SI Appendix

SI methods.

Conceptus recovery and dissection. Equine conceptuses from pregnant horse mares and jenny donkeys were collected on days 33-35 post ovulation by non-surgical uterine lavage using sterile Phosphate Buffered Saline (PBS) and a large bore catheter (1). After collection, the conceptuses were micro-dissected into distinct tissues (chorionic girdle, allantochorion, fetus, etc.) with the aid of a dissection microscope and ophthalmic instruments. Tissues were either snap frozen in liquid nitrogen and stored at -80^{o}C, or placed into tissue culture vessels and cultured under sterile conditions. Animal care was performed in accord with the guidelines set forth by the Institutional Animal Care and Use Committee of Cornell University under protocol #1986-0216. The invasive trophoblast cells of the equine chorionic girdle (CG) begin to form around day 25 of gestation and proliferate rapidly for 10 days before invading the endometrium at day 36-38, where they terminally differentiate and secrete equine chorionic gonadotrophin (2). On gestational days 33-34 complete equine conceptuses can be readily recovered by non-surgical uterine lavage, and the chorionic girdle tissue can be easily dissected and isolated free of maternal and fetal cell contamination (2).

Illumina mRNA sequencing of horse, donkey, mule and hinny transcriptomes. Our initial mRNA-Seq was performed on total RNA samples from one horse, one donkey, one mule and one hinny day 33-34 chorionic girdle sample (animal IDs: horse3879, donkey3689, mule3702 and hinny3703) using an Illumina Genome Analyzer (Illumina Inc., CA). The mRNA-Seq libraries were made with 3 μg of starting total RNA samples using the mRNA-Seq 8-Sample Prep Kit (Illumina Inc., CA), following the Illumina protocol for mRNA sequencing sample preparation. 20 Illumina GA lanes were sequenced for the horse library and 10 lanes for the donkey library. We also did 8 lanes each for the mule and hinny libraries. Image analysis and base calling were performed by the Illumina instrument software (Illumina pipeline v1.3). In total, we obtained 82.5 million short reads (read length 44 bp) for horse sample, 53.4 million for donkey, 68.4 million for mule and 58.7 million for the hinny sample. To check the imprinting status in fetus, we performed one Illumina GAIIx lane each for two mule fetus and one hinny fetus samples. The libraries were made from 6 μg of starting total RNA. 25-32 million reads were obtained for these fetus samples.

mRNA-Seq Alignment and quantification of total and allele-specific expression. Illumina sequencing reads were truncated to 40 bp and any reads containing one or more read position with Q-score less than 3 were filtered out. The reads were then aligned to the horse reference genome (equcab2, http://genome.ucsc.edu/) using BWA with a maximum of 5 mismatches (3). On average, 67.1% of the reads were mapped to exon regions in the reference genome. To identify reads that mapped to the exon-intron junctions, we built a junction database by extracting all possible junction sequences, based on the gene and exon models from the Ensembl database (www.ensembl.org). 5.7% of the total reads were mapped to the exon-intron junctions. The exon and junction alignment counts were summarized by custom scripts. Counts were normalized by the transcript length and the total number of mapped reads to compute RPKM (4). There were 10,937 autosomal Ensembl transcripts with RPKM $\geqslant 1$ in horse, donkey, mule and hinny chorionic girdle samples. In 3 mule and hinny fetus samples, 13,650 autosomal Ensembl transcripts were covered with RPKM $\geqslant 1$.

SNP information is needed to quantify the allelic expression from the two parental alleles in the hybrids. Specifically, we need informative SNPs which are homozygotes in both horse and donkey parents but different from each other (*i.e.* horse vs, donkey fixed differences). We performed de novo SNP calling in horse and donkey from the uniquely mapped reads using both Maq and SAMtools software (5). Besides horse 3879 and donkey 3689, RNA-Seq data on 6 additional horse and 4 additional donkey chorionic girdle samples were used to determine nucleotide sequence differences at sites that appeared monomorphic within each species. We called 48,125 fixed differences between horse and donkey chorionic girdle transcriptome in 7 horses and 5 donkey samples, 44,916 (93.3%) of which reside in known and predicted gene models. For the fetal samples, we do not have parental RNA-Seq data (horse and donkey). Additional fetus SNPs were called from the mule and hinny fetus data.

Exonic single nucleotide differences between horse and donkey were used to quantify allelic expression ratios. The exonic SNPs in horse transcripts were determined by two different sets of gene models: the Ensembl and the RefSeq gene models. The Ensembl horse gene model (v59) contains 29,159 transcripts and 26,954 genes. The horse RefSeq database has 18,446 RefSeq genes. Most of the gene models are predicted gene models (with XM and XR RefSeq IDs), therefore

most lack the 3' and 5' UTRs. Because the SNP density is higher in the UTR region due to relatively low level of evolutionary constraint, we need to cover SNPs in UTRs for allelic expression ratio quantification. To solve this problem, we generated an extended RefSeq database, based on conservation from other vertebrate species. We BLATed the RefSeq genes from mouse, human, dog and chicken to the horse reference genome, and selected the transcription start and end position in the horse genome from the longest mapped orthologous RefSeqs. In addition, we manually annotated the horse orthologs to mouse and human known imprinted genes.

To quantify the allele-specific expression in mule and hinny, at each identified SNP position we counted the reads with the reference/horse allele as well as reads with the alternative/donkey allele (6). In hybrids, since the RNA-Seq reads are only mapped to the horse genome, there will be genome mapping bias toward the horse allele if we use the same cut-off for both reads coming from the horse and donkey alleles. To remove this mapping bias, we generated a pseudo-genome, by replacing the reference allele in the horse genome with the alternative allele. Then we realigned the reads with the same cut-off to the pseudo-genome. The averaged counts from the reference and pseudo-genome were used as the final SNP count summary. Finally, the allelic expression ratio was quantified as the percentage of horse alleles at each SNP position. The allelic expression ratios were calculated on a per-gene basis by summarizing all informative SNP positions in the same transcript. In total, we covered 44,916 high quality autosomal SNPs with 4 or more counts in both mule and hinny chorionic girdle samples.

Detection of significant parent-of-origin effects. With the informative SNPs and the SNP counts, we were able to determine the allele-specific expression ratio by the relative counts from the reference and alternative alleles (6). We define p_1 as the expression percentage from the horse allele in mule and p_2 as the horse allele percentage in hinny. In regard to the direction of transmission, p_1 is the maternal allele percentage in mule because mule has a horse mother, and p_2 is the paternal percentage for hinny. For a non-imprinted gene with 50%:50% expression ratio in both mule and hinny, $p_1 = p_2 = 0.5$ ($p_2 - p_1 = 0$). For an imprinted gene with strictly paternal expression, we expect $p_1 = 0$ and $p_2 = 1$ ($p_2 - p_1 = 1$). To quantify the degree of genomic imprinting, we propose $p_2 - p_1$ as a measurement of the parent-or-origin effect, ranging from -1 (100% maternal expressed imprinted gene), to 0 (non-imprinted genes), to +1 (100% paternal expressed imprinted gene). The Storer-Kim test (7) was used as a formal statistical test of the null hypothesis that $(p_2 - p_1) = 0$. Rejections of this null hypothesis identify novel imprinted candidate genes. To include the significant partially imprinted candidates, we used an arbitrary cut-off of $p_1 > 0.65$ and $p_2 < 0.35$ for maternally expressed candidates, and $p_1 < 0.35$ and $p_2 > 0.65$ for paternally expressed ones.

Out of the 6,965 unique autosomal genes covered with one or more informative SNPs and SNP counts $\geqslant 10$ in both mule and hinny chorionic girdle transcriptome, with the above criteria, we found 93 candidates with q-value < 0.01 (*SI Appendix*, Table S2). Of these candidates, 40 have preferential maternal expression, and 53 have a paternally biased expression. To visualize the allelic expression ratio and the degree of parent-or-origin effect genome-wide, we made a plot for each autosome, and chromosome 10 is shown in Fig. 1C as an example. From these figures, we observed that most of the genes show nearly 50:50 allelic expression ratios. A number of significant candidate imprinted genes emerged from the parent-of-origin effect plot.

cDNA synthesis for Sanger and Pyrosequencing. First strand cDNA synthesis was carried out on one microgram (1 µg) of total RNA using M-MLV Reverse Transcriptase (USB, Cleveland, OH) in a final volume of 100 µl. cDNA quality was assayed by Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR) using primers to amplify within the coding sequence of the equine Beta-2 Microglobulin gene. Primers for the B2M PCR are positioned in neighboring exons separated by a 600 bp intron. All samples demonstrated a single strong 250 bp band in this assay, demonstrating lack of gDNA contamination. Equine genomic DNA was included in each assay in a separate well and gave the predicted 850 bp band.

Genotyping the SNPs in horse and donkey by Sanger sequencing. To genotype the SNPs in the hybrids and the horse-donkey parents, we designed PCR and sequencing primers for the candidate genes using primer3 (http://frodo.wi.mit.edu/primer3/). 25 µl PCR reactions were carried out using recombinant Taq DNA polymerase (Invitrogen) under the following cycling conditions: 1 cycle of 95°C for 10 min, 33 cycles of 95°C-30 sec, 58°C-30 sec, 72°C-1 min, followed by 1 cycle of 72°C for 10 min. The PCR products were purified by Exonuclease I and Shrimp Alkaline Phosphatase (USB from Affymetrix, CA) and sequenced bidirectionally on an ABI 3730xl DNA analyzer (Life

Technologies, CA) with BigDye Terminator v3.1. The sequence chromatograms were analyzed with CodonCode Aligner version 3.7.1 (CodonCode Corporation Software for DNA Sequencing).

Tissue Culture of placental tissue. Chorionic girdle cells were gently flaked off the basement membrane into cold PBS containing 200 U/ml penicillin and 200 µl/ml streptomycin using a #15 scalpel blade. PBS containing small clumps of cells was transferred into a 15 ml conical tube and placed on ice for 5 min, allowing cells to settle into a loose pellet. PBS was removed and cells were resuspended in culture medium (DMEM + 10% FBS, 100 U/ml penicillin, 100µg/ml streptomycin, 2 mM L-glutamine, 0.4 µg/ml insulin and 0.5 µg/ml ascorbic acid). Cells were plated in 15 x 100 mM petri dishes coated with 0.5% gelatin in PBS for 30 min at 37°C followed by a PBS rinse. Cultures were incubated at 37°C with 8% CO_2 and monitored daily for growth. Twice weekly ½ volume media was removed and replaced with fresh. When cells reached confluency they were removed from dishes by gentle scraping and passed into flasks or plates to decrease density. Cultures were maintained in this manner for 5 passages (33 days) at which time cells were collected, pelleted by centrifugation at 120 g, 5 min, 4°C, resuspended in DMEM + 10% FBS with 5% DMSO and gently transferred to cryovials at a concentration of 1.5×10^6 cells per vial. Vials were frozen at a rate of 1 degree per min to -80°C, and then transferred to liquid nitrogen for storage. To revive cells; one vial was rapidly thawed in a 37°C water bath, transferred to a conical tube containing 10 ml DMEM + 10% FBS and centrifuged as described above. Medium was discarded. Cells were resuspended in culture medium and plated in gelatin coated dishes. Fibroblast cultures were started by mincing the fetal tail with a razor blade in 1 ml trypsin-EDTA. Tissue was tritrated several times with a 1 ml pipette tip and transferred to a conical tube to incubate for 1 min at room temperature. 10 ml DMEM + 10% FBS was added to stop enzymatic activity. Cells were pelleted by centrifugation and medium was removed. Cells were resuspended in culture medium and plated in gelatin coated dishes. Cultures were maintained as chorionic girdle cell cultures described above, but cells were not frozen and revived.

Equine 44K Element Gene Expression Microarray. A 44,000 feature equine microarray (Agilent Technologies, Santa Clara, CA) was used to investigate gene expression in matching chorionic girdle samples from horse, donkey, mule, and hinny. 1ug of total RNA was linearly amplified and Cy3 labeled using the Ambion Amino Allyl MessageAmp™ II kit (Applied Biosystems, Foster City, CA), then hybridized to array slides using standard techniques. Experiments were analyzed with GeneSpring GX10 software (Agilent).

SI Text

Text S1. Verification of novel imprinted genes in mule and hinny CG samples.

To confirm the imprinting status of the 93 candidate imprinted genes we discovered, we performed independent gene-by-gene verification experiments using allele-specific pyrosequencing, on the top 40 candidates ranked by *q*-value and some additional genes further down the list. Among the 40 top candidates, we selected 22 for pyrosequencing (the candidate genes in numtDNA and the ones with X chromosome homology are excluded). 21/22 were verified to be imprinted in mule and hinny by pyrosequencing.

Candidate genes in numtDNA: From the 93 candidate imprinted genes we identified in mule and hinny CG samples, four genes, *ENSECAG00000016536*, *ENSECAG00000016730*, *CSMD1* and *NU1M* show 100% maternal expression (*SI Appendix*, Table S2), but they have >99% sequencing homology to the mitochondrial DNA. These predicted genes are located in the insertions of the mitochondrial DNA (mtDNA) in the nuclear genome, which is numtDNA (8). Due to the high sequencing identity, we cannot determine whether reads mapped to these regions are from the nuclear genome or the mtDNA. To determine how many numtDNA regions there are in the horse genome, we BLATed the horse mtDNA to the horse genome. With a cut-off of BLAT score >100, we detect 44 mtDNA insertions in the horse genome. Because mtDNA is circular, we joined matched regions head to tail and reduced the number of hits to 40. Then we searched the nearby candidate imprinted genes and found that 16/93 candidates are within +/- 5 Mbp of the numtDNAs. Among the verified candidate imprinted genes, 16 are orthologs to known imprinted genes in mouse and/or human. 6/16 are within 2 Mbp (9/16 are within 10 Mbp) of the numtDNA insertions. mtDNA insertion sequences may affect genomic imprinting in horse.

Candidate genes on chrUn: Six candidate imprinted genes are on chrUn (unmapped scaffolds), five of which (*MAGED2*, *TSR2*, *MAGED1*, *GNL3L* and *PHF8*) show 100% maternal expression (*SI Appendix*, Table S2). All five genes have orthologs on human X chromosome. We further investigated these genes in the horse genome assembly and we found *MAGED2*, *TSR2*, *GNL3L* and *PHF8* are on scaffold Un0004 and *MAGED1* is on scaffold Un0019. All SNPs in these 2 scaffolds show 100% maternal expression in male hybrids. These two scaffolds might be on horse X, so we expect 100% maternal expression in males.

