

Article

Analysis of the Whole-Genome Sequences from an *Equus* Parent-Offspring Trio Provides Insight into the Genomic Incompatibilities in the Hybrid Mule

Xiujuan Ren, Yuanyi Liu, Yiping Zhao, Bei Li, Dongyi Bai, Gerelchimeg Bou, Xinzhuang Zhang, Ming Du, Xisheng Wang, Tugeqin Bou, Yingchao Shen and Manglai Dugarjaviin *

Equine Research Center, College of Animal Science, Inner Mongolia Agricultural University, 306 Zhaowuda Road, Hohhot 010018, China

* Correspondence: dmanglai@imau.edu.cn



Citation: Ren, X.; Liu, Y.; Zhao, Y.; Li, B.; Bai, D.; Bou, G.; Zhang, X.; Du, M.; Wang, X.; Bou, T.; et al. Analysis of the Whole-Genome Sequences from an *Equus* Parent-Offspring Trio Provides Insight into the Genomic Incompatibilities in the Hybrid Mule. *Genes* **2022**, *13*, 2188. <https://doi.org/10.3390/genes13122188>

Academic Editors: Chunjiang Zhao and Bo Zuo

Received: 30 September 2022

Accepted: 21 November 2022

Published: 23 November 2022

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



Copyright: © 2022 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/>).

Abstract: Interspecific hybridization often shows negative effects on hybrids. However, only a few multicellular species, limited to a handful of plants and animals, have shown partial genetic mechanisms by which hybridization leads to low fitness in hybrids. Here, to explore the outcome of combining the two genomes of a horse and donkey, we analyzed the whole-genome sequences from an *Equus* parent-offspring trio using Illumina platforms. We generated $41.39\times$ and $46.21\times$ coverage sequences for the horse and mule, respectively. For the donkey, a $40.38\times$ coverage sequence was generated and stored in our laboratory. Approximately 24.86 million alleles were discovered that varied from the reference genome. Single nucleotide polymorphisms were used as polymorphic markers for assigning alleles to their parental genomic inheritance. We identified 25,703 Mendelian inheritance error single nucleotide polymorphisms in the mule genome that were not inherited from the parents through Mendelian inheritance. A total of 555 *de novo* single nucleotide polymorphisms were also identified. The rate of *de novo* single nucleotide polymorphisms was 2.21×10^{-7} in the mule from the *Equus* parent-offspring trio. This rate is obviously higher than the natural mutation rate for *Equus*, which is also consistent with the previous hypothesis that interracial crosses may have a high mutation rate. The genes associated with these single nucleotide polymorphisms are mainly involved in immune processes, DNA repair, and cancer processes. The results of the analysis of three genomes from an *Equus* parent-offspring trio improved our knowledge of the consequences of the integration of parental genomes in mules.

Keywords: *equus*; heterogeneous hybridization; genomic incompatibility; Mendelian inheritance error

1. Introduction

The caballine lineage and the stenonine lineage diverged quite recently, approximately 4.0–4.5 million years ago [1]. The earliest occurrence with gene flow was about 2.1–3.4 million years ago [2]. *Equus* speciation events were accomplished through acute chromosomal rearrangements, with the rearrangement rate ranging from 2.9 to 22.2 per million years [3,4]. Two sets of perfectly functional and structurally well-characterized genetic programs have evolved for horses (*Equus caballus*, $2n = 64$) and donkeys (*Equus asinus*, $2n = 62$) [5–7]. Although the crossing of horse and donkey is capable of producing offspring, the mule ($2n = 63$, offspring of male donkey and female horse) and hinny ($2n = 63$, offspring of male horse and female donkey) are scientifically considered to be incapable of natural mating to produce offspring.

According to Dobzhansky-Muller incompatibilities (DMIs), the evolutionary accumulation of mutational differences between species is attributed to the conflict and coevolution among and within the genomes of the organisms, which are deleterious genetic loci in their hybrid offspring [8,9]. Interactions between such two or more alleles left behind by parents will reduce the fitness of interspecific or interpopulation hybrids, including

hybrid sterility, inviability, lethality and weakness [10]. For young species or incomplete speciation, these disruptive interaction loci might arise at single genes evolving divergently across species through positive selection [11], at duplicate genes losing function in different paralogues from diverging populations [12], or at inversions and translocations [13,14], which ultimately drive speciation by causing intrinsic postzygotic reproductive isolation. For older species, such as the horse and donkey, except for the aforementioned detrimental positions driving speciation, more complex genetic incompatibilities have continuously accumulated across the genomes of two divergent species. We refer to these incompatibilities as generalized DMIs, which will produce interaction dysfunction when brought together in a hybrid background. It is possible that genes of the immune system have played important roles in generalized DMIs in vertebrates. Major histocompatibility complex (MHC) genes are characterized as trans-species polymorphisms maintained by pathogen-driven balanced selection [15,16]. Several studies in mice [17] and teleost fish [18] have implicated that these interspecific polymorphisms in the MHC genes will reduce the fitness of the F1 hybrids. Recent work about hybrid necrosis in three species of *Capsella* also showed that the polymorphisms fixed by balancing selection might become interspecific barriers [19]. Immunoglobulin genes (IGs) are not conserved across the closely related sister taxa, driven by the diversity of target microorganisms for IG responses [20]. The integration of divergent IG loci from *Mus musculus* subspecies into their F1 hybrids results in an apparent mismatch of heavy- and light-chain loci and reduced fitness, which predisposes them to autoimmune complications [21]. Allelic divergences between species may disrupt interactions between their products, such as parental regulatory divergences. This commonly results in hybrid misexpression, which will reduce the survival of offspring F1 [22]. Interspecific hybridization between two species of swordtail fish (e.g., *Xiphophorus birchmanni* × *Xiphophorus malinche* and *Xiphophorus maculatus* × *Xiphophorus hellerii*) causes melanoma in their hybrids [23]. This hybrid incompatibility is the result of an interaction between the melanoma receptor tyrosine-protein kinase (*xmrk*) gene and an unknown locus [24]. In addition to hybrid lethality and sterility, lethal melanocyte tumorigenesis in swordtail fish hybrids is an innovative mechanism for decreasing hybrid fitness.

Under this deleterious mutational stress during early embryonic development, the molecular mechanism by which heterozygotes from divergent species protect the survival of the F1 hybrids from self-correcting rejection remains unknown. Several studies have provided strong evidence that somatic mutations occur post-zygotically during early embryogenesis. These mutations are recognized as important raw materials for genetic diversity. However, they are generally considered to be a prominent cause of disease in normal individuals, and such diseases include cancer [25], autism [26], and some rare developmental disorders [27]. Nevertheless, some hybrids still survive their frequently mutated genomes, such as the high rate of offspring-specific mutations detected in the F1 hybrids of the goldfish × common carp cross [28]. Previous whole-genome mutation rate studies in plants showed that highly heterozygous lines have a higher mutation rate than homozygous lines. For example, there is a 3.6-fold higher rate in heterozygous thale cress, a 3.4-fold higher rate in heterozygous rice, and a more modest 1.6-fold increase in a hybrid peach tree (*Prunus davidiana* × *Prunus persica*) versus in a weakly heterozygous peach tree (*P. persica*) [29,30]. These early postzygotic mutations may reflect hidden genetic incompatibilities ascribed to evolutionarily accumulated mutational differences between species. However, whether these rapidly mutating single nucleotide polymorphisms (SNPs), triggered by interspecific hybridization stress, can buffer the imbalance between parental haploid genomes and thus protect the survival of hybrid individuals remains unknown.

Combining previous reports, more than 98% of the horse genome conserved sequence was covered (up to five depth) by the donkey genome data. (Table S1) [5,31]. The high similarity exposes one of the biggest challenges in sequencing the mule genome, which is the difficulty in discriminating homologous sequences inherited from progenitors. This limits the future analysis of the outcomes of combining two genomes. In this study, we sequenced the whole-genome sequences from an *Equus* parent-offspring trio using the

Illumina platform. Family-based sequencing is powerful in analyzing the inheritance patterns of genotypes. Guided by SNP markers, we assigned the alleles to their genomic inheritance and identified non-Mendelian SNPs and *de novo* SNPs in the mule genome. Through functional analysis of these rapidly mutating SNPs, we assessed the impact of hybridization on the survival of the mules.

2. Materials and Methods

2.1. Samples for Genomic DNA

Whole-genome sequences were characterized from three members of an *Equus* parent-offspring trio, consisting of a female horse, a male donkey and their hybrid offspring (a female mule). The three animals originated from the Xilingol League of Inner Mongolia, China. Data for donkey genome sequences used in this study was generated and stored in our laboratory [5]. For the horse and the mule, approximately 5 mL of peripheral blood was collected for DNA extraction. Blood samples were collected during veterinary examinations. No animal was hurt or captured as a result of these studies.

2.2. Genome Sequences

For the horse and the mule, DNA was extracted from peripheral blood cells. PE libraries were sequenced using the Illumina HiSeq X-ten (2 × 150 bp). Standard genomic library preparation and sequencing followed the manufacturer's instructions, and sequence reads were collected from the Illumina data processing pipeline.

