 95% confidence intervals. (B) Base preference of de novo mutation sites within 10 bp flanking sequences.

3.4. PCR Sanger Sequencing Validation of De Novo Mutations

To confirm the candidate DNMs predicted from high-throughput sequencing, we performed PCR-Sanger sequencing. Using 20 primer pairs, we were able to obtain reliable results from 14 regions using PCR-Sanger sequencing (Table S8), along with the six positive control results derived from the sequenced cow (Dt, i.e., Hereford cow named “L1 Dominette 01449”) bulk DNA. The Sanger sequencing results from 12 regions confirmed our recurrent DNM calls ($12/14 = 85.71\%$), therefore largely excluding the possibility of errors during PCR amplification, high-throughput sequencing and read mapping. Because these loci are inconsistent with the Mendelian inheritance rules, we successfully validated that they were DNM by using independent PCR Sanger sequencing.

4. Discussion

Ideally, scDNA-seq could provide information about all variants that occurred in a single cell, such as SNV and INDEL variants, together with structural variations. However, all existing whole-genome amplification (WGA) methods introduce errors and amplification biases and even complete dropouts of variant alleles [2,6,18]. Current single cell-specific SNV callers, like Monovar, SCcaller, and SCAN-SNV, were developed to deal with these errors and missing data [19–21]. They often incorporate amplification error rates and allele dropout in their models, take advantage of enhanced signals from multiple single cells and imputation of missing alleles, and integrate with different data types like bulk sequencing or RNA sequencing. However, a systematic comparison of these tools is still lacking, and most of them were developed for processing scDNA-seq generated from somatic cells in cancer research. We are not aware of any turnkey scDNA-seq variant calling pipeline available for DNM discovery in sperm. Another disadvantage of studying gamete cells like sperms compared to somatic cells is that recombination occurs during meiosis, introducing a major obstacle.

In their pioneer study of single human sperm, Wang et al. plotted the allele discordance ratio of sperm MDA products against the somatic genome. They found that a peak at 100% discordance illustrated a distinct group of loci violating the amplification errors model [1]. After excluding signals from repetitive regions or with low alignment confidence, they obtained 25–36 candidate DNMs in each sperm cell. They further validated 16 out of 18 selected DNMs using independent PCR-Sanger sequencing.

Using a similar approach, we generated 143 single sperm sequencing data and tentatively determined DNM rates ranging from 1.08 to 3.78×10^{-8} per nucleotide per generation in ten cattle sperm genomes with deep coverage, corresponding to 27–94 candidate point mutations per sperm. We also successfully validated 12 out of 14 selected recurrent DNMs using independent PCR-Sanger sequencing. Although there is a possibility that false-positive DNMs could be created during the sperm DNA amplification’s first cycle, it is rare to detect the same mutation to occur at the same position for different sperms. Therefore, our high confirmation rate (85.71%) of the recurrent DNMs argues

against this possibility. For example, a C to A mutation at chr7:66596324 was validated in 5 independent sperms except in WGA33 (Table S7). Agreeing with the abovementioned human result [1], Sample1's mutation rate (1.08 to 3.78×10^{-8} per nucleotide per generation) is higher than that derived from pedigree sequencing data ($\sim 1 \times 10^{-8}$ per nucleotide per generation) [22]. However, it is in line with evolutionary studies, suggesting more mutations in males than in females, possibly due to the larger number of germline cell divisions in males [23]. In our ten individual sperm cells, their respective mutation levels were broadly consistent (Figure 3A). Within each cell, most mutations reside in intergenic or intronic regions (Table 1). However, we did find one to two mutations affecting the coding sequence. The transition-to-transversion ratios of Sample 1 mutations varied from 1.08 to 3.27, compared to a population average of 2.1. The main reason for more transition than transversion is generally thought to be deamination of methylated cytosine, primarily at CpG and potentially in other sequence contexts.

When a DNM arises during the terminal differentiation of sperm, it will rarely be detected to be recurrent. If it arises in proliferating spermatogonial stem cells, it can be detected in multiple sperms [24]. A DNM can also arise early during paternal embryonic development, before primordial germ cell (PGC) specification, or late within the PGC population, causing mosaicism in sperms, and all of them be detected as recurrent events [25]. We detected 15 recurrent mutations in more than one sperm. Those recurrent DNMs might occur during paternal embryonic development, before PGC specification, or within the PGC population, as well as later within spermatogonial stem cells [25].