Candidate genes that are orthologous to known imprinted genes in human and/or mouse: Among the candidate imprinted genes we found in mule and hinny CG RNA-Seq data, 16 have an ortholog known to be imprinted in mouse and/or human (*SI Appendix*, Table S3). To confirm the imprinting status in mule and hinny, we used an independent method, allele-specific pyrosequencing, to verify them (9) (*SI Appendix*, SI Methods). In interspecific hybrids, differential allelic expression could be due to genome imprinting or random monoallelic expression (10). To exclude the possibility of stochastic monoallelic expression and confirm this is a parent-of-origin effect, we verified the candidates in multiple mule and hinny individuals. We also did Sanger sequencing on the parental gDNA from horse and donkey to confirm the SNP is homozygote in the parents.

5 of the 16 orthologs to known imprinted genes in mouse and human (please note they are all novel imprinted genes in equids) show preferential maternal expression. A recent study suggests that most of genes previously identified as imprinted and maternally expressed in human and mouse placenta are due to an artifact of maternal contamination (11). In our study, the samples are from the preimplantation stage, so there is zero maternal contamination and all five genes identified in our RNA-Seq data were verified to be imprinted in mule and hinny CG. *H19*, *PHLDA2* and *MEG3* have 100% expression from the maternal allele in both F1s. *IGF2R* and *NAP1L4* are partially imprinted genes with preferential expression from the mother (*SI Appendix*, Figs. S2-S6). The pyrosequencing results are consistent with the RNA-Seq data. The paternally expressed candidates, *IGF2*, *INS-IGF2*, *PEG10*, *MEST*, *PEG3*, *SNRPN*, *DLK1*, *NDN*, *PAR-SN* and *SGCE* all show 100% paternal expression. We confirmed these findings by allele-specific pyrosequencing (*SI Appendix*, Figs. S7-S13).

COMMD1 is a known imprinted gene in mouse with preferential maternal expression in adult brain and other tissues (12). It is reported to be not imprinted in human (13). Here, in our CG RNA-Seq data in mule and hinny, the SNP allele counts suggest preferential paternal expression (with expression ratio 35:65), which is the opposite direction in mouse. However,

the evidence is not as strong as other known imprinted genes. First, there are two SNPs in the gene region and only one is significant. Second, the *q*-value ranking is the last among the 16 known imprinted genes. We targeted the non-significant SNP, and the pyrosequencing results show biallelic expression in both mule and hinny, so *COMMD1* is not imprinted in D33 mule and hinny CG samples.

Candidate genes that are not known to be imprinted in any other species: using allele-specific pyrosequencing method, we verified six novel imprinted candidates that are not known to be imprinted in any other species.

-***LY6G6C***. *LY6G6C* (lymphocyte antigen 6 complex G6C, horse RefSeq predicted gene ID *XM_001917750*) is on chromosome 20 and is one of the two novel maternally expressed imprinted genes we discovered in our RNA-Seq data. It is a member of the LY6 superfamily (14). The LY6 family members are attached to the cell surface by a glycosylphosphatidylinositol (GPI) anchor which is directly involved in signal transduction. There is a retrotransposed gene in the horse genome, *LOC100053149* (*LY6G6C* like), which is on chromosome 16, with RefSeq ID XM_001915930. *LY6G6C* is NOT known to be imprinted in mouse or human or any other species. There are three informative SNPs covered in the Illumina RNA-Seq data and all 3 are significant (Table 2). We confirmed that this is a maternally expressed partially imprinted gene in 5 out of 5 mule/hinny individuals by pyrosequencing (*SI Appendix*, Fig. S14). *LY6G6C* is not in synteny with any known IP clusters. It is a novel imprinted gene in mule and hinny.

-***CFH***. *CFH* is another maternally expressed candidate imprinted gene. The *q*-value ranking is 54 (Table 2). We selected this gene for verification because it shows 100% maternal expression in RNA-Seq data. We checked the parental genotype of six mules and hinnies, and we found that the SNP we discovered in the RNA-Seq data is only informative in mules. However, there is a novel SNP adjacent to the original one which was missed because of the 100% maternal expression in mule and hinnies. The novel SNP is informative in the two hinnies. Pyrosequencing results confirmed 100% maternal expression in all six mule/hinnies we tested (*SI Appendix*, Fig. S15). *CFH* is complement factor H, which helps the complement system in immune defense. It has been reported to be associated with age-related macular degeneration (AMD) (15). It is not known to be imprinted in any other species.

-***HAT1***. Among the genes without mtDNA homology or X-linked contig location, *HAT1* is the highest ranking (13th) candidate imprinted gene that is not known to be imprinted in other species (Table 2) and is clearly biallelically expressed in mouse placenta. *HAT1* is histone acetyltransferase 1, which acetylates soluble histone H4 in the cytoplasm at Lys-5 and Lys-12 positions (16). *HAT1* is the first identified imprinted gene that is directly involved epigenetic modifications. *HAT1* is a conserved gene, and we only found two fixed nucleotide differences between horse and donkey. One in 5'-UTR and the other is in the 3'-UTR region. The RNA-Seq data show strong paternal expression, with >95% from the father in mule and 100% parental expression in hinny (Fig. 2A). We confirmed the SNP is homozygous in the parents by Sanger sequencing (Fig. 2B). We then verified the imprinted status in six different mule and hinny individuals by pyrosequencing, and they are all imprinted (Fig. 2C, D).

-***INSR***. *INSR* is a novel candidate imprinted gene that is not known to be imprinted in mouse or human or any other species. It is not in synteny with any known imprinted clusters either. There are 17 informative SNPs in the Illumina sequencing data and all are significant. In the RNA-Seq samples mule1 and hinny1, we observed asymmetry of the allelic expression ratio. The RNA-Seq data show 80% paternal expression from the donkey allele in mule1, with 20% leakage from the horse allele, whereas in hinny1 there is 100% paternal expression. Pyrosequencing results confirm the inferences from the Illumina RNA-Seq data (*SI Appendix*, Fig. S17). For the four additional verification samples, *INSR* is imprinted in mule4 and hinny2, with preferential paternal expression. Interestingly, in mule2 and mule3, we observed biallelic expression for *INSR*. Such variable imprinting status among individuals has been found in human placenta (17), but not in the inbred mouse studies. The mouse imprinted genes show stable imprinting status across individuals, partly because they are the same genetic background. To check whether the inter-individual variability of the imprinting status is also present within horse, or due to aberrant genomic imprinting in the hybrids, we need additional horse CG samples with known allelic transmission. *INSR* encodes insulin receptor, which can bind to insulin to stimulate glucose uptake. Among known imprinted genes, there is one famous ligand-receptor pair with opposite imprinting direction, which is *IGF2*. *IGF2* displays 100% paternal expression and *IGF2R* exhibits preferential maternal expression. Here, we discovered a second case, *INS* and *INSR*, both of which are imprinted in mule and hinny placenta but with the same direction (paternal expression). Defects in *INSR* can cause insulin resistance and noninsulin-dependent diabetes mellitus (NIDDM)

(OMIM number 125853) and a number of other human diseases, such as Rabson-Mendenhall syndrome (RMS), leprechaunism (LEPRCH) and familial hyperinsulinemic hypoglycemia type 5 (HHF5).

-***D7ERTD715E***. *D7ERTD715E* is the horse ortholog to the mouse *D7ertd715e* transcript (mouse RefSeq_ID NR_015456). *D7ertd715e* is a noncoding transcript without any CDS. It is located in a known imprinted gene cluster, about 10 kb downstream the mouse *Snrpn* gene. In the horse genome, *D7ERTD715E* is 15 kb downstream the *SNRPN* gene, which is also imprinted in our horse CG samples. We detect 100% paternal expression from *D7ERTD715E* in our RNA-Seq data. We verified the imprinting status in 6 mule/hinny individuals, and we observed consistent 100% paternal expression (*SI Appendix*, Fig. S18).

-***STON1***. *STON1* is a candidate imprinted gene with paternal expression in our RNA-Seq data. We tested in 6 mule and hinny individuals by pyrosequencing and 4 are consistent with preferential paternal expression (*SI Appendix*, Fig. S19).

Text S2. Identification and verification of imprinted genes in mule and hinny fetus by RNA-Seq and pyrosequencing.

To identify potential parent-of-origin effects in other tissues, we performed Illumina RNA-Seq on day 33-34 fetus samples (*SI Appendix*, SI Methods). Allelic expression ratios were quantified in both mule and hinny and the degree of parent-of-origin effect were calculated. 14 of the top 20 candidates are known to be imprinted in human or mouse, and the other 6 are either on potential X-linked scaffolds or mtDNA insertions. In total, we found 16 known imprinted genes: *H19*, *MEG3* and *IGF2R* with maternal expression; *NDN*, *SNRPN*, *PEG3*, *IGF2*, *INS-IGF2*, *NNAT*, *PEG10*, *MEST*, *ZIM2*, *DLK1*, *MAGEL2*, *DIRAS3* and *PON2* with preferential paternal expression. The imprinting status of all known genes matches the mouse and human status, suggesting conservation of the direction of these known imprinted genes. We verified the 16 known imprinted genes in 6 different mule and hinny individuals by allele-specific pyrosequencing and all of them are imprinted in mule and hinny fetus (Table 1). In addition, we tested four genes that are imprinted in CG but not covered in fetus (*PHLDA2*, *SGCE*, *PAR-SN* and *NAP1L4*) by pyrosequencing, and they are all imprinted in fetus with low total expression level.

We discovered 89 novel imprinting candidates in fetus, but most of them have inconsistent SNPs or low SNP coverage. We selected and tested three of them (*HBB*, *NRM* and *SSX2IP*) by pyrosequencing, and they are not imprinted. There are several reasons why we have a lower verification rate for the novel imprinted candidates in fetus. First, the novel candidates have low SNP coverage and higher *q*-value. All top 20 *q*-value ranking genes are known imprinted genes, and we have 100% verification rate for those genes. Second, for the between horse-donkey SNPs we used for the fetus transcriptome, we do not know whether they are homozygous in both parents. Heterozygous SNPs in horse or donkey will generate false positives. Third, unlike the CG sample, which consists of pure trophoblast cells, the fetus is a mixture of many different tissues, so the tissue-specific imprinted genes will be averaged out. Also, the total number of expressed genes is higher in the fetus samples. Therefore, we need more coverage to detect all possible candidates in the fetus.We rely on the high coverage genes in the fetus RNA-Seq data, most of which are known to be imprinted in human or mouse.

We found a total of 16 known imprinted genes in our mule and hinny CG samples (Table 1). In the fetus, we discovered three additional known imprinted genes, *NNAT*, *MAGEL2* and *DIRAS3*, which are not covered in the CG RNA-Seq data (*ZIM2* is also consistent with genomic imprinting, but its significant SNP overlaps with *PEG3* SNPs). 14 of the 16 imprinted genes in the CG are also imprinted in fetus, although some of them are expressed at different levels in CG vs. fetus. Two known imprinted genes show interesting tissue-specific imprinting patterns. *PHLDA2* is imprinted in CG with 100% maternal expression in both mule and hinny. However, in fetus, it is partially imprinting with 60% expression from the mother in mule and 85% from the mother in hinny (Table 1). *NAP1L4* is imprinted in CG with preferentially maternal expression, but it is not imprinted in fetus (Table 1). So for known imprinted genes in the fetus and CG, we observed both differences in imprinting status and differences in the degree of parent-of-origin effect. Overall, the imprinting status of most known imprinted genes is conserved between the two tissues.

For the 10 verified novel imprinted genes in CG samples, we checked their imprinting status in fetus, and only one is imprinted in fetus. *D7ERTD715E* is a non-coding transcript near the *SNRPN-SNURF* imprinting cluster. Its imprinting status in fetus is not surprising because the genes in the *SNRPN-SNURF* cluster (*SNRPN*, *PAR-SN* and *NDN*) are also imprinted in fetus. For the other 9 verified novel imprinted genes, three are either not expressed or have very low expression level in fetus. The remaining six are not imprinted in fetus (Table 1). The dramatic degree of tissue-specificity of the novel imprinted genes we discovered is a novel finding compared to the known imprinted genes, most of which are expressed and imprinted in both tissues. This could be one possible reason why these imprinted genes had not yet been discovered in mouse or human.

Text S3. Methylation profiling of Differentially Methylated Regions for known and novel imprinted genes.

The mechanism of genomic imprinting is not fully understood, but allele-specific differential epigenetic modifications, including differential DNA methylation, clearly play a role in silencing the inactive allele (18, 19). The imprinting regulatory element, known as the Differentially Methylated Regions (DMR), was first discovered in mouse (20, 21). To date, about one third of the mouse imprinted genes have a known DMR associated with them. If the DMR is located in the promoter region of an imprinted gene, the methylated allele will be inactive. If the DMR is located further upstream and regulates the differential allelic expression through an enhancer blocker mechanism (via CTCF or other proteins), the methylated allele will prevent the binding of CTCF, allowing the enhancer region to be folded close to promoter region (22). Thus, the methylated allele is active. Although the DMR mechanism is well studied in mouse, the horse and mouse lineages diverged about 70 million years ago. We would like to ask whether the DMR CpG islands are still present in the horse lineage, and if the DMR is present, whether the function (differential methylation pattern) is conserved.