2.3. Data Filtering

AdapterRemoval (version 2.0, Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark) [32] was used to trim adapter sequences from sequence reads generated by Illumina HiSeq X-ten. Low-quality sequences were defined using sliding windows of 5 bp with a step size of 1 bp. If the average quality value (Q) was <20 for five consecutive bases or the Phred quality score was ≤2 for the last base, we trimmed reads from the last base in the windows. Finally, only high-quality PE reads with ≥50 nucleotides were selected.

2.4. Mapping Reads to the Thoroughbred Horse Reference Sequence

High-quality PE reads from the horse, donkey and mule were mapped to the reference genome sequence of the thoroughbred horse (*EquCab 3.0*) using BWA (version 0.7.5a-r416, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge, UK) software with default parameters [33]. Unique alignments were generated using the SAMtools (Version 0.1.19-44428cd, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge, UK) package with the parameter options “-q 30” [34].

2.5. SNP Calling

The unique alignments were used for calling SNPs and Indels. SNPs were detected by comparison of mapped sequences between the reference genome and each sample. Loci with heterozygous and homozygous genotypes in each sample but different from the reference bases were detected as SNPs. Picard (version 1.93, The Broad Institute of Harvard and MIT, Cambridge, MA, USA) software was used to mark potential duplicates. SAMtools and GATK (version 3.5-0-g36282e4, Program in Medical and Population Genetics, The Broad Institute of Harvard and MIT, Cambridge, MA, USA) [35] were used to call SNPs for each individual separately, and only the intersection of SNPs identified by the two software programs was used for subsequent analysis [36]. The HaplotypeCaller program of GATK was used to obtain SNPs with the parameters: -stand_call_conf 30 and -stand_emit_conf 10. To obtain the high-confidence variants, the initially obtained candidate sites were strictly filtrated with the following parameters: -clusterSize 3 -clusterWindowSize 10, -filterExpression “QUAL < 30.0”, -filterExpression “QD < 2.0”, -filterName “FilterFS” -filterExpression “FS > 20.0”, -filterName “FilterMQ” -filterExpression “MQ < 20.0”, -filterName “Fil-

terMQRankSum", --filterExpression "MQRankSum < -3.0", --filterName "FilterReadPos-RankSum", --filterExpression "ReadPosRankSum < -3.0", --filterName "HaplotypeScore", --filterExpression "HaplotypeScore > 13.0" [37].

For the above candidate SNPs, further filtering was performed if they exhibited the following characteristics: (i) SNPs were located in low-complexity or simple-repeat regions, (ii) the read depth at the variant position was lower than 4 or higher than 50, (iii) as high frequency of false-positives (FPs) occurring around InDels, the adjacent 50 bp of target genomic regions were excluded, and (iv) variant sites located from the gap within 3 bp and adjacent 10 bp sites.

2.6. Genotype Inheritance State and Mutation Analysis

"Genotype" refers to both alleles at one position. There is a genotype position in each genome we sequenced that corresponds to each position in the reference genome. As genotype callers, such as GATK, assign diploid genotypes to autosomal loci, regions of heterozygous deletion are erroneously assigned as homozygous genotypes. In the context of triple designs, variants within heterozygous deletions frequently exhibit Mendelian errors as a result of this genotype misassignment [38]. Therefore, we masked these SNPs located in copy number variation (CNV) regions detected by CNVnator [39] and repetitive sequence regions detected by RepeatMasker [40].

Two methods were chosen to analyze genotype transmission patterns in *Equus* progenitor-progeny, and only the intersection of non-Mendelian inheritance loci identified by the two methods were used for subsequent analysis. First, referring to Roach's method [41], we analyzed the genotype transmission patterns in three family members using bcftools (version 1.3.1, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge, UK) software and vcftools (version 0.1.16-15, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge, UK) software. Second, we directly screened SAMtools mpileup data to identify *de novo* SNPs and Mendelian inheritance error (MIE) SNPs using the "trio" command in VarScan (version 2.4.2) software with the following parameters: --min-coverage 10, --min-var-freq 0.20, --p-value 0.05, --adj-var-freq 0.05, and --adj-p-value 0.15 [42]. These non-Mendelian SNPs were retained, which met the following two requirements: (1) every SNP was supported by no less than ten reads; (2) the read counts of SNPs from the two parental alleles were not fewer than five [43].

2.7. Gene Annotation of MIE SNPs and De Novo SNPs

ANNOVAR software was used to annotate the function of SNPs in the mule genome. For SNPs located in intergenic regions, only genes within 5 kb were retained. The clusterProfiler package of R software was used to KEGG enrichment analysis with the following parameters: organism = "ecb", keyType = "kegg".

3. Results

3.1. The *Equus* Parent-Offspring Trio Genomes

Whole-genome sequences from an *Equus* parent-offspring trio were analyzed. The donkey genome used in this study was stored in our laboratory (Table S2) [5]. For mule and horse genomes, PE libraries were constructed and sequenced on the Illumina HiSeq X-ten platform (Table S3). In total, 103.78 Gb, 100.39 Gb, and 114.36 Gb of high-quality genome sequences were generated for the horse, donkey and mule, respectively, after stringent filtering (Table S4). Approximately equal amounts of data were intercepted to avoid compromising analytical accuracy due to the difference in the original input data. Approximately 97 Gb bases with an average depth coverage of 38× were aligned to the thoroughbred horse reference genome (*EquCab3.0*) (Tables S5 and S6).

Unique alignments were retained for subsequent analysis after filtering with mapping quality < 30. In total, 28,272,430, 5,918,968, and 28,287,142 SNPs were identified in the genomes of the donkey, horse and mule, respectively, by both the SAMtools and GATK software programs (Figures S1 and S2). These initially obtained candidate sites were further

examined to minimize systematic errors and false-positives for the accuracy. Furthermore, we only considered autosomes, for which SNPs were evenly distributed across chromosomes (Figure S3). Of these, there were approximately 24,861,384 positions at which at least one family member had an allele that varied from the reference genome (Figure S4).

As shown in Table 1, we identified 5,012,403 SNPs in the Mongolian horse. This was comparable to the results of previous reports, which reported a range from 3,639,479 to 6,040,778 [44,45]. Heterozygous SNPs in the Mongolian horse were more abundant than homozygous SNPs, which was also consistent with the previous reports above. We identified 23,819,055 SNPs in the domestic donkey (thoroughbred horse genome as reference), and this value was also comparable to previous reports of approximately 24,076,918 SNPs [2]. The frequency of SNPs in the donkey genome (0.9501%) and the mule genome (0.9344%) was considerably higher than that in the horse genome (0.1999%), which could be explained by the identification of SNPs using the thoroughbred horse genome as the reference.

Table 1. Summary of SNPs aligning to the horse reference genome (*EquCab3.0*).

Samples	Donkey	Horse	Mule
Depth	$4 \leq \text{depth} \leq 50$	$4 \leq \text{depth} \leq 50$	$4 \leq \text{depth} \leq 50$
Heter. SNPs	1,996,879	3,387,403	21,771,865
Homo. SNPs	21,822,176	1,625,000	1,654,376
Total SNPs	23,819,055	5,012,403	23,426,241
%SNP	0.950115	0.199939	0.934446
% Heterozygosity	0.0797	0.135	0.868
Transitions	16,302,515	3,386,154	16,029,003
Transversions	7,516,540	1,626,249	7,397,238
Ti/Tv (autosome)	2.17	2.08	2.17

3.2. Characterization of SNP Transmission from Parents to Offspring

The purpose of the present study was to identify and analyze the genetic signature of point mutations in the mule genome from the parent-offspring trio. Referring to Roach's method [41], in this *Equus* trio family, the transmission pattern for each variant position was grouped into three inheritance states. First, the mule received one allele from the horse and the other from the donkey, whose variant position was classified as a Mendelian inheritance SNP (Table S7). Second, the mule received a pair of alleles from one parent, whose variant position was classified as a Mendelian inheritance error (MIE) SNP (Table S8). Third, the mule received at least one allele that was not from the parents, whose variant position was classified as a *de novo* SNP (Table S9).

Of the 24,861,384 variant positions, 14,154,723 were inherited through the Mendelian inheritance pattern in parent-offspring members (Table S7). A total of 25,703 MIE SNPs transmitted from parental genomes were identified in the mule genome (Figure 1, Tables S8 and S10), which is almost 32 times more than the number of natural MIE mutations in chimpanzees (794) from a chimpanzee parent-offspring trio [46]. Compared with parental SNPs, 555 SNPs in the mule genome were characterized as novel, which is almost 12 times more than the natural *de novo* mutations in the chimpanzees (45) mentioned above (Figure 1, Tables S9 and S10). The rate of *de novo* SNPs is 2.21×10^{-7} in the *Equus* parent-offspring trio. The natural mutation rate for *Equus* is 7.24×10^{-9} (per site per generation) [1] and ranges between 0.82 and 1.70×10^{-8} in humans [47,48].