Limitations and future directions: As discussed above, much progress is needed before turnkey software pipelines can routinely make reliable calls for DNM from single sperm sequencing. As the sequencing coverage was critically low, DNMs reported here are probably less reliable. Furthermore, our sperm number was not large enough to draw general conclusions. In summary, we tried to generate single sperm whole-genome sequencing data and detected occurrences of DNM in cattle. In the meantime, our results also highlighted the importance of uniform whole genome amplification, deep sequence coverage, and dedicated software pipelines for variant detection. To our knowledge, this is the first single sperm sequencing attempt in livestock, which could open the door for studying point mutations and male fertility.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/dairy2040050/s1>, Table S1: Statistics of sequencing data for each sperm. Table S2: Statistics of genotyping data for each sperm. Table S3: Statistics of sequencing data for Sample1's somatic genome. Table S4: Statistics of genotyping data for Sample1's somatic genome. Table S5: De novo mutation sites from Sample1 to sperms. Table S6: De novo mutation rate of ten sperms. Table S7: Gene feature annotation of de novo mutation sites from Sample1 to sperms. Table S8: PCR primers and Sanger sequencing validations of candidate de novo mutations.

Author Contributions: G.E.L. and L.M. conceived the study. L.Y., Y.G. and C.B. analyzed and interpreted data. L.Y., L.M. and G.E.L. wrote the manuscript. L.L. and H.Z. contributed tools and materials. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by AFRI grant numbers 2016-67015-24886, 2019-67015-29321, 2020-67015-31398, and 2021-67015-33409 from the USDA National Institute of Food and Agriculture (NIFA) Animal Genome and Reproduction Programs and BARD grant number US-4997-17 from the US-Israel Binational Agricultural Research and Development (BARD) Fund. This research used resources provided by the SCINet project of the USDA ARS project number 0500-00093-001-00-D. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the results of this research are available within the article and its Supplementary Information Files. All other sequence data can be tracked in

Supplemental Files. The single sperm sequencing data were submitted to GEO under the accession number PRJNA691741.

Acknowledgments: We thank Reuben Anderson, Ki-Eun Park, Adam Oswalt, Bhanu P. Telugu, and Charles G. Sattler for their technical assistances and sample donations. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture. The USDA is an equal opportunity provider and employer.

Conflicts of Interest: All other authors declare that they have no competing interests.

Disclaimer: Ethics approval and consent to participate. The need for ethics approval was waived as the current study did not involve whole animals.

Abbreviations

AGIL	Animal Genomics and Improvement Laboratory
BQSR	base quality score recalibration
CDS	coding sequence
DNM	de novo mutations
DPR	Daughter pregnancy rate
FS	Fisher strand
INDEL	insertion and deletion
Kb	kilobase pairs
MALBAC	multiple annealing and looping based amplification cycles
Mb	megabase pairs
MQ	mapping quality
PCR	polymerase chain reaction
PGC	primordial germ cell
QC	quality control
QD	quality by depth
QUAL	quality score
scDNA-seq	single-cell DNA-seq
SNP	single nucleotide polymorphism
SNV	single nucleotide variation
WGA	whole-genome amplification