To do this, we selected two well-known imprinted loci, *H19* and *PEG3*. *H19* is a maternally expressed imprinted gene at the *Igf2-H19* imprinting cluster in mouse. There is a DMR 2 kb upstream of *H19* and 90 kb downstream from *IGF2*. The paternal allele is methylated at the *H19* DMR. Methylation at the paternal allele at this DMR blocks CTCF binding to the *Igf2* paternal allele, allowing the *Igf2* gene to access the shared enhancers and repress the *H19* expression from the paternal allele (23, 24). In the horse genome, the *H19* DMR is present and located 2.5 kb upstream from the horse *H19* gene. We discovered in this study that *H19* is imprinted with 100% maternal expression in mule and hinny CG samples. If the DMR function and imprinting mechanism are conversed between mouse and horse, then we expect differential methylation at the horse *H19* DMR. To check this, we performed bisulfite sequencing targeting the *H19* DMR region in horse, donkey, mule and hinny CG gDNA samples, as well as in horse sperm. We found that in horse CG sample across all 14 CpG sites we tested in the *H19* DMR, about half of the alleles are fully methylated and the other half are fully unmethylated at all CpGs in the DMR (Fig. 3A). Therefore, we concluded that the horse DMR is differentially methylated in CG sample. To check whether the differential methylation is parent-specific, we examined the horse sperm gDNA and found it to be 100% methylated. Since most of the imprinted genes are not subject to the wave of de- and re-methylation after fertilization during epigenetic reprogramming (25), we infer that the paternal allele is methylated and maternal allele is unmethylated. This is consistent with what is known in mouse. We also checked the DMR methylation profile in donkey CG samples, and it is also differentially methylated. The advantage of having both reciprocal F1s and their parental species is that we can directly quantify the allele-specific methylation in the DMR between the two parental species, provided there is a SNP nearby. This is exactly the case for *H19*. At CpG site #6, there is a G/A SNP between horse and donkey. At this nucleotide position, horse has genotype G/G and donkey has genotype A/A (Fig. 3A). Both mule and hinny are G/A heterozygotes at this position, but A is transmitted from the donkey father in mule, and the horse father in hinny. We could infer the paternal and maternal allele transmission for our methylation date in mule and hinny. From the methylation profile, we observed that the paternal alleles are methylated in both mule and hinny CG samples (Fig. 3A).

We found that both the DMR and the differential methylation status are conserved in horse for a maternally-expressed, paternally-methylated gene (*H19*). To be complete, we also checked a paternally expressed imprinted gene, with the maternal allele methylated in mouse. *PEG3* is a paternally expressed known imprinted gene in mouse, and a CpG island spanning the 5' upstream and the first exon has been shown to be differentially methylated (26). In our mule and hinny RNA-Seq data, we discovered that *PEG3* is imprinted with 100% paternal expression. We found a CpG islands with 24 CpG sites in horse about 550 bp upstream the gene. Bisulfite sequencing results show that the horse CG CpG island is differentially methylated and horse sperm is 100% unmethylated (Fig. 3B). The maternal allele-specific methylation is consistent with the paternal allelic expression we observed. The donkey CpG island in CG sample is also differentially methylated. There is a T/G SNP between horse and donkey in *PEG3* DMR (CpG site #21), which is homozygous T/T in horse, homozygous G/G in donkey, and heterozygous T/G in both mule and hinny. We are able to quantify allele-specific methylation in mule and hinny. Both mule and hinny CG DMR are differentially methylated, with paternal-specific methylation (Fig. 3B). Therefore, the differential methylation status is conserved in hybrids as well.

Although the sequence divergence between horse and donkey is relatively low, we found sequence differences between horse and donkey DMR for both *H19* and *PEG3*. These sequence differences were used for detecting allele transmission direction in mule and hinny. We further examined the sequence and found there are turnovers of CpG sites between horse and donkey. In *H19* DMR, in a CpG context (CpG site #6), there is a G (horse) to A (donkey) substitution, abolishing the CpG site in donkey (CpA in donkey). Because most of the time only cytosines in the context of a CpG could be methylated in differentiated cell in mammals, we expect 100% unmethylation in donkey for this CpG site. This is exactly what we saw in our data (Fig. 3A). In the hybrids, the hinny has a horse father and a donkey mother. For CpG site #6, because the donkey mother is unmethylated (CpA) and the horse father is methylated (paternal methylation), we expect a differential methylation pattern. However, in the mule sample with a donkey father and a horse mother, since both the horse mother (paternal methylation) and donkey father (CpA) are unmethylated, we expect 100% unmethylation. The bisulfite sequencing results for this site are entirely consistent with these expectations (Fig. 3A). In the *PEG3* DMR, we have the opposite situation: a horse-donkey T/G SNP in CpG context (CpG site #21) abolishes the CpG site in horse (Fig. 3b). At this CpG site, we expect a differential methylation pattern in hinny because the donkey mother is methylated (maternal methylation) and horse father is unmethylated (CpT); whereas in the hinny sample we expect 100% unmethylation. Again, the bisulfite sequencing results agree with these predictions. In the *PEG3* DMR, there is another horse-donkey nucleotide difference at CpG site #22, with C->T substitution in donkey. Because donkey does not have this CpG (TpG instead), we observed 100% unmethylation (Fig. 3B). The methylation patterns in the hybrids are also as expected. This is also a segregating polymorphism in our donkey CG samples (donkey 3689) at CpG site #12. We sequenced the gDNA for 3689 and it has C/T genotype. To determine the direction of transmission, we genotyped its parents by sequencing, and the father is C/C (donkey 3485) and the mother's allele is C/T (donkey 3418). Assuming mendelian transmission, the C allele is paternal and the T allele is maternal. Because we observed 100% unmethylation at this site, we know that the paternal C allele is unmethylated, which is consistent with maternal methylation for *PEG3*. We checked the DMR methylation profile for the *H19* and *PEG3* DMR in horse, donkey, mule and hinny CG, as well as in horse sperm gDNA. We found all four species to show differential methylation, indicating that the imprinting mechanism is conserved between mouse and horse for the two tested known imprinted genes, despite of the 70 million years divergence. The horse-donkey nucleotide differences served as a marker for allele-specific methylation quantification. We also observed CpG site turnovers between horse and donkey, and in each case the CpG acquired the methylation status appropriate to its neighboring sites. This suggests that the methylation and epigenetic reprogramming is functioning normally in hybrids. The turnovers of CpG sites lend insight as to why differential methylation at many CpG sites in the DMR region is required for the imprinting regulation. Apparently there is considerable tolerance to sequence substitutions at the DMR CpG sites during evolution.

HAT1 is one of the more interesting novel imprinted genes discovered in our RNA-Seq data, and we profiled the spatial and temporal distribution of imprinting status in a number of different placental, fetal and adult tissues. To check whether the imprinting status is linked to differential allelic methylation, we searched for CpG islands 50 kb upstream and 50 kb downstream from the gene. Three CpG islands were found, two of which are the promoter CpG islands of the upstream and downstream genes (*SI Appendix*, SI Methods). The *HAT1* CpG is 1.2 kb in length and covers the promoter region as well as the first exon. We assayed the methylation status in horse, donkey, mule and hinny CG gDNA samples, and found it is differentially methylated in all four species (Fig. 3C), suggesting that the *HAT1* promoter CpG island is the *HAT1* DMR. The *HAT1* DMR is 100% unmethylated in horse sperm. The maternal methylation is consistent with paternal expression. There is a horse-donkey nucleotide difference in the DMR, with A/A in horse and G/G in donkey. Based on the direction of parental transmission inferred from the heterozygous SNP, we confirmed that the maternal copy of the DMR is methylated in both mule and hinny CG gDNA samples. We also examined the methylation status in horse fetus and horse adult lymphocytes, in which we observed biallelic expression of *HAT1* from both parental alleles. In these two tissues where *HAT1* is not imprinted, we found that the DMR is not methylated at all (Fig. 3C). These results demonstrate that the tissue-specific imprinting status of *HAT1* is entirely consistent with the allele-specific methylation this gene.

Text S4. Verification of known and novel imprinted genes within species.

We discovered 93 candidate imprinted genes in our RNA-Seq data from mule and hinny CG samples. Because mule and hinnies have odd number of chromosomes ($2N = 63$) and they are generally sterile, this aberrant genome configuration may lead to dysregulation of gene expression and genomic imprinting in these interspecific hybrids. To exclude the possibility of random monoallelic expression, we verified the top 40 candidates in six different individuals by pyrosequencing as independent biological replicates. The next question is whether the novel imprinted genes we found are imprinted within the parental species, or they are just due to hybrid-specific effects. To assess this, we tested three selected genes in horse x horse or donkey x donkey crosses.

-IGF2. *IGF2* is one of the 16 candidate genes known to be imprinted in human and/or mouse, and it is the first imprinted gene discovered in mammals. The imprinting status of these known imprinted genes is conserved across different species. We discovered that these genes are also imprinted in mule and hinny placenta CG samples (Table 1). Based on conservation, we expect that they will be imprinted within species as well. To prove this, we selected a well-studied imprinted gene in human and mouse, *IGF2*, and checked its imprinting status in horse x horse crosses. We sequenced all the exons in a number of horses and could not find a single SNP. However, we found a (TG) dinucleotide repeat polymorphism in the 3'-UTR, about 1.2 kb after the stop codon. We cloned the region containing the dinucleotide repeats and genotyped by Sanger sequencing. We found a homozygous stallion (3105) of the reference allele (10/10), and a mare (2994) to be homozygous for the alternate 12/12 allele (*SI Appendix*, Fig. S20). In the heterozygous conceptus (10/12), there is only the paternal (10) allele in the CG cDNA sample, indicating that the *IGF2* is 100% allelic expression from the father (*SI Appendix*, Fig. S20). Therefore, we confirmed that *IGF2* is imprinted in horse x horse crosses.

-HAT1. *HAT1* is a novel imprinted gene we discovered in mule and hinny CG samples. We wanted to know whether *HAT1* is imprinted within species. To check this, we selected 22 horse CG samples, and fully sequenced the entire gene region (12 exons) in them and their parents by Sanger sequencing. Three horse SNPs were found in these 22 horse CG samples. SNP2 is in an intron. SNP1 is 647 bp downstream the gene model and SNP3 is 315 bp upstream. Because the predicted gene model could miss the UTRs, these two SNPs may be in the transcript. Therefore, we designed pyrosequencing primers to target them and both SNPs show no pyrosequencing signal, indicating that they are either too close to the transcription start (or end), or they are not in the transcript at all.

Because *HAT1* is a conserved gene, we failed to find informative SNPs in horse x horse crosses. We then Sanger sequenced 6 donkey CG samples, and found one SNP (SNP4) in exon 1 (*SI Appendix*, Fig. S21). We checked the direction of transmission by genotyping their parents and we found that SNP4 (T/G) is informative in 3 of the 6 donkey individuals, with G allele transmitted from the father in one individual (donkey 4106) and from the mother in the other two (donkey 3689 and 3693). The advantage of having both transmission directions is that we can rule out the possibility of 100% eQTL explanation. We quantified the allelic expression ratios by pyrosequencing in three informative samples, with one homozygote individual as control (donkey 3643). We confirmed that *HAT1* is imprinted in donkey CG samples (*SI Appendix*, Fig. S21).

-INSR. *INSR* is another novel imprinted gene we discovered on chromosome 7 in mule and hinny CG samples, with about 80% expression from the paternal allele. The imprinting status in hybrids is variable (*SI Appendix*, Text S2 and Table S5). We verified in six individuals and two show biallelic expression (*SI Appendix*, Text S1). We Sanger sequenced the *INSR* exons of 17 horse CGs samples as well as their parents and we found several useful SNPs. Two of these within horse SNPs are informative in two individuals and one (CUHSNP00055036) is informative in five individuals. CUHSNP00055036 is in the 3'UTR of *INSR* gene, 2 bp after the stop codon. We tested the allelic expression ratios in 3 of 5 informative individuals for which we have cDNA samples, and discovered preferential paternal expression in all three individual CG samples (*SI Appendix*, Fig. S22). We concluded that *INSR* is also imprinted in horse day-33 CG samples. We did observe some variability in allele-specific expression ratios, but the imprinting status is stable in all three individuals we tested (*SI Appendix*, Fig. S22).

The low SNP density (1 per kb on average), precludes us from testing allelic expression for every single known imprinted gene in horse x horse crosses, especially genes that are more conserved than average. However, we can use another method to check the within-species imprinting status. It had been shown in mouse and human, for a subset of the imprinted genes, the imprinting status is regulated by differentially methylated regions (DMRs). We found that the DMR regulation mechanism is conserved between mouse and equids for several known imprinted genes in the imprinting

clusters (*SI Appendix*, Text S3). If we could show differential methylation at horse and donkey DMRs, we will be able to provide indirect evidence to support the imprinting status within species. We checked syntenic region of the mouse *H19* and *PEG3* DMRs, as well as the novel *HAT1* DMR we discovered in horse, donkey, mule and hinny (*SI Appendix*, Text S3). Differential methylation was observed for all four species, suggesting that they are imprinted in horse and donkey.

In summary, we tested the imprinting status of 5 imprinted genes discovered in the hybrids, *IGF2*, *INSR*, *HAT1*, *PEG3* and *H19* within parental species, by allele-specific pyrosequencing on informative SNPs and/or bisulfite sequencing of the DMRs. 5 of 5 are imprinted in parental species, suggesting that the imprinting status is properly regulated in hybrids.