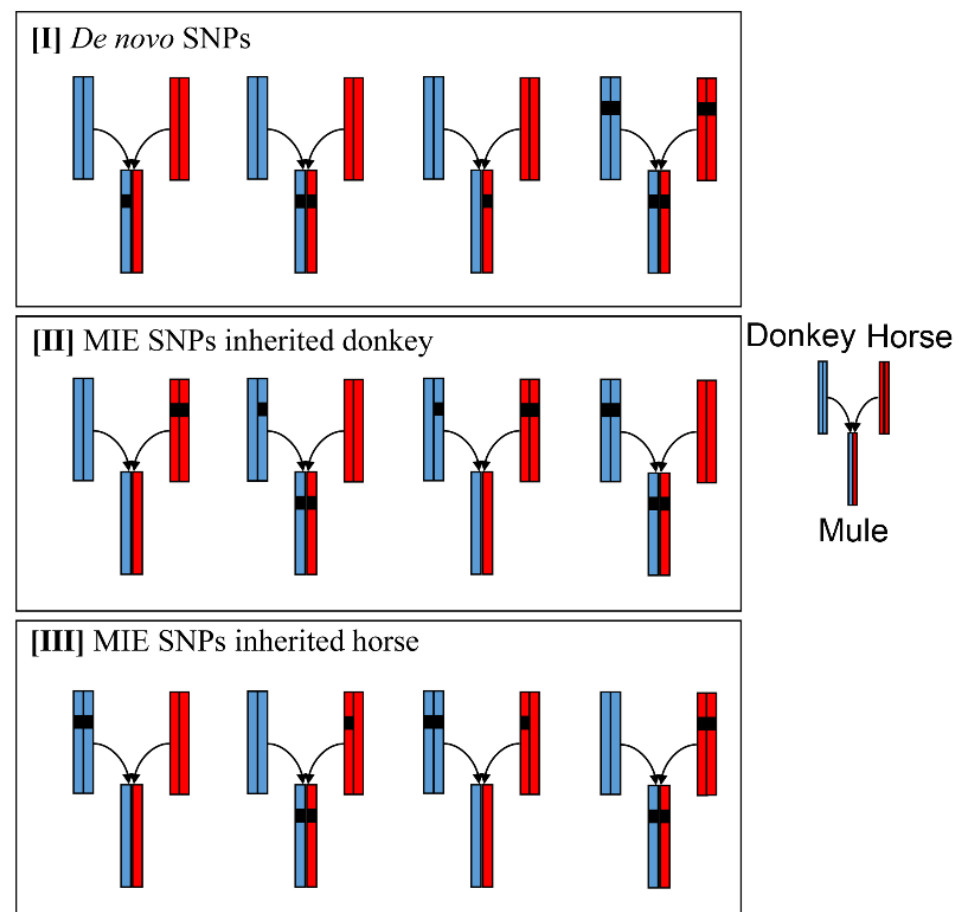


Figure 1. Analysis of non-Mendelian inheritance SNPs. Classification of non-Mendelian inheritance SNPs. When variant alleles were identified only in the mule, they were classified as [I] *de novo* SNPs. MIE SNPs were classified into [II] inherited from the donkey and [III] inherited from the horse.

3.3. Functional Annotation of SNPs

To investigate the potential functions of these SNPs in the mule genome, we searched for genes within approximately 5 kb of the SNPs using ANNOVAR software. A total of 5625 genes were annotated. In total, 11,595 MIE SNPs inheriting from the donkey were embraced by 1453 genes, the other 14,108 MIE SNPs from the horse were embraced by 4557 genes, and 555 *de novo* SNPs were embraced by 658 genes (Table S10). As shown in Figure 2, a high frequency of SNPs was located in intergenic regions (14,927, 56.85%); this was followed by intronic regions (6506, 24.78%), ncRNA intronic regions (1665, 6.34%), exons (956, 3.64%), upstream regions (673, 2.56%), downstream regions (608, 2.32%), UTR3 regions (386, 1.47%), ncRNA exons (374, 1.42%), and UTR5 regions (149, 0.58%). These findings are consistent with the general distribution reported previously.

Of these, lots of point mutant SNPs were located directly within or adjacent to immune genes, including MHC class I genes (e.g., *MHCX1*, *EQMCE1*) and MHC class II genes (e.g., *DRA*, *DQB*), immunoglobulin family genes (e.g., *LOC100054595*) and various critical cytokines, such as interleukin genes (e.g., *IL-1 β* , *IL-12 β*), interferon genes (e.g., *IFN- γ*), and intercellular adhesion molecule 1 (*ICAM1*). In the F1 heterozygous mule, genes with mutations in DNA sequences were also highly concentrated in DNA replication (e.g., *PRIM1*, *POLD2*), DNA repair (e.g., *PMS2* and *MSH2*), and cancer, such as proto-oncogenes (e.g., *KRAS*, *HRAS*) and tumor-suppressor genes (e.g., *APC*, *PTEN*).

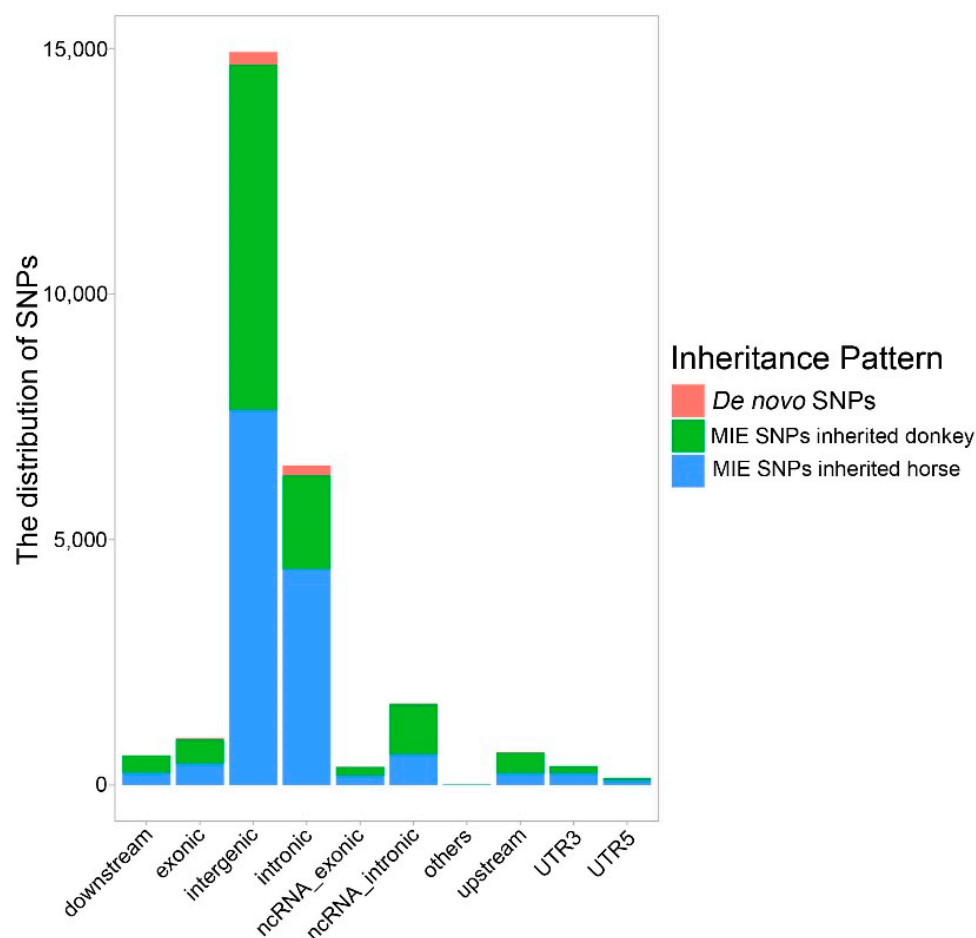


Figure 2. The statistics of SNP distribution.

3.4. KEGG Pathway Enrichment Analysis

Pathway analysis showed that the mutant genes were enriched in 344 KEGG pathways (Table S11). A large number of pathways playing crucial roles in the immune response processes, including antigen processing and presentation, specific antigen recognition, naive lymphocyte activation, proliferation and differentiation, and effector cell generation and production of effects, were enriched (Table S11). The antigen processing and presentation pathway (ecb04612) was enriched by mutant genes in this study, including MHC class I genes (e.g., *MHCX1*, *EQMCE1*) and MHC class II genes (e.g., *DRA*, *DQB1*), by which peptide antigens derived from cytosolic proteins, both self and foreign, are processed for presentation at the cell surface to restricted T lymphocytes and initiate adaptive immune responses (Figure 3) [49]. In addition, endocytosis (ecb04144) and FcγR-mediated phagocytosis (ecb04666), which are essential for antigen presenting cells to recognize and internalize exogenous antigens, were enriched. Numerous pathways involving the subsequent processing of antigens were enriched, including phagosome (ecb04145), lysosome (ecb04142), and the ubiquitin-proteasome degradation process (ecb04120, ecb03050). Through these processes, peptide substrates are degraded into suitable antigenic peptides for presentation to T cells. The T cell receptor signaling pathway (ecb04660, e.g., *FYN*, *GRB2*) was enriched. The essential function of this pathway is to transduce extracellular stimuli into intracellular signals and then initiates signaling cascades, culminating in the response to antigen receptor engagement. The MAPK signaling pathway (ecb04010) and NF-κB signaling pathway (ecb04064), two major downstream pathways of TCR signal propagation, were enriched. The Th1 and Th2 cell differentiation pathway (ecb04658) and Th17 cell differentiation pathway (ecb04659) were enriched, which directly coordinate CD4⁺ T cell differentiation. These differentiation programs are also dynamically regulated by the JAK-STAT signaling

pathway (ecb04630). An appropriate cellular energy environment ensures that naive T cells successfully proliferate and differentiate into specific T cell subsets for an effective immune response. Multiple metabolism-related signaling pathways were enriched in this study, such as PI3K-Akt (ecb04151), mTOR (ecb04150), AMPK (ecb04152), and Foxo (ecb04068) pathways, which are required to coordinate T cell metabolic activity [50–53].