References

1. Wang, J.; Fan, H.C.; Behr, B.; Quake, S.R. Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. *Cell* **2012**, *150*, 402–412.
2. Lu, S.; Zong, C.; Fan, W.; Yang, M.; Li, J.; Chapman, A.R.; Zhu, P.; Hu, X.; Xu, L.; Yan, L.; et al. Probing meiotic recombination and aneuploidy of single sperm cells by whole-genome sequencing. *Science* **2012**, *338*, 1627–1630.
3. Shalek, A.K.; Satija, R.; Shuga, J.; Trombetta, J.J.; Gennert, D.; Lu, D.; Chen, P.; Gertner, R.S.; Gaublot, J.T.; Yosef, N.; et al. Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* **2014**, *510*, 363–369.
4. Smith, G.P. Evolution of repeated DNA sequences by unequal crossover. *Science* **1976**, *191*, 528–535.
5. Han, X.; Zhou, Z.; Fei, L.; Sun, H.; Wang, R.; Chen, Y.; Chen, H.; Wang, J.; Tang, H.; Ge, W.; et al. Construction of a human cell landscape at single-cell level. *Nature* **2020**, *581*, 303–309.
6. Zong, C.; Lu, S.; Chapman, A.R.; Xie, X.S. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* **2012**, *338*, 1622–1626.
7. Gawad, C.; Koh, W.; Quake, S.R. Single-cell genome sequencing: Current state of the science. *Nat. Rev. Genet.* **2016**, *17*, 175–188.
8. Bell, A.D.; Mello, C.J.; Nemesh, J.; Brumbaugh, S.A.; Wysoker, A.; McCarroll, S.A. Insights into variation in meiosis from 31,228 human sperm genomes. *Nature* **2020**, *583*, 259–264.
9. O’Roak, B.J.; Deriziotis, P.; Lee, C.; Vives, L.; Schwartz, J.J.; Girirajan, S.; Karakoc, E.; MacKenzie, A.P.; Ng, S.B.; Baker, C.; et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat. Genet.* **2011**, *43*, 585–589.
10. O’Roak, B.J.; Vives, L.; Girirajan, S.; Karakoc, E.; Krumm, N.; Coe, B.P.; Levy, R.; Ko, A.; Lee, C.; Smith, J.D.; et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* **2012**, *485*, 246–250.
11. Kong, A.; Frigge, M.L.; Masson, G.; Besenbacher, S.; Sulem, P.; Magnusson, G.; Gudjonsson, S.A.; Sigurdsson, A.; Jonasdottir, A.; Jonasdottir, A.; et al. Rate of de novo mutations and the importance of father’s age to disease risk. *Nature* **2012**, *488*, 471–475.
12. Zhou, Y.; Shen, B.; Jiang, J.; Padhi, A.; Park, K.E.; Oswalt, A.; Sattler, C.G.; Telugu, B.P.; Chen, H.; Cole, J.B.; et al. Construction of PRDM9 allele-specific recombination maps in cattle using large-scale pedigree analysis and genome-wide single sperm genomics. *DNA Res.* **2018**, *25*, 183–194.

13. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120.
14. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760.
15. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303.
16. Quinlan, A.R.; Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841–842.
17. Ye, J.; Coulouris, G.; Zaretskaya, I.; Cutcutache, I.; Rozen, S.; Madden, T.L. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinform.* **2012**, *13*, 134.
18. Hou, Y.; Fan, W.; Yan, L.; Li, R.; Lian, Y.; Huang, J.; Li, J.; Xu, L.; Tang, F.; Xie, X.S.; et al. Genome analyses of single human oocytes. *Cell* **2013**, *155*, 1492–1506.
19. Zafar, H.; Wang, Y.; Nakhleh, L.; Navin, N.; Chen, K. Monovar: Single-nucleotide variant detection in single cells. *Nat. Methods* **2016**, *13*, 505–507.
20. Dong, X.; Zhang, L.; Milholland, B.; Lee, M.; Maslov, A.Y.; Wang, T.; Vijg, J. Accurate identification of single-nucleotide variants in whole-genome-amplified single cells. *Nat. Methods* **2017**, *14*, 491–493.
21. Luquette, L.J.; Bohrsen, C.L.; Sherman, M.A.; Park, P.J. Identification of somatic mutations in single cell DNA-seq using a spatial model of allelic imbalance. *Nat. Commun.* **2019**, *10*, 3908.
22. Conrad, D.F.; Bird, C.; Blackburne, B.; Lindsay, S.; Mamanova, L.; Lee, C.; Turner, D.J.; Hurles, M.E. Mutation spectrum revealed by breakpoint sequencing of human germline CNVs. *Nat. Genet.* **2010**, *42*, 385–391.
23. Makova, K.D.; Li, W.H. Strong male-driven evolution of DNA sequences in humans and apes. *Nature* **2002**, *416*, 624–626.
24. Acuna-Hidalgo, R.; Veltman, J.A.; Hoischen, A. New insights into the generation and role of de novo mutations in health and disease. *Genome Biol.* **2016**, *17*, 241.
25. Breuss, M.W.; Antaki, D.; George, R.D.; Kleiber, M.; James, K.N.; Ball, L.L.; Hong, O.; Mitra, I.; Yang, X.; Wirth, S.A.; et al. Autism risk in offspring can be assessed through quantification of male sperm mosaicism. *Nat. Med.* **2020**, *26*, 143–150.