SI References

1. Antczak DF, *et al.* (1987) Differentiation molecules of the equine trophoblast. *J Reprod Fertil Suppl* 35:371-378.
2. de Mestre AM, *et al.* (2009) Glial cells missing homologue 1 is induced in differentiating equine chorionic girdle trophoblast cells. *Biol Reprod* 80(2):227-234.
3. Li H & Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754-1760.
4. Mortazavi A, Williams BA, McCue K, Schaeffer L, & Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5(7):621-628.
5. Li H, *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078-2079.
6. Wang X, *et al.* (2008) Transcriptome-wide identification of novel imprinted genes in neonatal mouse brain. *PLoS One* 3(12):e3839.
7. Storer BE & Kim C (1990) Exact Properties of Some Exact Test Statistics for Comparing 2 Binomial Proportions. *Journal of the American Statistical Association* 85(409):146-155.
8. Richly E & Leister D (2004) NUMTs in sequenced eukaryotic genomes. *Mol Biol Evol* 21(6):1081-1084.
9. Marsh S (2007) *Pyrosequencing protocols* (Humana Press, Totowa, N.J.) pp xii, 196 p.
10. Gimelbrant A, Hutchinson JN, Thompson BR, & Chess A (2007) Widespread monoallelic expression on human autosomes. *Science* 318(5853):1136-1140.
11. Proudhon C & Bourc'his D (2010) Identification and resolution of artifacts in the interpretation of imprinted gene expression. *Brief Funct Genomics* 9(5-6):374-384.
12. Wang Y, *et al.* (2004) The mouse Murr1 gene is imprinted in the adult brain, presumably due to transcriptional interference by the antisense-oriented U2af1-rs1 gene. *Molecular and cellular biology* 24(1):270-279.
13. Zhang Z, *et al.* (2006) Comparative analyses of genomic imprinting and CpG island-methylation in mouse Murr1 and human MURR1 loci revealed a putative imprinting control region in mice. *Gene* 366(1):77-86.
14. Mallya M, Campbell RD, & Aguado B (2006) Characterization of the five novel Ly-6 superfamily members encoded in the MHC, and detection of cells expressing their potential ligands. *Protein Sci* 15(10):2244-2256.
15. Klein RJ, *et al.* (2005) Complement factor H polymorphism in age-related macular degeneration. *Science* 308(5720):385-389.
16. Verreault A, Kaufman PD, Kobayashi R, & Stillman B (1998) Nucleosomal DNA regulates the core-histone-binding subunit of the human Hat1 acetyltransferase. *Curr Biol* 8(2):96-108.
17. Bjornsson HT, *et al.* (2008) SNP-specific array-based allele-specific expression analysis. *Genome research* 18(5):771-779.
18. Bartolomei MS & Ferguson-Smith AC (2011) Mammalian Genomic Imprinting. *Cold Spring Harb Perspect Biol* 3(7).
19. Pauler FM & Barlow DP (2006) Imprinting mechanisms--it only takes two. *Genes Dev* 20(10):1203-1206.
20. Delaval K & Feil R (2004) Epigenetic regulation of mammalian genomic imprinting. *Curr Opin Genet Dev* 14(2):188-195.
21. Spahn L & Barlow DP (2003) An ICE pattern crystallizes. *Nat Genet* 35(1):11-12.
22. Reik W & Walter J (2001) Genomic imprinting: parental influence on the genome. *Nat Rev Genet* 2(1):21-32.
23. Thorvaldsen JL, Fedoriw AM, Nguyen S, & Bartolomei MS (2006) Developmental profile of H19 differentially methylated domain (DMD) deletion alleles reveals multiple roles of the DMD in regulating allelic expression and DNA methylation at the imprinted H19/Igf2 locus. *Molecular and cellular biology* 26(4):1245-1258.
24. Tremblay KD, Duran KL, & Bartolomei MS (1997) A 5' 2-kilobase-pair region of the imprinted mouse H19 gene exhibits exclusive paternal methylation throughout development. *Molecular and cellular biology* 17(8):4322-4329.
25. Reik W, Dean W, & Walter J (2001) Epigenetic reprogramming in mammalian development. *Science* 293(5532):1089-1093.
26. Li LL, Szeto IY, Cattanach BM, Ishino F, & Surani MA (2000) Organization and parent-of-origin-specific methylation of imprinted Peg3 gene on mouse proximal chromosome 7. *Genomics* 63(3):333-340.
27. Wang X, Soloway PD, & Clark AG (2011) A survey for novel imprinted genes in the mouse placenta by mRNA-seq. *Genetics* 189(1):109-122.

Figure S1. Chromosomal location of known and novel imprinted genes in mule and hinny chorionic girdle.

The candidate imprinted genes are labeled on the horse chromosomes accoridng to their physical location. The mtDNA insertion genes and genes on chrUn were not shown. Maternal expressed imprinted candidates are plotted in red and paternally expressed candidates in blue. The verified imprinted genes are shown in bold. Ideogram of horse chromosomes are color-coded with human chromosome synteny.

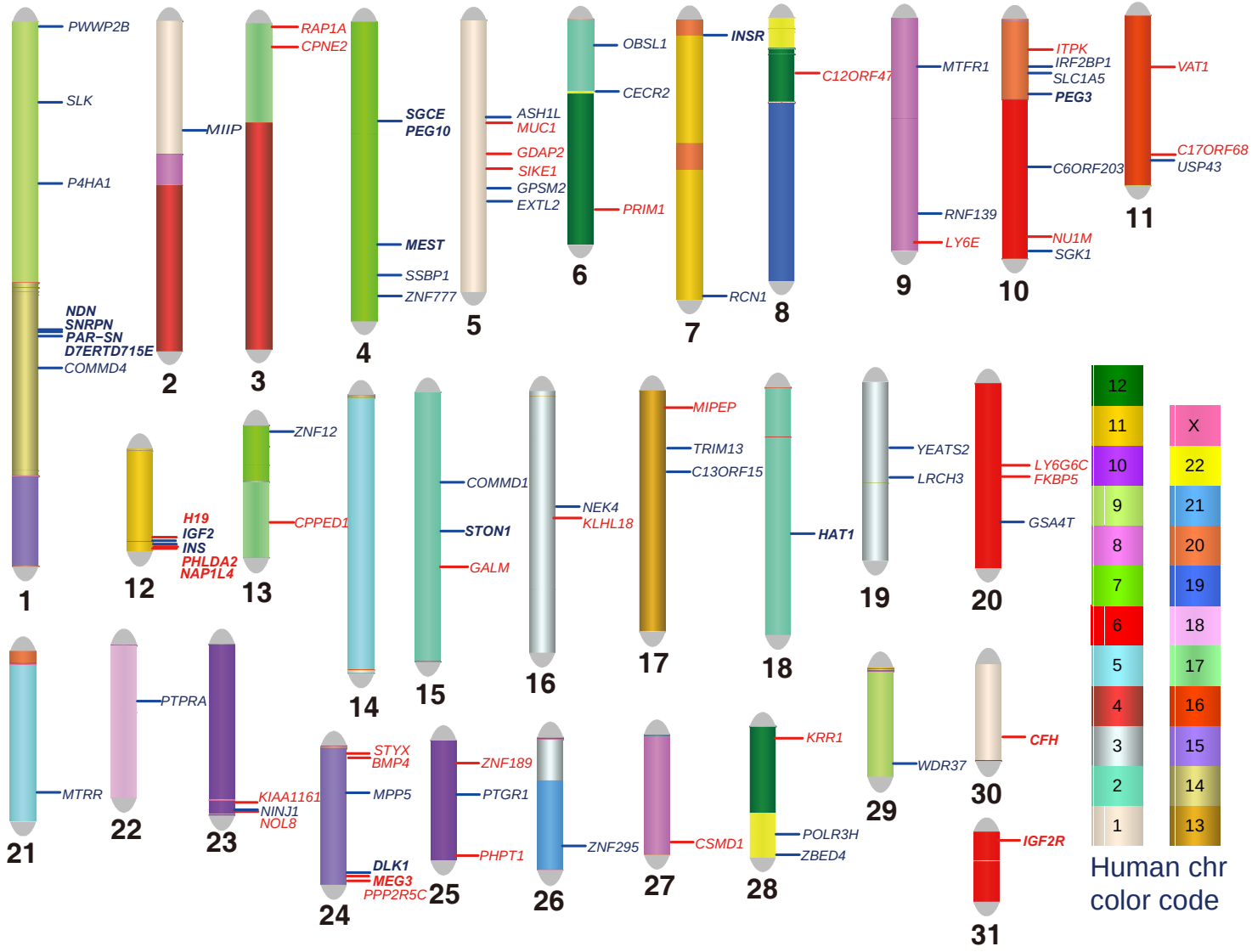


Figure S2. Verification of *H19* imprinting status in mule and hinny CG samples by allele-specific pyrosequencing. (A) The SNP allele counts in mule and hinny in RNA-Seq data. **(B)** Genotyping results for the parental alleles for informative SNP CUHSNP00007410. **(C)** Pyrosequencing results for *H19*. We confirmed 100% maternal expression in mule and hinny from the pyrosequencing results.

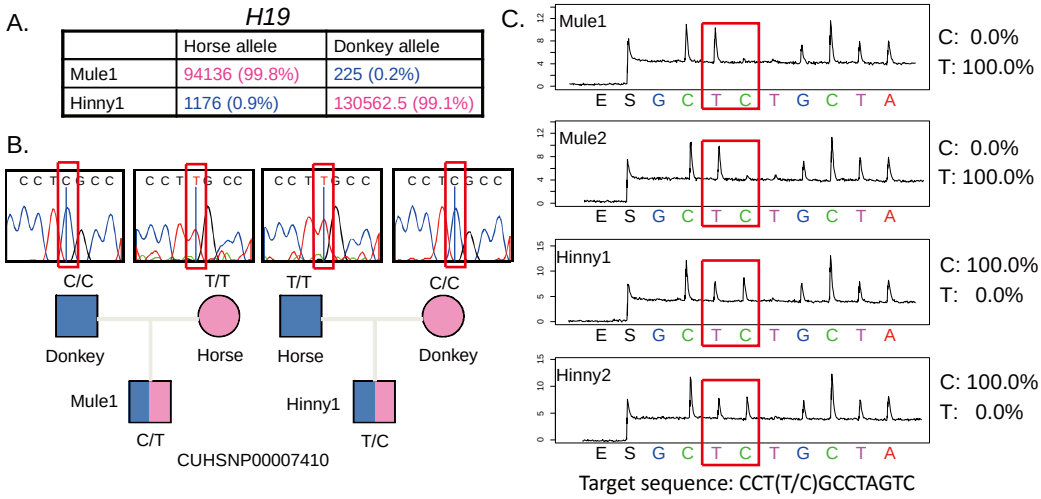


Figure S3. Verification of *PHLDA2* imprinting status in mule and hinny CG samples by allele-specific pyrosequencing. (A) The SNP allele counts in mule and hinny in Illumina RNA-Seq data. **(B)** Genotyping results for the parental alleles for informative SNP CUHSNP00034727. **(C)** Pyrosequencing results for *PHLDA2*. We confirmed 100% maternal expression in mule and hinny.

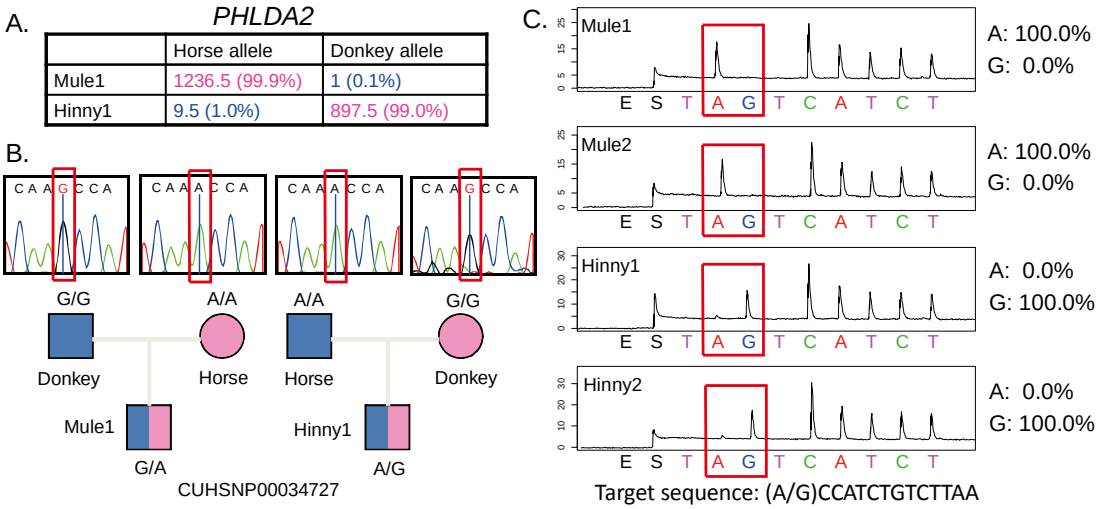


Figure S4. Verification of *MEG3* imprinting status in mule and hinny CG samples by allele-specific pyrosequencing. (A). The SNP allele counts in mule and hinny in Illumina RNA-Seq data. **B.** Genotyping results for the parental alleles for informative SNP CUHSNP00007837 (the second SNP shown in the Sanger sequencing trace). **C.** Pyrosequencing results for *MEG3*. We confirmed 100% maternal expression in mule and hinny from the pyrosequencing results.

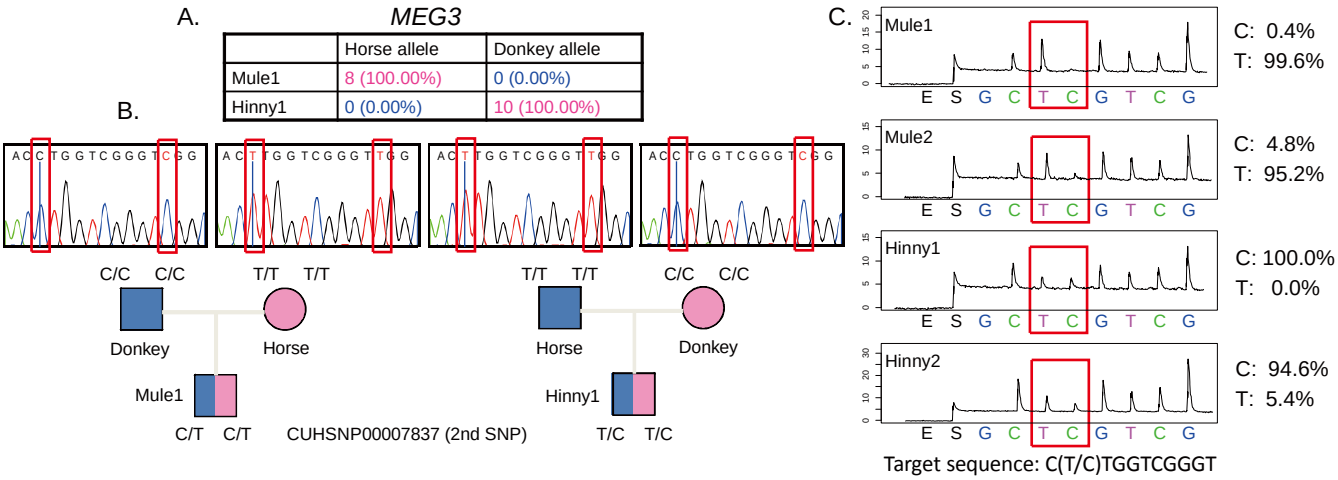
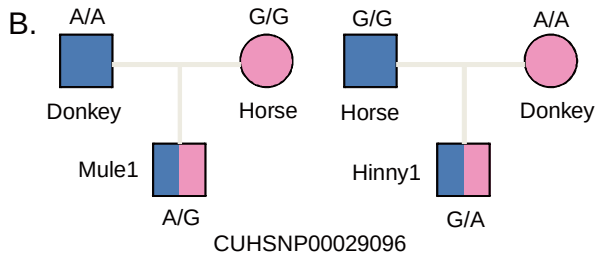


Figure S5. Verification of *IGF2R* imprinting status in mule and hinny CG samples by allele-specific pyrosequencing. (A) The SNP allele counts in mule and hinny in RNA-Seq data. **(B)** Genotyping results for the parental alleles for informative SNP CUHSNP00029096. **(C)** Pyrosequencing results for *IGF2R*. The pyro primer is targeting the opposite strand. We confirmed preferential expression of maternal allele in mule and hinny from the pyrosequencing results.