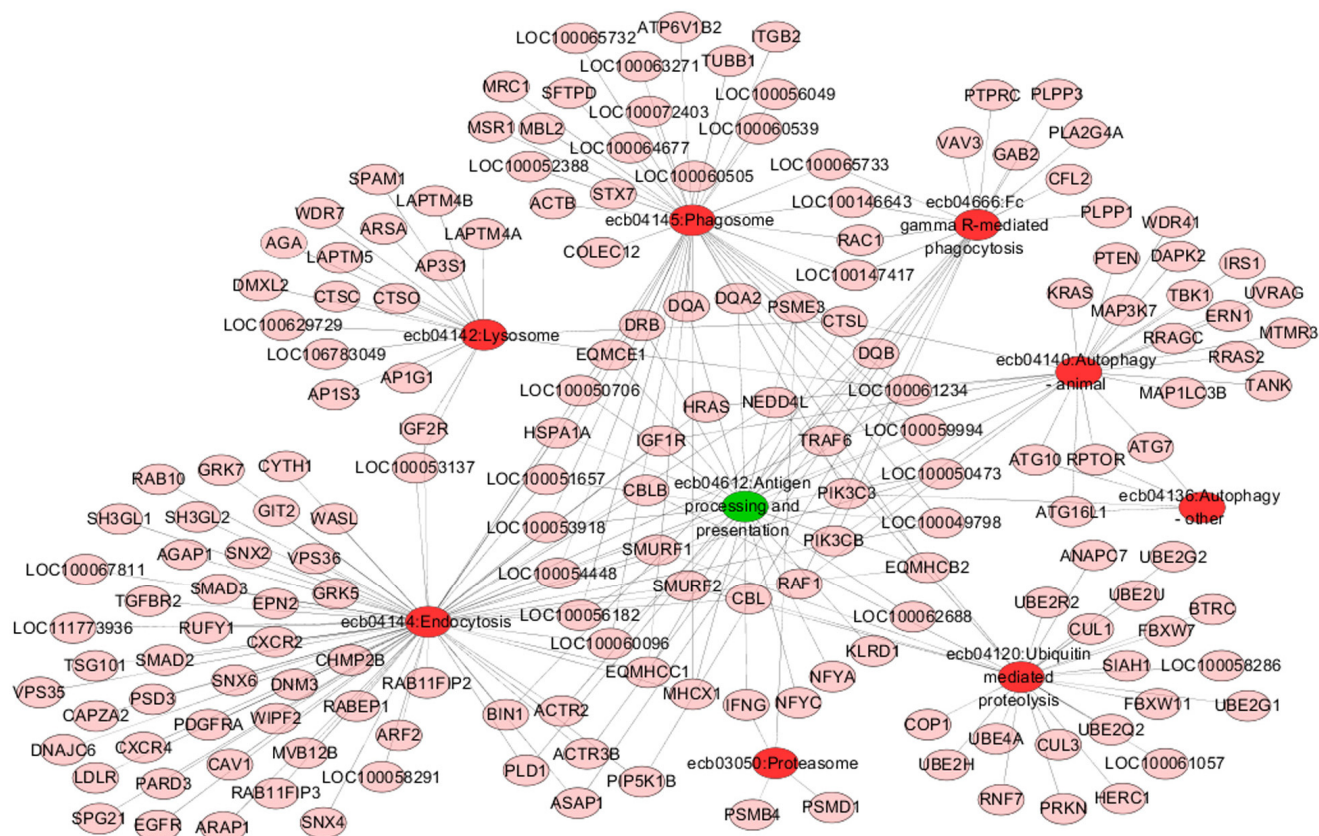


Figure 3. Integrated pathway network analysis of point mutation genes involved in antigen processing and presentation in the mule genome. Green and red represent pathways; pink represents genes.

Many mutant genes are directly involved in cell fate processes, such as the cell cycle (ecb04110, e.g., *CCNB1*, *CCND1*, *CDC25A*) and cellular senescence (ecb04218, e.g., *ATM*) pathways. These mutant genes were also involved in the cAMP signaling pathway (ecb04024, e.g., *PDE10A*, *PDE4D*) and the cGMP-PKG signaling pathway (ecb04022, e.g., *PDE5A*, *PDE3A*). These interacting kinases are critical for regulating cell fate by dynamically regulating the concentration of cAMP [54]. The DNA replication pathway (ecb03030, e.g., *PRIM1*, *POLD2*), which regulates the DNA replication process, was enriched. Mutant genes were also enriched in DNA damage responses, including mismatch repair (ecb03430, e.g., *MSH2*), nucleotide excision repair (ecb03420, e.g., *ERCC4*), base excision repair (ecb03410, e.g., *PARP4*), and two major double-strand break repair pathways: non-homologous end-joining (ecb03450) and homologous recombination (ecb03440) (Figure S5) [55,56].

Pathways in cancer (ecb05200) and another 17 cancer- or tumor-related pathways (e.g., colorectal cancer, renal cell carcinoma and pancreatic cancer) were directly enriched by mutant genes in the mule genome (Table S11). In addition, transcriptional regulation of cancer pathways, such as transcriptional misregulation in cancer (ecb05202, e.g., *MYC*, *MLLT10*) and microRNAs in cancer (ecb05206, e.g., *MIR125B*), were enriched. Several metabolic regulatory pathways participated in cancer, such as the proteoglycans in cancer pathway (ecb05205, e.g., *TWIST2*, *HGF*, *EGFR*), choline metabolism in cancer pathway (ecb05231, e.g., *PLA2G4A*, *PLD1*) and central carbon metabolism in cancer pathway (ecb05230, e.g., *LDHB*, *HK2*), were also enriched. The viral carcinogenesis pathway (ecb05203) and chemical

carcinogenesis pathway (ecb05204, ecb05207, ecb05208, e.g., the cytochrome P450 gene family, the glutathione-S-transferase gene family, and the uridine diphosphate-glucuronosyl transferase gene family) were enriched. Many cancer-related pathways, including the wnt (e.g., *NLK*, *DAAM1*) and p53 signaling pathways, were also specifically enriched by mutant genes in the mule [57,58].

4. Discussion

In this study, the rates of SNP in the domestic donkey and Mongolian horse are comparable to previous reports [2,44,45]. Referring to the reports of Tatsumoto [46] and Roach [41], point mutations, including MIE SNPs and *de novo* SNPs, were identified in the mule genome through the whole-genome sequences of an *Equus* parent-offspring trio. The mule has a higher mutation rate than the natural mutation rate in chimpanzees and humans, as well as in *Equus* [1,46,47], which is consistent with the hypothesis tested experimentally by Duncan in 1915 that interracial crosses may have a high mutation rate [59]. A limited number of plant genome-wide studies also indicated that heterozygotes have higher mutation rates than homozygotes [29,30]. We cannot deny that there were incorrectly identified SNPs due to limited sequencing data, although the mutation rate in mules conformed to the rule that heterozygotes have higher mutation rate.

In previous reports by Liu and his colleagues, they identified 12,146 (the genomes of goldfish as reference,) and 58,587 (the genomes of common carp as reference) offspring-specific mutations in the hybrid fish transcriptome [28]. The mutation rate in hybrid fish is obviously higher than that observed in mules. However, we observed that the mutation rate in the mule was ten times higher than the one observed in heterozygous plants [30]. That is contrary to the previous conclusion that plants have higher point mutation rates than animals, depending on the transgenic tissue analyzed [60]. Mutation rates vary among species, which may be attributed to the species itself or the length of separation time. These results were generated by a few studies on a limited number of experimental samples. In the future, we need to further verify these conclusions by relying on more samples.

In summary of previous reports [48,61,62], the origin of these point mutations may be as follows: (i) germline events induced by parental meiosis during gametogenesis, (ii) postzygotic events that occurring in offspring mitosis early during embryogenesis, or (iii) specific mutations in blood cells of the offspring after tissue differentiation. Our working speculates that the higher mutation rate in the mule genome is attributable to rapid post-zygotic somatic mutation triggered by shock stress integrating the horse and donkey haploids into the mule genome. The molecular mechanism of the high mutation rate in mule genome may be the defect of DNA mismatch repair caused by sequence divergence in horses and donkeys [63], but the data in this study is not enough to explain the detailed mechanism. According to the above hypothesis, these point mutations in mule may reflect the incompatible loci in mutational differences accumulated between the horse and the donkey. We also speculate that the rapidly mutating SNP may act as a buffer to balance the incompatibility of parental haploid genomes and thus protect the survival of mule.