A. *IGF2R*

	Horse allele	Donkey allele
Mule1	2233 (73.6%)	801.5 (26.4%)
Hinny1	669 (24.1%)	2109 (75.9%)



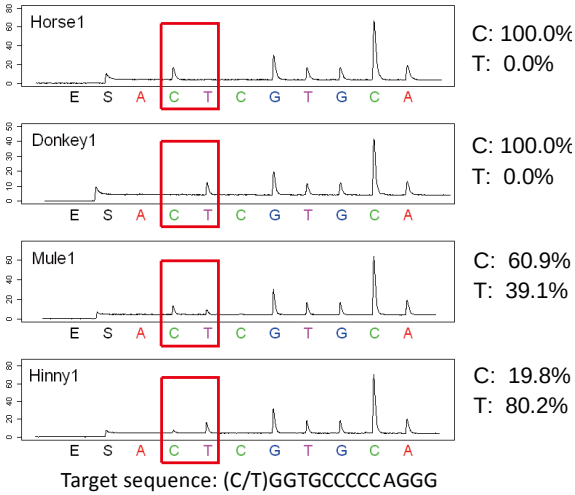
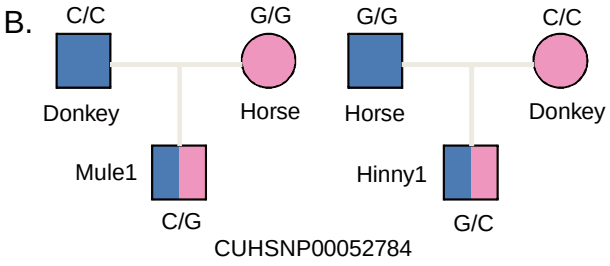


Figure S6. Verification of *NAP1L4* imprinting status in mule and hinny CG samples by allele-specific pyrosequencing. (A) The SNP allele counts in mule and hinny in RNA-Seq data. **(B)** Genotyping results for the parental alleles for informative SNP CUHSNP00052784. **(C)** Pyrosequencing results for *NAP1L4*. We confirmed preferential maternal expression.

A. *NAP1L4*

	Horse allele	Donkey allele
Mule1	375.5 (72.7%)	141 (27.3%)
Hinny1	46.5 (17.9%)	213 (82.1%)



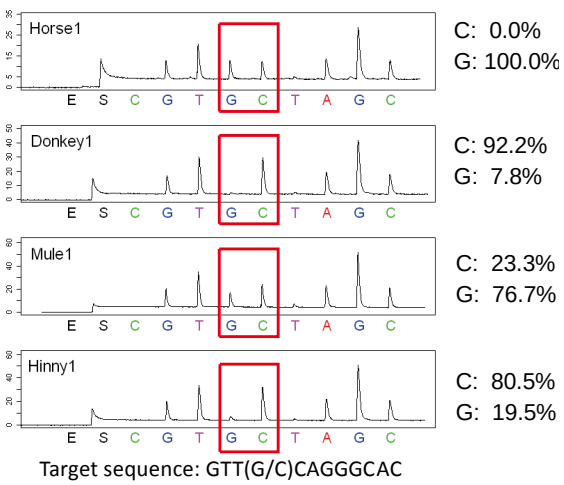
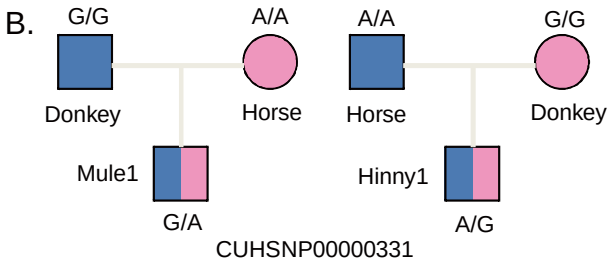


Figure S7. Verification of *IGF2* imprinting status in mule and hinny CG samples by allele-specific pyrosequencing. (A) The SNP allele counts in mule and hinny in RNA-Seq data. **(B)** Genotyping results for the parental alleles for informative SNP CUHSNP00000331. **(C)** Pyrosequencing results for *IGF2*. We confirmed 100% paternal expression.

A. *IGF2*

	Horse allele	Donkey allele
Mule1	1.5 (0.4%)	421.5 (99.6%)
Hinny1	171 (100.0%)	0 (0%)



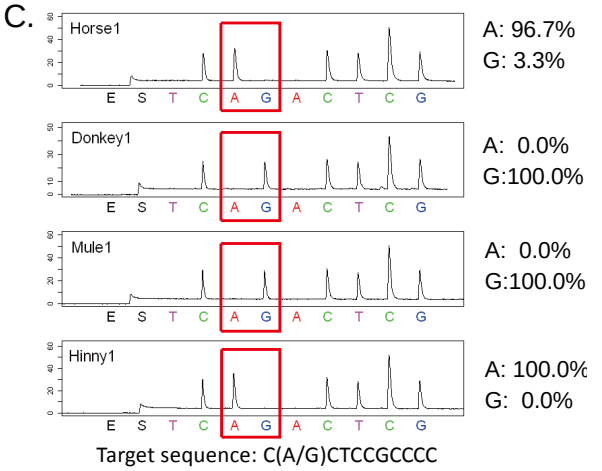
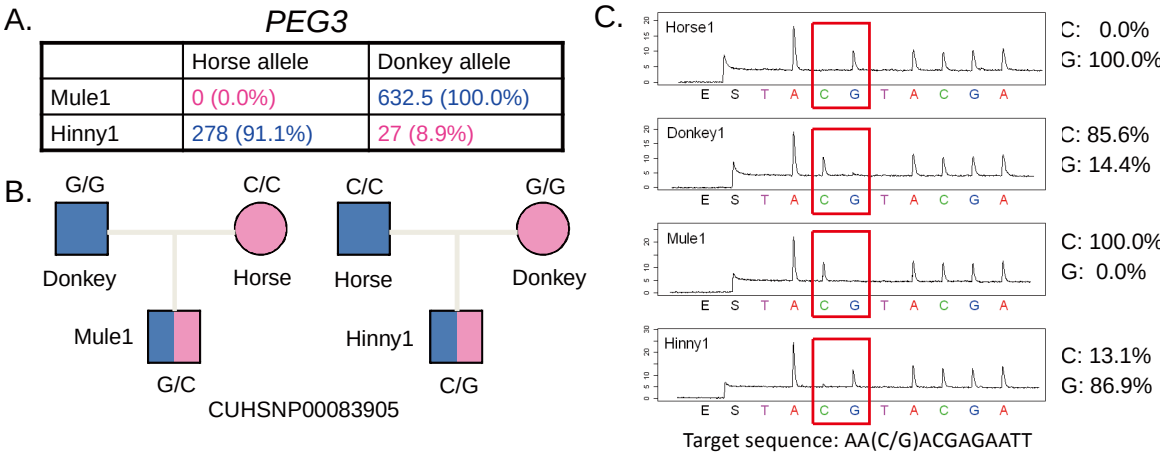
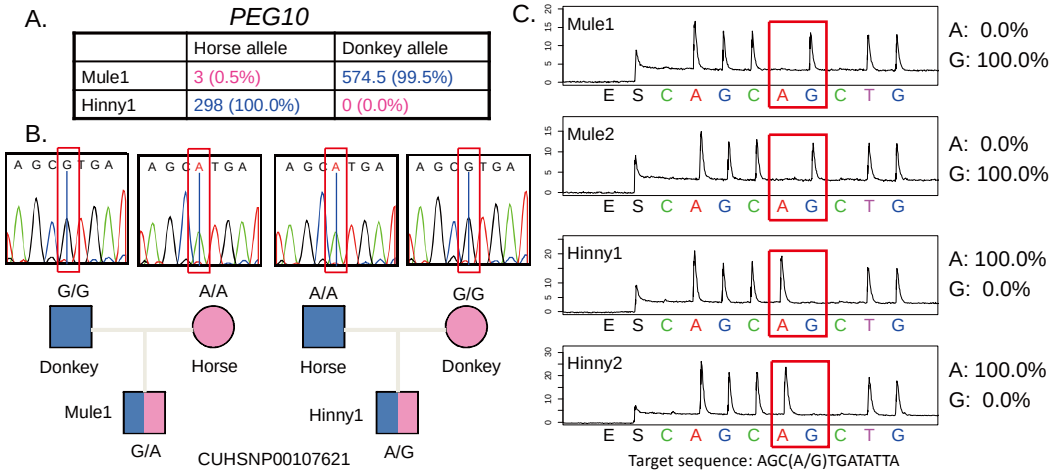


Figure S8. Verification of *PEG3* imprinting status in mule and hinny CG samples by allele-specific pyrosequencing. **(A)** The SNP allele counts in mule and hinny in RNA-Seq data. **(B)** Genotyping results for the parental alleles for informative SNP CUHSNP00083905. **(C)** Pyrosequencing results for *PEG3*. We confirmed preferential paternal expression in hybrids.



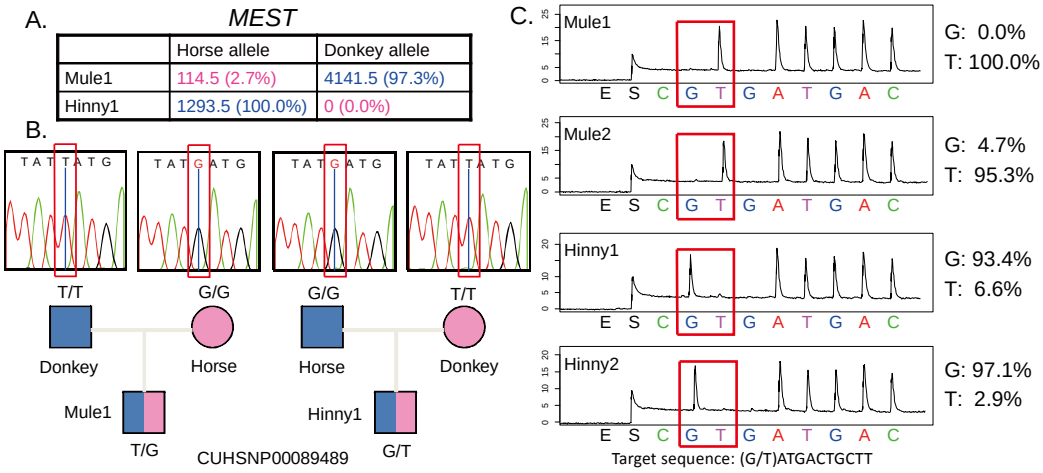
	Horse allele	Donkey allele
Mule1	0 (0.0%)	632.5 (100.0%)
Hinny1	278 (91.1%)	27 (8.9%)

Figure S9. Verification of *PEG10* imprinting status in mule and hinny CG samples by allele-specific pyrosequencing. **(A)** The SNP allele counts in mule and hinny in Illumina RNA-Seq data. **(B)** Genotyping results for the parental alleles for informative SNP CUHSNP00107621. **(C)** Pyrosequencing results for *PEG10*. We confirmed 100% expression from the paternal allele in mule and hinny from the pyrosequencing results.



	Horse allele	Donkey allele
Mule1	3 (0.5%)	574.5 (99.5%)
Hinny1	298 (100.0%)	0 (0.0%)

Figure S10. Verification of *MEST* imprinting status in mule and hinny CG samples by allele-specific pyrosequencing. **(A)** The SNP allele counts in mule and hinny in RNA-Seq data. **(B)** Genotyping results for the parental alleles for informative SNP CUHSNP00089489. **(C)** Pyrosequencing results for *MEST*. We confirmed 100% paternal expression.



	Horse allele	Donkey allele
Mule1	114.5 (2.7%)	4141.5 (97.3%)
Hinny1	1293.5 (100.0%)	0 (0.0%)

Figure S11. Verification of *SNRPN* imprinting status in mule and hinny CG samples by allele-specific pyrosequencing. (A) The SNP allele counts in mule and hinny in RNA-Seq data. **(B)** Genotyping results for the parental alleles for informative SNP CUHSNP00035546. **(C)** Pyrosequencing results for *SNRPN*. We confirmed 100% paternal expression.

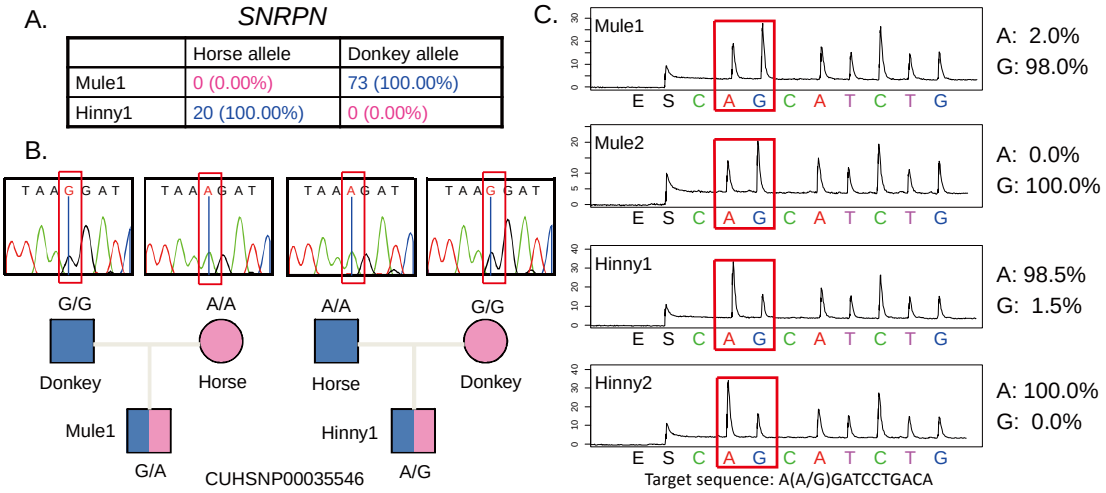


Figure S12. Verification of *DLK1* imprinting status in mule and hinny CG samples by allele-specific pyrosequencing. (A). The SNP allele counts in mule and hinny in RNA-Seq data. **(B).** Genotyping results for the parental alleles for informative SNP CUHSNP00121873. **(C).** Pyrosequencing results for *DLK1*. We confirmed 100% paternal expression.