MHC genes play critical roles in maintaining appropriate immune homeostasis and self-tolerance, and are central to the vertebrate immune response, which is necessary for health. Similarly to most mammals, *Equus* MHC genes are characterized as extreme polymorphisms and trans-species polymorphisms that are maintained by pathogen-driven balanced selection, and lineage-specific alleles in the antigen binding site (ABS) facilitated by geographic subdivision between horses and donkeys [64–66]. This genetic heterozygosity may reduce the fitness, survival, or reproductive success of mules, as reported in a study on the crossbreeding of mice from natural populations by Petteri [17]. Genetic data from various species demonstrate that immunoglobulin loci driven by the diversity of target microorganisms for IG responses are not conserved across the closely related sister taxa. The uncoordinated evolution of heavy- and light-chain gene sets of IGs between the horse and donkey may result in poor interactions in the mule, ultimately reducing fitness, as illustrated in a study using hybrid F1 mice by Watson [21].

It is possible that genes of the immune system have played important roles in inter-specific hybridization incompatibility. Autoimmune-like responses induced by DMI have been described in hybrid necrosis in plants [67]. These incompatibilities occur when gene divergence affects loci encoding interacting products, such as receptors and their ligands. During negative selection in the thymus, the strength of the pMHC:TCR interaction is higher than the threshold required for negative selection due to MHC polymorphisms located at or near the peptide-binding groove, resulting in failing to trigger depletion of autoreactive T cells in the medulla [68]. Polymorphisms at MHC class II (e.g., *DQA*, *DQB*, *DRB*) loci account for a higher risk for all autoimmune diseases, such as rheumatoid arthritis [69], type I diabetes mellitus [70] and systemic lupus erythematosus [71], than any other loci in the genome. The impeccable performance of immune responses and immune homeostasis reflects the accurate delivery of signals from external stimuli, which is affected by various factors, such as the intensity and duration of TCR or BCR signaling, the strength of downstream cascade signaling, the specificity of the binding cytokine with its receptor and appropriate gene expression regulated by transcription factors [72]. ICAM1 cooperates with TCR-pMHC complexes to form immune synapses and generate costimulatory signals in response to antigenic stimulation [73]. Polymorphisms of the *ICAM1* gene also play a crucial role in the pathological process of rheumatoid arthritis [74]. TGF- β and IL-1 β have the capacity to instruct naive T cells to develop into IL-17-producing Th17 cells which trigger inflammatory responses, including neutropenia, tissue remodeling, and the production of antimicrobial proteins [75]. As previously reported, SNPs in the proinflammatory cytokine genes *IL-1 β* and *IFN- γ* are associated with susceptibility to rheumatoid arthritis [76]. SNPs in interleukin *IL-12 β* is associated with susceptibility to inflammatory bowel diseases [77], and SNPs in anti-inflammatory *IL-4/IL-4R* is associated with susceptibility to type I diabetes mellitus [78].

Mutations might change the enzyme activity or fidelity by affecting standard transcriptional processing. The accumulation of mutational differences in critical kinases that regulate the cell cycle, DNA replication, and DNA repair between horses and donkeys may alter enzyme specificity in mules. Ultimately, this would severely restrict cell fate, manifesting as genomic instability, apoptosis or carcinogenesis. Periodic and phase-specific cyclin expression regulates the cell cycle by activating cyclin-dependent kinases. Overexpression of the cyclin D1 (*CCND1*) gene leads to uncontrolled cell cycle regulation and abnormal cell proliferation, eventually resulting in tumorigenesis [79]. It is well-accepted that *PRIM1* is responsible for DNA replication initiation by synthesizing short RNA-DNA primers, and *POLD*, a high-fidelity DNA polymerase, is used for processive elongation on the leading and lagging strands of DNA [80,81]. DNA polymerase also plays a crucial role in DNA repair. DNA polymerase inefficiency may cause detrimental consequences, such as chromosomal instability and oncogene activation, which lead to carcinogenesis. Previous reports by Ranga et al. in mice indicated that heterozygous mutations at the polymerase active site of DNA polymerase reduces lifespan, increases genomic instability, and accelerates tumorigenesis in an allele-specific manner [82]. As implied by Lynch in the *MLH1* (also known as *PMS2*)-deficient Baker's yeast strain, DNA mismatch repair genes repair over 90% of replication errors [83]. The heterozygosity of enzymes, integrating the divergent parental haploids, may alter enzyme efficacy, which may also be a reason for the high frequency of mutations in mule.

A complex regulatory network integrating gene expression, energy metabolism, and extracellular matrix controls cell fate. However, dysregulation of this network remodeling can lead to tumorigenesis and cancer development by providing favorable conditions for tumor cells [84,85]. Mutations in proto-oncogenes (e.g., *KRAS*, *HRAS*) and tumor-suppressor genes (e.g., *APC*, *PTEN*) are the primary drivers in various cancers [86–88]. Altered gene expression is another primary molecular mechanism of cancer pathology. The *MYC* gene and the *MIR125B* microRNA are directly involved in establishing the specific programs of gene expression in cancer cells, comprising almost all aspects of cancer biology, such as proliferation, apoptosis, invasion/metastasis, and angiogenesis [85,89]. The proteoglycans

in cancer pathway (ecb05205) regulates proteoglycans in the extracellular matrix, which influences the behavior of cancer cells and their microenvironment by interacting with various cytokines (e.g., *TWIST2*), growth factors (e.g., *HGF*) and cell surface receptors (e.g., *EGFR*) [90]. The choline metabolism in cancer pathway (ecb05231) leads to increased levels of choline-containing precursors via the modulation of enzymes (e.g., *PLA2G4A*, *PLD1*) [91]. The central carbon metabolism in cancer pathway (ecb05230, e.g., *LDHB*, *HK2*) controls cancer metabolic adaptations through aerobic glycolysis, elevated glutaminolysis, dysregulated tricarboxylic acid cycle and pentose phosphate pathways [92]. Chemical compound exposure and viral infection are two common exogenous factors responsible for carcinogenesis. In general, chemical carcinogens require metabolic activation to produce reactive intermediates capable of binding to cellular macromolecules, which are necessary for chemicals to cause cancer. The cytochrome P450 gene family, the glutathione-S-transferase gene family, and the uridine diphosphate-glucuronosyl transferase gene family are critical enzymes responsible for the metabolism of chemical carcinogens [93].

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes13122188/s1>, Figure S1: Venn diagram showing unique and shared SNPs identified by two software in (A) donkey, (B) horse, and (C) mule genomes; Figure S2: Venn diagram showing unique and shared InDels identified by two software in (A) donkey, (B) horse, and (C) mule genomes; Figure S3: the distribution of SNPs on autosome for (A) the donkey, (B) the horse and (C) the mule; Figure S4: the annotation of SNPs; Figure S5: summary of point mutation genes involved in DNA repair processes in the mule genome; Table S1: statistics regarding mapping of donkey raw reads to the thoroughbred horse genome (*EquCab3.0*); Table S2: summary of genome sequencing for the male donkey; Table S3: summary of genome sequencing for the female horse and the female mule; Table S4: qualified data for the horse and mule genomes; Table S5: summary of mapping against the thoroughbred horse reference genome (*EquCab3.0*); Table S6: sequencing depth coverage; Table S7: genotype combinations for the Mendelian inheritance; Table S8: genotype combinations for Mendelian inheritance errors; Table S9: genotype combinations for *de novo* SNPs in the mule genome; Table S10: the detail of non-Mendelian inheritance variants; Table S11: Pathway analysis of MIE and De novo mutant genes.

Author Contributions: M.D. (Manglai Dugarjaviin), D.B., Y.Z., X.Z., M.D. (Ming Du) and X.R. designed and managed the project. X.W., Y.S., T.B. and X.R. collected samples and prepared the nucleic acid samples. B.L., G.B., Y.S., T.B. and X.R. performed the genomes sequencing. X.W. and X.R. designed and performed the formal analysis. X.R., Y.L. and M.D. (Manglai Dugarjaviin) wrote and revised the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the international (regional) cooperation and exchange project (31961143025), the National Natural Science Foundation of China and the Mongolian foundation of science and technology (NSFC-MFST) Joint Project (3191101008), National Natural Science Foundation of China (31960657), Inner Mongolia Autonomous Region Major Science and Technology Project (2021) (2021ZD0018) and Natural Science Foundation of Inner Mongolia Autonomous Region (2020BS030345). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Institutional Review Board Statement: This study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Animal Care and Use Committee at Inner Mongolia Agricultural University.

Informed Consent Statement: Not applicable.

Data Availability Statement: The genome data were submitted to NCBI as project accession PRJNA842856 (SRA accession: SRR19427107, SRR19427108) for the horse and mule. The remaining data are available within the article and its supplementary materials files or available from the authors upon request.

Acknowledgments: We thank Jinlong Huang (Department of Pathology, Memorial Sloan Kettering Cancer Center, USA) for valuable comments to this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Orlando, L.; Ginolhac, A.; Zhang, G.; Froese, D.; Albrechtsen, A.; Stiller, M.; Schubert, M.; Cappellini, E.; Petersen, B.; Moltke, I.; et al. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **2013**, *499*, 74–78. [\[CrossRef\]](#) [\[PubMed\]](#)
- Jonsson, H.; Schubert, M.; Seguin-Orlando, A.; Ginolhac, A.; Petersen, L.; Fumagalli, M.; Albrechtsen, A.; Petersen, B.; Korneliussen, T.S.; Vilstrup, J.T.; et al. Speciation with gene flow in equids despite extensive chromosomal plasticity. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 18655–18660. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bush, G.L.; Case, S.M.; Wilson, A.C.; Patton, J.L. Rapid speciation and chromosomal evolution in mammals. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 3942–3946. [\[CrossRef\]](#) [\[PubMed\]](#)
- Trifonov, V.A.; Stanyon, R.; Nesterenko, A.I.; Fu, B.; Perelman, P.L.; O'Brien, P.C.; Stone, G.; Rubtsova, N.V.; Houck, M.L.; Robinson, T.J.; et al. Multidirectional cross-species painting illuminates the history of karyotypic evolution in Perissodactyla. *Chromosome Res. Int. J. Mol. Supramol. Evol. Asp. Chromosome Biol.* **2008**, *16*, 89–107. [\[CrossRef\]](#) [\[PubMed\]](#)
- Huang, J.; Zhao, Y.; Bai, D.; Shiraigol, W.; Li, B.; Yang, L.; Wu, J.; Bao, W.; Ren, X.; Jin, B.; et al. Donkey genome and insight into the imprinting of fast karyotype evolution. *Sci. Rep.* **2015**, *5*, 14106. [\[CrossRef\]](#) [\[PubMed\]](#)
- Renaud, G.; Petersen, B.; Seguin-Orlando, A.; Bertelsen, M.F.; Waller, A.; Newton, R.; Paillot, R.; Bryant, N.; Vaudin, M.; Librado, P.; et al. Improved de novo genomic assembly for the domestic donkey. *Sci. Adv.* **2018**, *4*, eaaq0392. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wade, C.M.; Giulotto, E.; Sigurdsson, S.; Zoli, M.; Gnerre, S.; Imsland, F.; Lear, T.L.; Adelson, D.L.; Bailey, E.; Bellone, R.R.; et al. Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* **2009**, *326*, 865–867. [\[CrossRef\]](#)
- Muller, H.J. Isolating mechanisms, evolution and temperature. *Biol. Symp.* **1942**, *6*, 71–125.
- Dobzhansky, T. Genetics and the Origin of Species. *Nature* **1959**, *184*, 587–588. [\[CrossRef\]](#)
- Fishman, L.; Sweigart, A.L. When Two Rights Make a Wrong: The Evolutionary Genetics of Plant Hybrid Incompatibilities. *Annu. Rev. Plant Biol.* **2018**, *69*, 707–731. [\[CrossRef\]](#)
- Presgraves, D.C.; Balagopalan, L.; Abmayr, S.M.; Orr, H.A. Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. *Nature* **2003**, *423*, 715–719. [\[CrossRef\]](#) [\[PubMed\]](#)
- Zuellig, M.P.; Sweigart, A.L. gene duplicates cause hybrid lethality between sympatric species of *mimulus*. *PLoS Genet* **2018**, *14*, e1007130. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hoffmann, A.A.; Rieseberg, L.H. Revisiting the Impact of Inversions in Evolution: From Population Genetic Markers to Drivers of Adaptive Shifts and Speciation? *Annu. Rev. Ecol. Evol. Syst.* **2008**, *39*, 21–42. [\[CrossRef\]](#) [\[PubMed\]](#)
- Stathos, A.; Fishman, L. Chromosomal rearrangements directly cause underdominant F1 pollen sterility in *Mimulus lewisii*-*Mimulus cardinalis* hybrids. *Evol. Int. J. Org. Evol.* **2014**, *68*, 3109–3119. [\[CrossRef\]](#)
- Klein, J.; Sato, A.; Nikolaidis, N. MHC, TSP, and the origin of species: From immunogenetics to evolutionary genetics. *Annu. Rev. Genet.* **2007**, *41*, 281–304. [\[CrossRef\]](#)
- Gillingham, M.A.; Courtiol, A.; Teixeira, M.; Galan, M.; Bechet, A.; Cezilly, F. Evidence of gene orthology and trans-species polymorphism, but not of parallel evolution, despite high levels of concerted evolution in the major histocompatibility complex of flamingo species. *J. Evol. Biol.* **2016**, *29*, 438–454. [\[CrossRef\]](#)
- Ilmonen, P.; Penn, D.J.; Damjanovich, K.; Morrison, L.; Ghotbi, L.; Potts, W.K. Major histocompatibility complex heterozygosity reduces fitness in experimentally infected mice. *Genetics* **2007**, *176*, 2501–2508. [\[CrossRef\]](#)
- Malmstrom, M.; Matschiner, M.; Torresen, O.K.; Star, B.; Snipen, L.G.; Hansen, T.F.; Baalsrud, H.T.; Nederbragt, A.J.; Hanel, R.; Salzburger, W.; et al. Evolution of the immune system influences speciation rates in teleost fishes. *Nat. Genet.* **2016**, *48*, 1204–1210. [\[CrossRef\]](#)
- Sicard, A.; Kappel, C.; Josephs, E.B.; Lee, Y.W.; Marona, C.; Stinchcombe, J.R.; Wright, S.I.; Lenhard, M. Divergent sorting of a balanced ancestral polymorphism underlies the establishment of gene-flow barriers in *Capsella*. *Nat. Commun.* **2015**, *6*, 7960. [\[CrossRef\]](#)
- Collins, A.M.; Watson, C.T.; Breden, F. Immunoglobulin genes, reproductive isolation and vertebrate speciation. *Immunol. Cell Biol.* **2022**, *100*, 497–506. [\[CrossRef\]](#)
- Watson, C.T.; Kos, J.T.; Gibson, W.S.; Newman, L.; Deikus, G.; Busse, C.E.; Smith, M.L.; Jackson, K.J.; Collins, A.M. A comparison of immunoglobulin IGHV, IGHD and IGHJ genes in wild-derived and classical inbred mouse strains. *Immunol. Cell Biol.* **2019**, *97*, 888–901. [\[CrossRef\]](#) [\[PubMed\]](#)
- Mack, K.L.; Nachman, M.W. Gene Regulation and Speciation. *Trends Genet. TIG* **2017**, *33*, 68–80. [\[CrossRef\]](#) [\[PubMed\]](#)
- Lu, Y.; Sandoval, A.; Voss, S.; Lai, Z.; Kneitz, S.; Boswell, W.; Boswell, M.; Savage, M.; Walter, C.; Warren, W.; et al. Oncogenic allelic interaction in *Xiphophorus* highlights hybrid incompatibility. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 29786–29794. [\[CrossRef\]](#) [\[PubMed\]](#)
- Powell, D.L.; Garcia, M.; Keegan, M.; Reilly, P.; Schumer, M. Natural hybridization reveals incompatible alleles that cause melanoma in swordtail fish. *Cold Spring Harb. Lab.* **2019**, *368*, 731–736. [\[CrossRef\]](#) [\[PubMed\]](#)
- Levin-Sparenberg, E.; Bylsma, L.C.; Lowe, K.; Sangare, L.; Fryzek, J.P.; Alexander, D.D. A Systematic Literature Review and Meta-Analysis Describing the Prevalence of KRAS, NRAS, and BRAF Gene Mutations in Metastatic Colorectal Cancer. *Gastroenterol. Res.* **2020**, *13*, 184–198. [\[CrossRef\]](#) [\[PubMed\]](#)