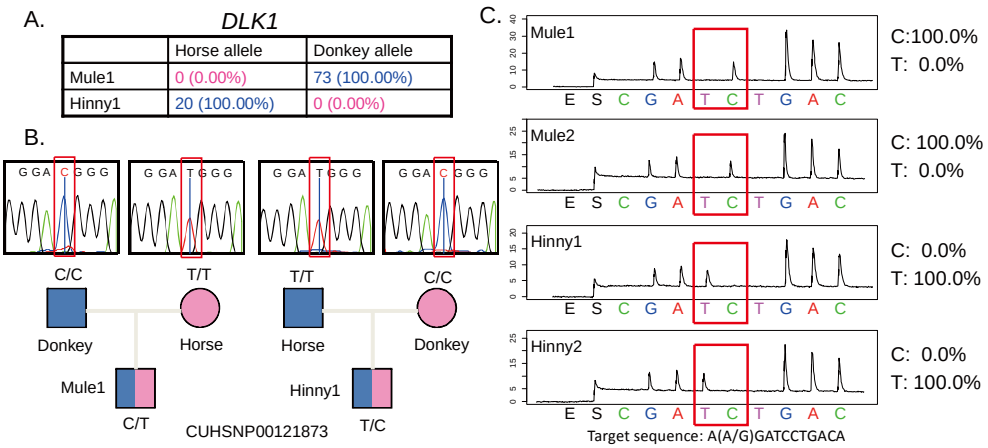


Figure S13. Verification of *PAR-SN* imprinting status in mule and hinny CG samples by allele-specific pyrosequencing. (A) The SNP allele counts in mule and hinny in RNA-Seq data. **(B)** Genotyping results for the parental alleles for informative SNP CUHSNP00007879. **(C)** Pyrosequencing results for *PAR-SN*. The pyro primer is targeting the opposite strand. We confirmed 100% expression from the paternal allele in mule and hinny from the pyrosequencing results.

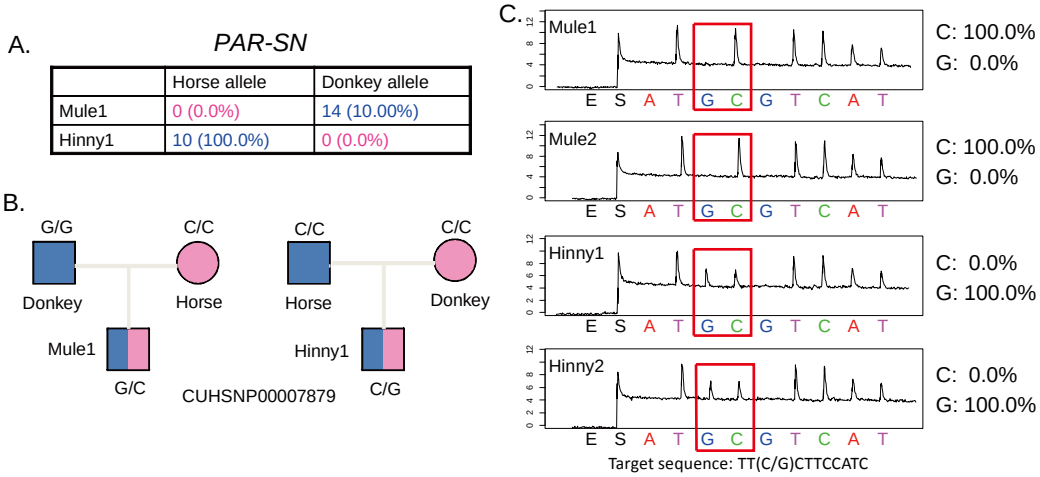
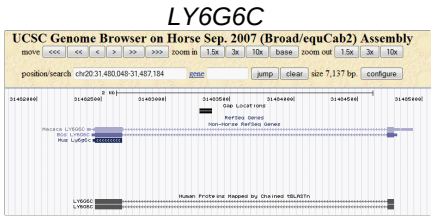


Figure S14. Verification of *LY6G6C* imprinting status in mule and hinny CG samples by allele-specific pyrosequencing. (A) Screenshot from UCSC genome browser, showing the chromosomal location and gene model for *LY6G6C*. **(B)** The SNP allele counts in mule and hinny in RNA-Seq data. **(C)** Genotyping results for the parental alleles for SNP CUHSNP00123114 and CUHSNP00123115. CUHSNP00123115 is not informative in hinny2. **(D)** Pyrosequencing results for *LY6G6C* at SNP position CUHSNP00123114. We confirmed preferential maternal expression in 5 of 5 mule/hinny individuals.

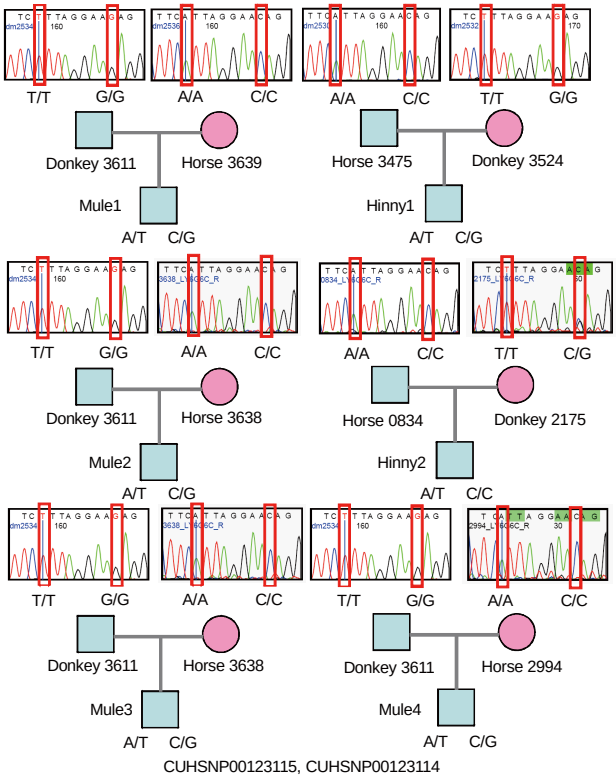
A.



B.

	Horse allele	Donkey allele
Mule1	91 (68.4%)	42 (31.6%)
Hinny1	9 (9.0%)	93 (91.0%)

C.



D.

Pyrosequencing results

Individual	Mother	Father	horse allele %	donkey allele %
horse1	horse	horse	100.00%	0.00%
donkey1	donkey	donkey	0.00%	100.00%
mule1	horse	donkey	83.95%	16.05%
hinny1	donkey	horse	11.00%	89.00%
mule2	horse	donkey	81.95%	18.05%
hinny2	donkey	horse	33.70%	66.30%
mule3	horse	donkey	99.50%	0.50%
mule4	horse	donkey	NA	NA

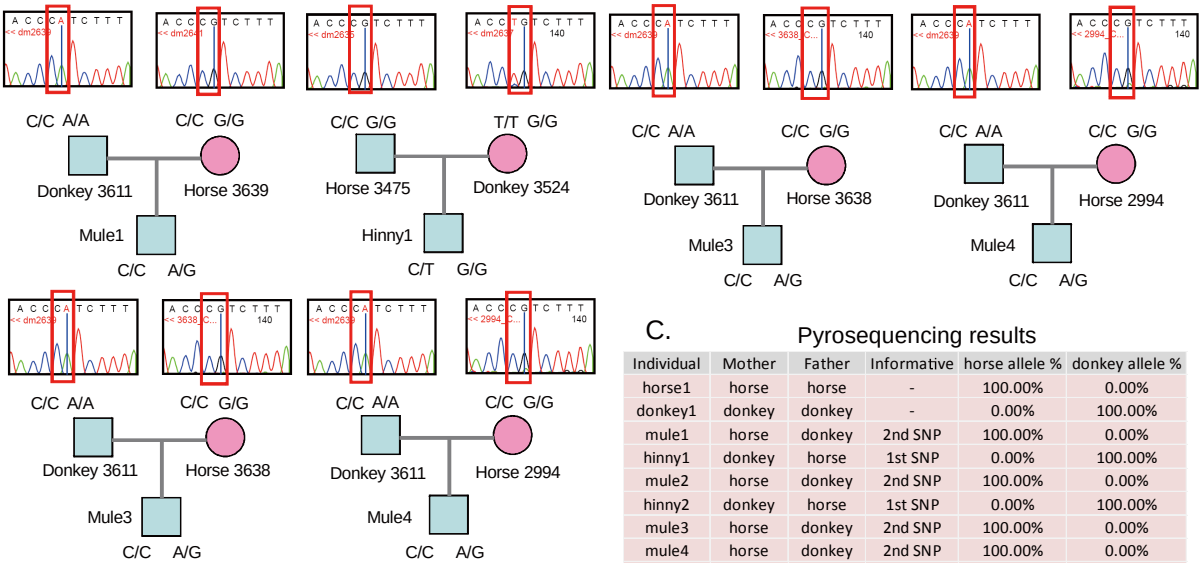
Figure S15. Verification of *CFH* imprinting status in mule and hinny CG samples by allele-specific pyrosequencing. (A) The SNP allele counts in mule and hinny in RNA-Seq data. **(B)** Genotyping results for the parental alleles for SNP CUHSNP00030016 and an adjacent novel SNP. CUHSNP00030016 is informative in all four mules and the novel SNP is informative in two hinnies. **(C)** Pyrosequencing results for *CFH* at SNP position CUHSNP00030016 and the novel SNP. We confirmed 100% maternal expression in 6 of 6 mule/hinny individuals. The pyro signal is relatively low because the expression level is low in the tissue we tested.

A.

CFH

	Horse allele	Donkey allele
Mule1	8 (100.00%)	0 (0.00%)
Hinny1	0 (0.00%)	11 (100.00%)

B.



C.

Pyrosequencing results

Individual	Mother	Father	Informative	horse allele %	donkey allele %
horse1	horse	horse	-	100.00%	0.00%
donkey1	donkey	donkey	-	0.00%	100.00%
mule1	horse	donkey	2nd SNP	100.00%	0.00%
hinny1	donkey	horse	1st SNP	0.00%	100.00%
mule2	horse	donkey	2nd SNP	100.00%	0.00%
hinny2	donkey	horse	1st SNP	0.00%	100.00%
mule3	horse	donkey	2nd SNP	100.00%	0.00%
mule4	horse	donkey	2nd SNP	100.00%	0.00%

Figure S16. Verification of *HAT1* imprinting status in mule and hinny CG samples by allele-specific pyrosequencing. (A) The SNP allele counts in mule and hinny in RNA-Seq data. **(B)** Genotyping results for the parental alleles for SNP CUHSNP00046513. **(C)** Pyrosequencing results for *HAT1* at SNP position CUHSNP00046513. The pyro primer is targeting the opposite strand. We confirmed 90%-97% paternal expression in 4 mules and 100% paternal expression in 2 hinnies.

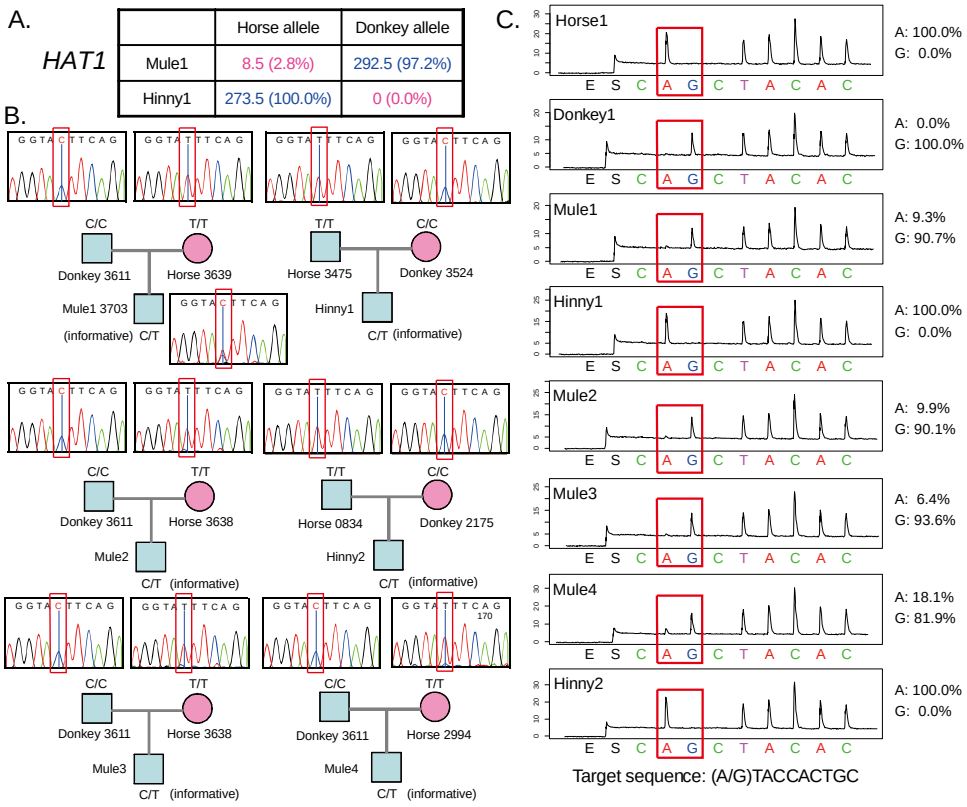


Figure S17. Verification of *INSR* imprinting status in mule and hinny CG samples by allele-specific pyrosequencing. (A) The SNP allele counts in mule and hinny in Illumina RNA-Seq data. **(B)** Genotyping results for the parental alleles for SNP CUHSNP00055072. **(C)** Pyrosequencing results for *INSR* at SNP position CUHSNP00055072. The pyro primer is targeting the opposite strand. We verified preferential paternal expression in 4 of 6 mule/hinny individuals.