26. 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. [[CrossRef](#)] [[PubMed](#)]
27. Hoischen, A.; van Bon, B.W.; Gilissen, C.; Arts, P.; van Lier, B.; Steehouwer, M.; de Vries, P.; de Reuver, R.; Wieskamp, N.; Mortier, G.; et al. De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nat. Genet.* **2010**, *42*, 483–485. [[CrossRef](#)] [[PubMed](#)]
28. Liu, S.; Luo, J.; Chai, J.; Ren, L.; Zhou, Y.; Huang, F.; Liu, X.; Chen, Y.; Zhang, C.; Tao, M.; et al. Genomic incompatibilities in the diploid and tetraploid offspring of the goldfish x common carp cross. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 1327–1332. [[CrossRef](#)] [[PubMed](#)]
29. Yang, S.; Wang, L.; Zhang, X.; Yuan, Y.; Chen, J.-Q.; Hurst, L.D.; Tian, D. Parent-progeny sequencing indicates higher mutation rates in heterozygotes. *Nature* **2015**, *523*, 463–467. [[CrossRef](#)] [[PubMed](#)]
30. Xie, Z.; Wang, L.; Wang, L.; Wang, Z.; Lu, Z.; Tian, D.; Yang, S.; Hurst, L.D. Mutation rate analysis via parent-progeny sequencing of the perennial peach. I. A low rate in woody perennials and a higher mutagenicity in hybrids. *Proc. R. Soc. Biol. Sci.* **2016**, *283*, 1016. [[CrossRef](#)] [[PubMed](#)]
31. Cho, Y.S.; Hu, L.; Hou, H.; Lee, H.; Xu, J.; Kwon, S.; Oh, S.; Kim, H.M.; Jho, S.; Kim, S.; et al. The tiger genome and comparative analysis with lion and snow leopard genomes. *Nat. Commun.* **2013**, *4*, 2433. [[CrossRef](#)] [[PubMed](#)]
32. Schubert, M.; Lindgreen, S.; Orlando, L. AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **2016**, *9*, 88. [[CrossRef](#)]
33. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)] [[PubMed](#)]
34. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)]
35. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernysky, 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. [[CrossRef](#)]
36. Poplin, R.; Chang, P.; Alexander, D.; Schwartz, S.; Colthurst, T.; Ku, A.; Newburger, D.; Dijamco, J.; Nguyen, N.; Pt, A. Creating a universal SNP and small indel variant caller with deep neural networks. *Cold Spring Harb. Lab.* **2016**, *36*, 983–987. [[CrossRef](#)]
37. Auwera, G.A.V.D.; Carneiro, M.O.; Hartl, C.; Poplin, R.; Depristo, M.A. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Curr. Protoc. Bioinform.* **2013**, *43*, 11.10.11–11.10.33. [[CrossRef](#)] [[PubMed](#)]
38. Manheimer, K.B.; Patel, N.; Richter, F.; Gorham, J.; Sharp, A.J. Robust identification of deletions in exome and genome sequence data based on clustering of Mendelian errors. *Hum. Mutat.* **2018**, *39*, 870–881. [[CrossRef](#)]
39. Abyzov, A.; Urban, A.E.; Snyder, M.; Gerstein, M. CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* **2011**, *21*, 974–984. [[CrossRef](#)]
40. Tempel, S. Using and understanding RepeatMasker. *Methods Mol. Biol.* **2012**, *859*, 29–51. [[CrossRef](#)]
41. Roach, J.C.; Glusman, G.; Smit, A.F.; Huff, C.D.; Hubley, R.; Shannon, P.T.; Rowen, L.; Pant, K.P.; Goodman, N.; Bamshad, M.; et al. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* **2010**, *328*, 636–639. [[CrossRef](#)]
42. Koboldt, D.C.; Larson, D.E.; Wilson, R.K. Using VarScan 2 for Germline Variant Calling and Somatic Mutation Detection. *Curr. Protoc. Bioinform.* **2013**, *44*, 15.4.1–15.4.17. [[CrossRef](#)] [[PubMed](#)]
43. Shao, L.; Xing, F.; Xu, C.; Zhang, Q.; Che, J.; Wang, X.; Song, J.; Li, X.; Xiao, J.; Chen, L.L.; et al. Patterns of genome-wide allele-specific expression in hybrid rice and the implications on the genetic basis of heterosis. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 5653–5658. [[CrossRef](#)] [[PubMed](#)]
44. Do, K.T.; Kong, H.S.; Lee, J.H.; Lee, H.K.; Cho, B.W.; Kim, H.S.; Ahn, K.; Park, K.D. Genomic characterization of the Przewalski’s horse inhabiting Mongolian steppe by whole genome re-sequencing. *Livest. Sci.* **2014**, *167*, 86–91. [[CrossRef](#)]
45. Huang, J.; Zhao, Y.; Shiraigol, W.; Li, B.; Bai, D.; Ye, W.; Daidiikhoo, D.; Yang, L.; Jin, B.; Zhao, Q.; et al. Analysis of horse genomes provides insight into the diversification and adaptive evolution of karyotype. *Sci. Rep.* **2014**, *4*, 4958. [[CrossRef](#)]
46. Tatsumoto, S.; Go, Y.; Fukuta, K.; Noguchi, H.; Hayakawa, T.; Tomonaga, M.; Hirai, H.; Matsuzawa, T.; Agata, K.; Fujiyama, A. Direct estimation of de novo mutation rates in a chimpanzee parent-offspring trio by ultra-deep whole genome sequencing. *Sci. Rep.* **2017**, *7*, 13561. [[CrossRef](#)]
47. Campbell, C.D.; Chong, J.X.; Malig, M.; Ko, A.; Dumont, B.L.; Han, L.; Vives, L.; O’Roak, B.J.; Sudmant, P.H.; Shendure, J.; et al. Estimating the human mutation rate using autozygosity in a founder population. *Nat. Genet.* **2012**, *44*, 1277–1281. [[CrossRef](#)] [[PubMed](#)]
48. 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. [[CrossRef](#)]
49. Neefjes, J.; Jongsma, M.L.; Paul, P.; Bakke, O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat. Rev. Immunol.* **2011**, *11*, 823–836. [[CrossRef](#)]
50. Gibson, S.A.; Wei, Y.; Yan, Z.; Qin, H.; Benveniste, E.N. Ck2 controls th17 and regulatory T cell differentiation through inhibition of foxo1. *J. Immunol.* **2018**, *201*, 383–392. [[CrossRef](#)]

51. Rostamzadeh, D.; Yousefi, M.; Haghshenas, M.R.; Ahmadi, M.; Babaloo, Z. mTOR Signaling pathway as a master regulator of memory CD8 + T-cells, Th17, and NK cells development and their functional properties: ROSTAMZADEH et al. *J. Cell. Physiol.* **2019**, *234*, 12353–12368. [[CrossRef](#)] [[PubMed](#)]
52. Son, J.; Cho, Y.W.; Woo, Y.J.; Baek, Y.A.; Kim, E.J.; Cho, Y.; Kim, J.Y.; Kim, B.S.; Song, J.J.; Ha, S.J. Metabolic Reprogramming by the Excessive AMPK Activation Exacerbates Antigen-Specific Memory CD8(+) T Cell Differentiation after Acute Lymphocytic Choriomeningitis Virus Infection. *Immune Netw.* **2019**, *19*, e11. [[CrossRef](#)] [[PubMed](#)]
53. Abdullah, L.; Hills, L.B.; Winter, E.B.; Huang, Y.H. Diverse Roles of Akt in T cells. *Immunometabolism* **2021**, *3*, e210007. [[CrossRef](#)] [[PubMed](#)]
54. Koyama, H.; Bornfeldt, K.E.; Fukumoto, S.; Nishizawa, Y. Molecular pathways of cyclic nucleotide-induced inhibition of arterial smooth muscle cell proliferation. *J. Cell. Physiol.* **2001**, *186*, 1–10. [[CrossRef](#)] [[PubMed](#)]
55. Li, Z.; Pearlman, A.H.; Hsieh, P. DNA mismatch repair and the DNA damage response. *DNA Repair* **2016**, *38*, 94–101. [[CrossRef](#)]
56. Lavrik, O.I. PARPs' impact on base excision DNA repair. *DNA Repair* **2020**, *93*, 102911. [[CrossRef](#)]
57. Meek, D.W. Regulation of the p53 response and its relationship to cancer. *Biochem. J.* **2015**, *469*, 325–346. [[CrossRef](#)] [[PubMed](#)]
58. Nguyen, A.V.; Albers, C.G.; Holcombe, R.F. Differentiation of tubular and villous adenomas based on Wnt pathway-related gene expression profiles. *Int. J. Mol. Med.* **2010**, *26*, 121–125. [[CrossRef](#)] [[PubMed](#)]
59. Duncan, F.N. An attempt to produce mutations through hybridization. *Am. Nat.* **1915**, *49*, 575–582. [[CrossRef](#)]
60. Kovalchuk, I.; Kovalchuk, O.; Hohn, B. Genome-wide variation of the somatic mutation frequency in transgenic plants. *EMBO J.* **2000**, *19*, 4431–4438. [[CrossRef](#)]
61. Dal, G.M.; Erguner, B.; Sagiroglu, M.S.; Yuksel, B.; Onat, O.E.; Alkan, C.; Ozcelik, T. Early postzygotic mutations contribute to de novo variation in a healthy monozygotic twin pair. *J. Med. Genet.* **2014**, *51*, 455–459. [[CrossRef](#)] [[PubMed](#)]
62. Acuna-Hidalgo, R.; Bo, T.; Kwint, M.P.; van de Vorst, M.; Pinelli, M.; Veltman, J.A.; Hoischen, A.; Vissers, L.E.; Gilissen, C. Post-zygotic Point Mutations Are an Underrecognized Source of De Novo Genomic Variation. *Am. J. Hum. Genet.* **2015**, *97*, 67–74. [[CrossRef](#)] [[PubMed](#)]
63. Lujan, S.A.; Kunkel, T.A. Stability across the Whole Nuclear Genome in the Presence and Absence of DNA Mismatch Repair. *Cells* **2021**, *10*, 1224. [[CrossRef](#)] [[PubMed](#)]
64. Kamath, P.L.; Getz, W.M. Adaptive molecular evolution of the Major Histocompatibility Complex genes, DRA and DQA, in the genus Equus. *BMC Evol. Biol.* **2011**, *11*, 128. [[CrossRef](#)]
65. Liu, C.; Lei, H.; Ran, X.; Wang, J. Genetic variation and selection in the major histocompatibility complex Class II gene in the Guizhou pony. *PeerJ* **2020**, *8*, e9889. [[CrossRef](#)]
66. Radwan, J.; Babik, W.; Kaufman, J.; Lenz, T.L.; Winternitz, J. Advances in the Evolutionary Understanding of MHC Polymorphism. *Trends Genet. TIG* **2020**, *36*, 298–311. [[CrossRef](#)] [[PubMed](#)]
67. Atanasov, K.E.; Liu, C.; Erban, A.; Kopka, J.; Parker, J.E.; Alcázar, R. NLR Mutations Suppressing Immune Hybrid Incompatibility and Their Effects on Disease Resistance. *Plant Physiol.* **2018**, *177*, 1152–1169. [[CrossRef](#)]
68. Yang, X.; Mariuzza, R.A. Pre-T-cell receptor binds MHC: Implications for thymocyte signaling and selection. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 8166–8167. [[CrossRef](#)]
69. Okada, Y.; Wu, D.; Trynka, G.; Raj, T.; Terao, C.; Ikari, K.; Kochi, Y.; Ohmura, K.; Suzuki, A.; Yoshida, S.; et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **2014**, *506*, 376–381. [[CrossRef](#)]
70. Hu, X.; Deutsch, A.J.; Lenz, T.L.; Onengut-Gumuscu, S.; Han, B.; Chen, W.M.; Howson, J.; Todd, J.A.; Bakker, P.D.; Rich, S.S. Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. *Nat. Genet.* **2015**, *47*, 898–905. [[CrossRef](#)]
71. Selvaraja, M.; Too, C.L.; Tan, L.K.; Koay, B.T.; Abdullah, M.; Shah, A.M.; Arip, M.; Amin-Nordin, S. Human leucocyte antigens profiling in Malay female patients with systemic lupus erythematosus: Are we the same or different? *Lupus Sci. Med.* **2022**, *9*, e000554. [[CrossRef](#)] [[PubMed](#)]
72. Gaud, G.; Lesourne, R.; Love, P.E. Regulatory mechanisms in T cell receptor signalling. *Nat. Rev. Immunol.* **2018**, *18*, 485–497. [[CrossRef](#)]
73. Lebedeva, T.; Dustin, M.L.; Sykulev, Y. ICAM-1 co-stimulates target cells to facilitate antigen presentation. *Curr. Opin. Immunol.* **2005**, *17*, 251–258. [[CrossRef](#)] [[PubMed](#)]
74. Lee, E.B.; Kim, J.Y.; Kim, E.H.; Nam, J.H.; Park, K.S.; Song, Y.W. Intercellular adhesion molecule-1 polymorphisms in Korean patients with rheumatoid arthritis. *Tissue Antigens* **2004**, *64*, 473–477. [[CrossRef](#)]
75. Huang, G.; Wang, Y.; Chi, H. Regulation of TH17 cell differentiation by innate immune signals. *Cell. Mol. Immunol.* **2012**, *9*, 287–295. [[CrossRef](#)] [[PubMed](#)]
76. Silva, I.; Lima, C.; Monteiro, M.; Barboza, D.; Maia, M. IL1 β , IL18, NFKB1 and IFNG gene interactions are associated with severity of rheumatoid arthritis: A pilot study. *Autoimmunity* **2020**, *53*, 95–101. [[CrossRef](#)]
77. Wang, J.; Liu, H.; Wang, Y.; Wu, J.; Wang, C.; Liu, K.; Qin, Q. The Polymorphisms of Interleukin-12B Gene and Susceptibility to Inflammatory Bowel Diseases: A Meta-analysis and Trial Sequential Analysis. *Immunol. Investig.* **2021**, *50*, 987–1006. [[CrossRef](#)] [[PubMed](#)]
78. Osman, A.E.; Brema, I.; AlQurashi, A.; Al-Jurayyan, A.; Bradley, B.; Hamza, M.A. Single nucleotide polymorphism rs 2070874 at Interleukin-4 is associated with increased risk of type 1 diabetes mellitus independently of human leukocyte antigens. *Int. J. Immunopathol. Pharmacol.* **2022**, *36*, 3946320221090330. [[CrossRef](#)]

79. Rosenberg, E.; Demopoulos, R.I.; Zeleniuch-Jacquotte, A.; Yee, H.; Sorich, J.; Speyer, J.L.; Newcomb, E.W. Expression of cell cycle regulators p57KIP2, cyclin D1, and cyclin E in epithelial ovarian tumors and survival. *Hum. Pathol.* **2001**, *32*, 808–813. [[CrossRef](#)]
80. Dang, T.T.; Morales, J.C. Involvement of POLA2 in Double Strand Break Repair and Genotoxic Stress. *Int. J. Mol. Sci.* **2020**, *21*, 4245. [[CrossRef](#)]
81. Shiratori, A.; Okumura, K.; Nogami, M.; Taguchi, H.; Onozaki, T.; Inoue, T.; Ando, T.; Shibata, T.; Izumi, M.; Miyazawa, H. Assignment of the 49-kDa (PRIM1) and 58-kDa (PRIM2A and PRIM2B) Subunit Genes of the Human DNA Primase to Chromosome Bands 1q44 and 6p11.1-p12. *Genomics* **1995**, *28*, 350–353. [[CrossRef](#)] [[PubMed](#)]
82. Venkatesan, R.N.; Treuting, P.M.; Fuller, E.D.; Goldsby, R.E.; Norwood, T.H.; Gooley, T.A.; Ladiges, W.C.; Preston, B.D.; Loeb, L.A. Mutation at the Polymerase Active Site of Mouse DNA Polymerase Increases Genomic Instability and Accelerates Tumorigenesis. *Mol. Cell. Biol.* **2007**, *27*, 7669–7682. [[CrossRef](#)] [[PubMed](#)]
83. Zanders, S.; Ma, X.; Roychoudhury, A.; Hernandez, R.D.; Demogines, A.; Barker, B.; Gu, Z.; Bustamante, C.D.; Alani, E. Detection of Heterozygous Mutations in the Genome of Mismatch Repair Defective Diploid Yeast Using a Bayesian Approach. *Genetics* **2010**, *186*, 493–503. [[CrossRef](#)]
84. Magon, K.L.; Parish, J.L. From infection to cancer: How DNA tumour viruses alter host cell central carbon and lipid metabolism. *Open Biol.* **2021**, *11*, 210004. [[CrossRef](#)]
85. Ali Syeda, Z.; Langden, S.S.S.; Munkhzul, C.; Lee, M.; Song, S.J. Regulatory Mechanism of MicroRNA Expression in Cancer. *Int. J. Mol. Sci.* **2020**, *21*, 1723. [[CrossRef](#)] [[PubMed](#)]
86. Liaw, D.; Marsh, D.J.; Li, J.; Dahia, P.L.; Wang, S.I.; Zheng, Z.; Bose, S.; Call, K.M.; Tsou, H.C.; Peacocke, M.; et al. Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome. *Nat. Genet.* **1997**, *16*, 64–67. [[CrossRef](#)]
87. Sansom, O.J.; Meniel, V.; Wilkins, J.A.; Cole, A.M.; Oien, K.A.; Marsh, V.; Jamieson, T.J.; Guerra, C.; Ashton, G.H.; Barbacid, M.; et al. Loss of Apc allows phenotypic manifestation of the transforming properties of an endogenous K-ras oncogene in vivo. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 14122–14127. [[CrossRef](#)]
88. Murugan, A.K.; Grieco, M.; Tsuchida, N. RAS mutations in human cancers: Roles in precision medicine. *Semin. Cancer Biol.* **2019**, *59*, 23–35. [[CrossRef](#)]
89. Lee, T.I.; Young, R.A. Transcriptional regulation and its misregulation in disease. *Cell* **2013**, *152*, 1237–1251. [[CrossRef](#)]
90. Wei, J.; Hu, M.; Huang, K.; Lin, S.; Du, H. Roles of Proteoglycans and Glycosaminoglycans in Cancer Development and Progression. *Int. J. Mol. Sci.* **2020**, *21*, 5983. [[CrossRef](#)]
91. Glunde, K.; Bhujwala, Z.M.; Ronen, S.M. Choline metabolism in malignant transformation. *Nat. Rev. Cancer* **2011**, *11*, 835–848. [[CrossRef](#)] [[PubMed](#)]
92. Wang, Y.; Chen, Y.; Fang, J. Post-Transcriptional and Post-translational Regulation of Central Carbon Metabolic Enzymes in Cancer. *Anti-Cancer Agents Med. Chem.* **2017**, *17*, 1456–1465. [[CrossRef](#)] [[PubMed](#)]
93. Oliveira, P.A.; Colaco, A.; Chaves, R.; Guedes-Pinto, H.; De-La-Cruz, P.L.; Lopes, C. Chemical carcinogenesis. *An. Da Acad. Bras. De Cienc.* **2007**, *79*, 593–616. [[CrossRef](#)] [[PubMed](#)]