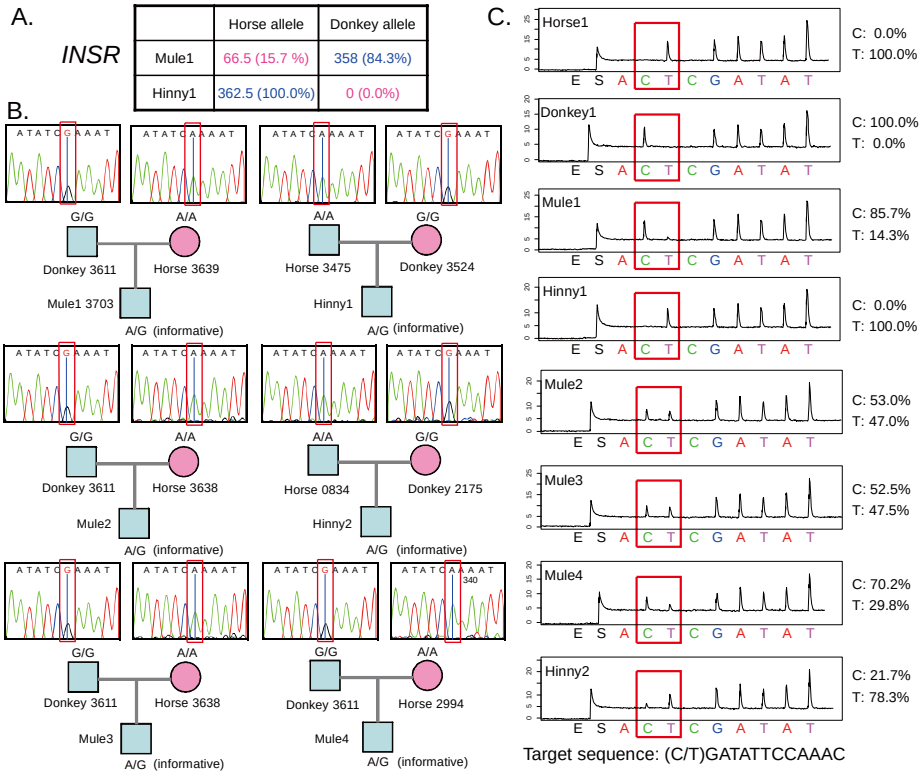


Figure S18. Verification of *D7ERTD715E* imprinting status in mule and hinny CG samples by allele-specific pyrosequencing. (A) The SNP allele counts in mule and hinny in RNA-Seq data. **(B)** Genotyping results for the parental alleles for SNP CUHSNP00194382 and CUHSNP00194383. **(C)** Pyrosequencing results for *D7ERTD715E*. The pyro primer is targeting the opposite strand.

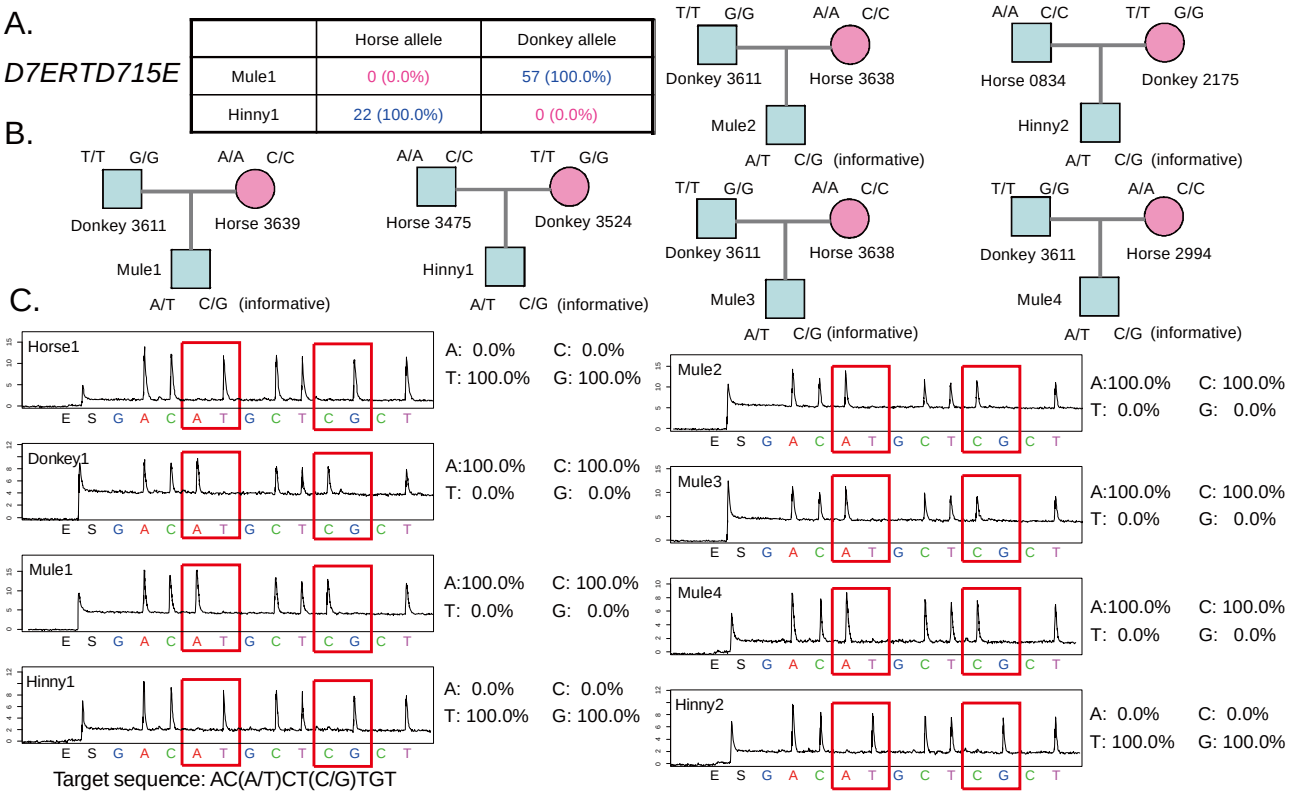


Figure S19. Verification of *STON1* imprinting status in mule and hinny CG samples by allele-specific pyrosequencing. (A) The SNP allele counts in mule and hinny in Illumina RNA-Seq data. **(B)** Genotyping results for the parental alleles for SNP CUHSNP00192769. **(C)** Pyrosequencing results for *STON1* at SNP positions CUHSNP00192769.

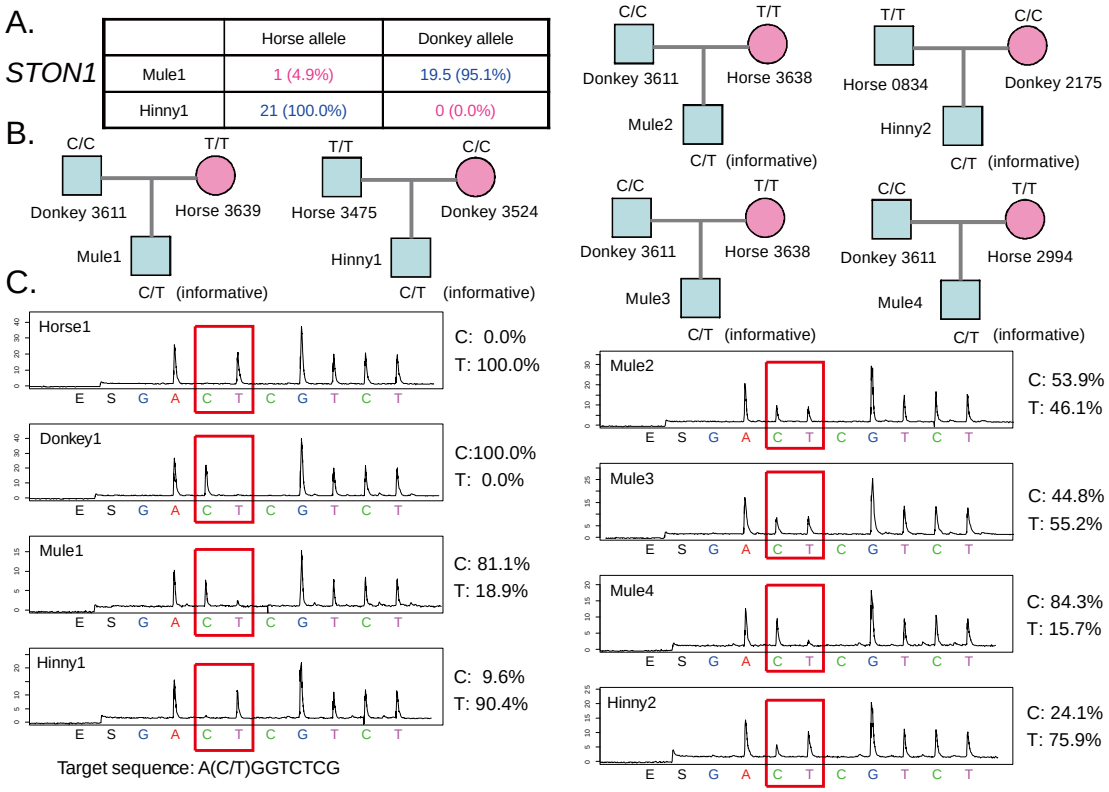


Figure S20. Verification of *IGF2* imprinting status in horse x horse crosses. (A) Genotypes of the Tracy x Sparky crosses for the TG dinucleotide repeats in *IGF2* 3'-UTR. (**B)** Sanger sequencing alignment results for the TG dinucleotide repeats the conceptus CG cDNA sample and the parental gDNA samples.

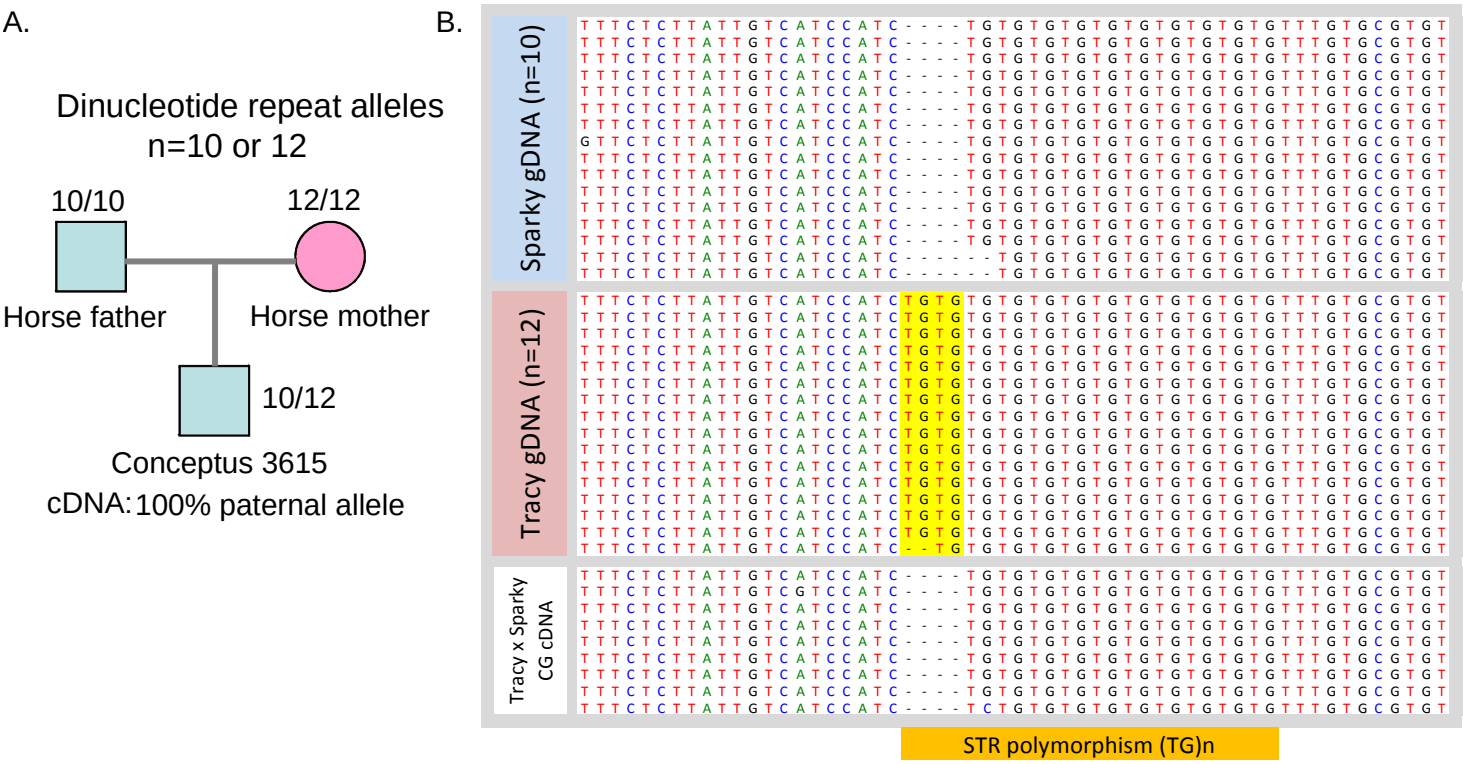


Figure S21. SNP discovery and verification of HAT1 imprinting status in parental crosses.

(A) SNP discovery results for *HAT1* in horse x horse crosses. The highlighted heterozygous positions are not within the transcripts.

(**B)** Donkey CG and parental gDNA genotyping results. The heterozygous positions in the CG samples are highlighted.

(C) Quantification of allelic expression ratio in one homozygote control donkey 3643 by allele-specific pyrosequencing. The top left panel is the SNP genotyping results in 3643 CG and parental blood gDNA samples, showing that 3643 and its parents all have homozygous G/G alleles. The top-right panel is the pyrosequencing results using both oligo-dT and random decamer primed cDNA samples. We observed 100% of G allele, indicating that our pyrosequencing reaction is working properly. The bottom panel is the RNA-Seq alignments for donkey 3643 sample. All alleles from reads aligned to the SNP position are G alleles, which is consistent with the pyrosequencing results.

(D). Quantification of allelic expression ratio in an informative heterozygous donkey 4106 (G/T) by allele-specific pyrosequencing. The top left panel is the SNP genotyping results in 4106 CG and parental blood gDNA samples, showing that the G allele is transmitted from the father (1992: G/G) and T allele is transmitted from the mother (1987: G/T). 4106 has the opposite transmission direction compared to the two previous informative hets. The top-right panel is the pyrosequencing results using both oligo-dT and random decamer primed cDNA samples. We observed >100% allelic expression ratio from the G allele, suggesting 100% paternal expression.

(E) Quantification of allelic expression ratio in an informative heterozygous donkey 3689 (G/T) by allele-specific pyrosequencing. The top left panel is the SNP genotyping results in 3689 CG and parental blood gDNA samples, showing that the T allele is transmitted from the father (3489: G/T) and G allele is transmitted from the mother (3418: G/G). The top-right panel is the pyrosequencing results using both oligo-dT and random decamer primed cDNA samples. We observed >98% allelic expression ratio from the T allele, suggesting close to 100% paternal expression. The bottom panel is the RNA-Seq alignments for donkey 3689. 11/12 reads aligned to the SNP position have the T alleles, which is consistent with the pyrosequencing results.

(F) Quantification of allelic expression ratio in an informative heterozygous donkey 3693 (G/T) by allele-specific pyrosequencing. The top left panel is the SNP genotyping results in 3693 CG and parental blood gDNA samples, showing that the T allele is transmitted from the father (3489: G/T) and G allele is transmitted from the mother (3418: G/G). The top-right panel is the pyrosequencing results using both oligo-dT and random decamer primed cDNA samples. We observed >96% allelic expression ratio from the T allele, suggesting close to 100% paternal expression. The bottom panel is the RNA-Seq alignments for donkey 3693. All 12 reads aligned to the SNP position have the T alleles, which is consistent with the pyrosequencing results.

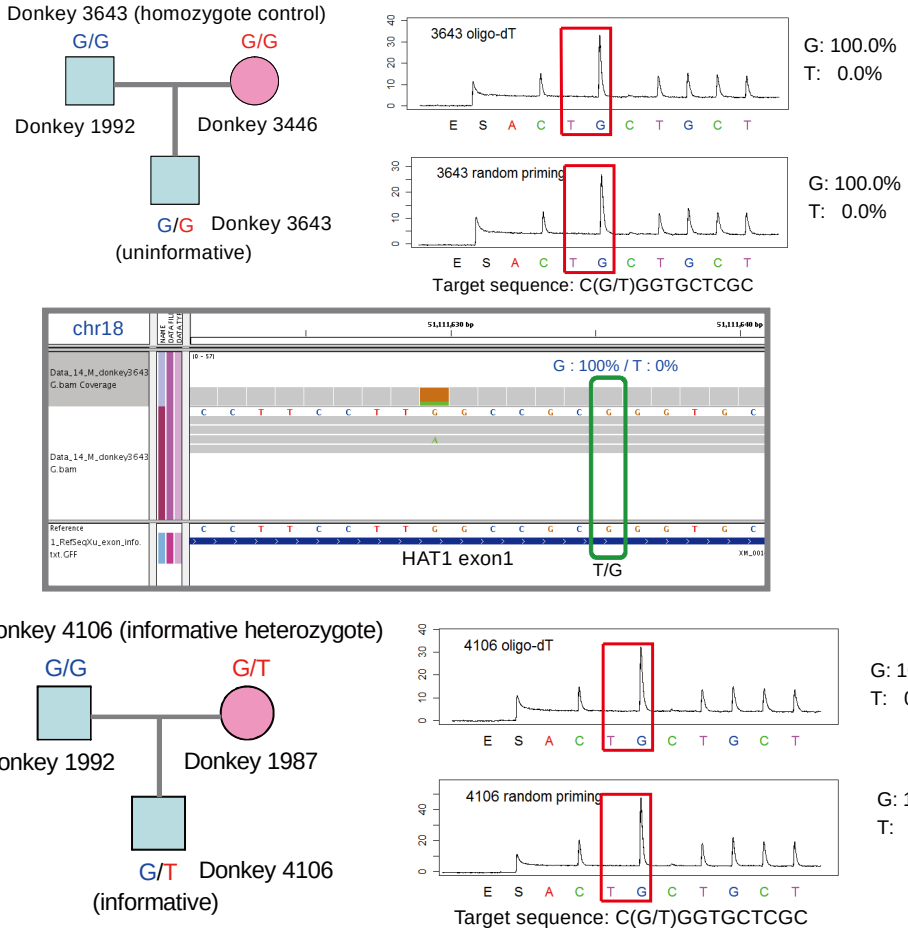
A.

HAT1 Horse					SNP1			SNP3		
Sample #.	ID	Cross	Father	Mother	genotype	Father	Mother	genotype	Father	Mother
1.	3868	3474 x 3639	3474	3639	T/C	T/T	T/C	T/T	T/T	T/C
2.	3881	3474 x 3641	3474	3641	T/C	T/T	T/C	T/T	T/T	T/C
3.	3885	3474 x 3640	3474	3640	T/T	T/T	T/C		T/T	T/C
4.	3894	3475 x 3640	3475	3640	T/C	T/T	T/C	T/C	C/C	T/C
5.	3907	3474 x 3641	3474	3641	T/C	T/T	T/C	T/T	T/T	T/C
6.	3917	3475 x 3640	3475	3640	T/T	T/T	T/C		C/C	T/C
7.	3940	3474 x 3639	3474	3639	T/C	T/T	T/C	T/T	T/T	T/C
8.	3948	3474 x 3641	3474	3641	T/T	T/T	T/C		T/T	T/C
9.	3959	3474 x 3639	3474	3639	T/T	T/T	T/C		T/T	T/C
10.	3960	3474 x 3640	3474	3640	T/T	T/T	T/C		T/T	T/C
11.	3994	3474 x 3640	3474	3640	T/C	T/T	T/C	T/T	T/T	T/C
12.	3999	3475 x 3641	3475	3641	T/C	T/T	T/C	T/C	C/C	T/C
13.	4057	3474 x 3640	3474	3640	T/T	T/T	T/C		T/T	T/C
14.	4060	3474 x 3641	3474	3641	T/T	T/T	T/C		T/T	T/C
15.	4061	3475 x 3639	3475	3639	T/C	T/T	T/C	T/C	C/C	T/C
16.	4070	3474 x 3640	3474	3640	T/C	T/T	T/C	T/T	T/T	T/C
17.	4071	3474 x 3639	3474	3639	T/C	T/T	T/C	T/T	T/T	T/C
18.	4077	3474 x 3641	3474	3641	T/T	T/T	T/C		T/T	T/C
19.	4079	3474 x 3639	3474	3639	T/C	T/T	T/C	T/T	T/T	T/C
20.	4083	3474 x 3640	3474	3640	T/T	T/T	T/C		T/T	T/C
21.	4095	3474 x 3639	3474	3639	T/T	T/T	T/C		T/T	T/C
22.	4096	3474 x 3641	3474	3641	T/C	T/T	T/C		T/T	T/C

B.

HAT1 Donkey					SNP4		
Sample #.	ID	Cross	Father	Mother	genotype	Father	Mother
1.	4106	1992x1987	1992	1987	G/T	G/G	G/T
2.	3574	1992x3418	1992	3418	G/G	G/G	G/G
3.	3643	1992x3446	1992	3446	G/G	G/G	G/G
4.	3694	3489x3446	3489	3446	G/G	G/T	G/G
5.	3689	3489x3418	3489	3418	G/T	G/T	G/G
6.	3693	3489x3418	3489	3418	G/T	G/T	G/G

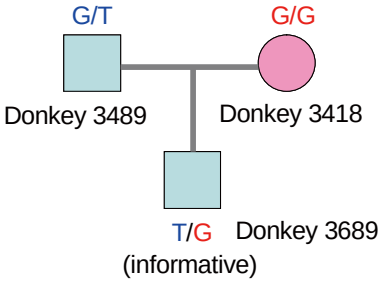
C.

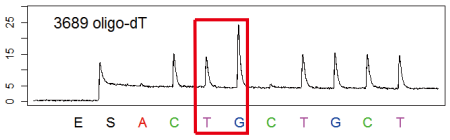


D.

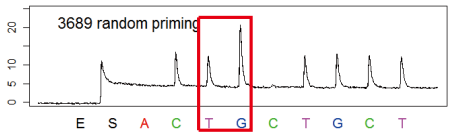
E.

Donkey 3689 (informative heterozygote)



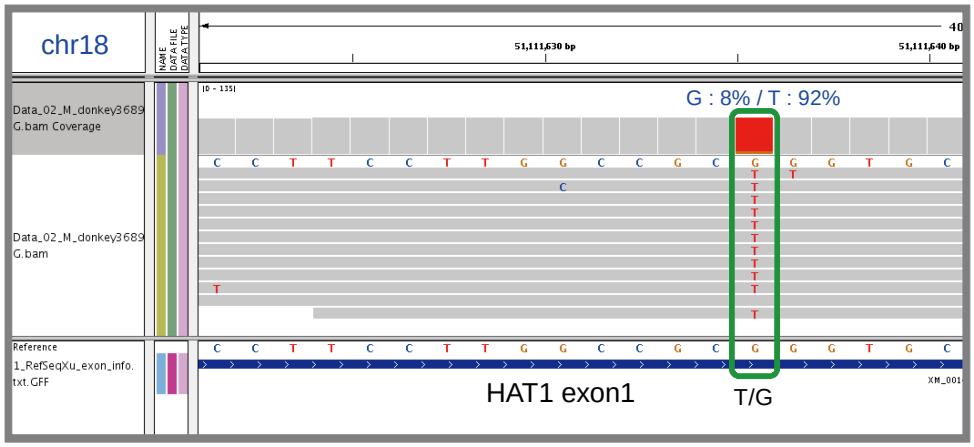


G: 1.9%
T: 98.1%



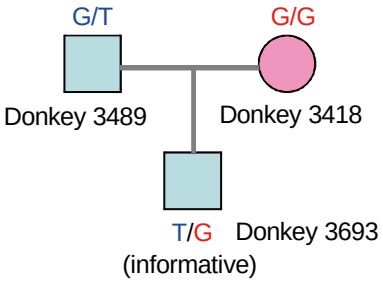
G: 1.1%
T: 98.9%

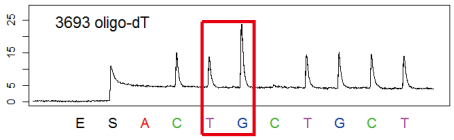
Target sequence: C(G/T)GGTGCTCGC



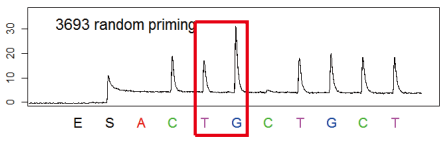
F.

Donkey 3693 (informative heterozygote)





G: 2.1%
T: 97.9%



G: 3.1%
T: 96.9%

Target sequence: C(G/T)GGTGCTCGC

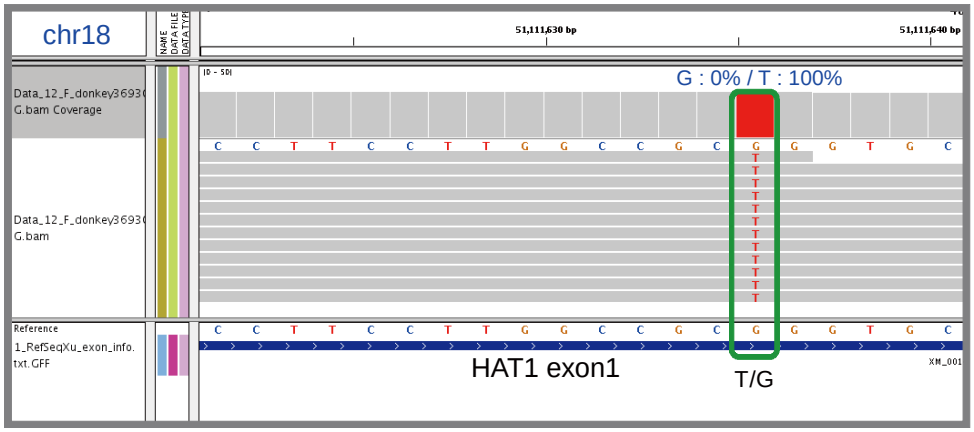


Figure S22. SNP discovery and verification of *INSR* imprinting status in horse x horse crosses.

(A) Horse CG and parental gDNA genotyping results for horse SNP CUHSNP00055036. The heterozygous positions in the CG samples are highlighted.

(B) Quantification of allelic expression ratio in one homozygote control (horse 4060), and three informative heterozygotes (horse 3885, 4061 and 4096) by allele-specific pyrosequencing. The allelic expression percentages are calculated by the relative signal intensity after background subtraction. For horse 3885 and 4096, the G allele is transmitted from the father. For horse 4061, the A allele is transmitted from the father. In all three individuals, we observed preferentially paternal expression, with 15-35% leakage from the maternal allele.

A.

INSR					CUHSNP00055036		
Lane #.	ID	Cross	Father	Mother	genotype	Father	Mother
1.	3881	3474 x 3641	3474	3641	NA	G/A	G/G
2.	3885	3474 x 3640	3474	3640	G/A	G/A	G/G
3.	3896	3474 x 3725	3474	3725	G/G	G/A	NA
4.	3907	3474 x 3641	3474	3641	G/G	G/A	G/G
5.	3948	3474 x 3641	3474	3641	G/G	G/A	G/G
6.	3960	3474 x 3640	3474	3640	G/G	G/A	G/G
7.	3961	3474 x 3725	3474	3725	G/G	G/A	NA
8.	3994	3474 x 3640	3474	3640	G/G	G/A	G/G
9.	4057	3474 x 3640	3474	3640	G/A	G/A	G/G
10.	4060	3474 x 3641	3474	3641	G/G	G/A	G/G
11.	4061	3475 x 3639	3475	3639	G/A	G/G	G/A
12.	4070	3474 x 3640	3474	3640	G/G	G/A	G/G
13.	4077	3474 x 3641	3474	3641	G/A	G/A	G/G
14.	4080	3474 x 3725	3474	3725	G/G	G/A	NA
15.	4083	3474 x 3640	3474	3640	G/G	G/A	G/G
16.	4096	3474 x 3641	3474	3641	G/A	G/A	G/G
17.	4098	3474 x 3725	3474	3725	NA	G/A	NA

B.

INSR				CUHSNP00055036		Pyro		
ID	Father	Mother	genotype	Father	Mother	G%	A%	pat:mat
4060	3474	3641	G/G	G/A	G/G	100.00%	0.00%	-
3885	3474	3640	G/A	G/A	G/G	15.48%	84.52%	85:15
4061	3475	3639	G/A	G/G	G/A	63.59%	36.41%	65:35
4096	3474	3641	G/A	G/A	G/G	24.02%	75.98%	75:25

Table S1. Horse, donkey mule and hinny samples and Illumina mRNA-Seq summary.

Sample ID	Species	Tissue	Stage	platform	# of lanes	# reads	read length	total	genome alignments		junction alignments	
									counts	%	counts	%
3879	Horse	CG	D34	GA	20	82501499	44bp	3.6 Gbp	50906563	61.70%	4411370	6.45%
3689	Donkey	CG	D33	GA	10	53391141	44bp	2.3 Gbp	33142526	62.07%	2596278	5.65%
3703	Mule	CG	D33	GA	8	68413399	44bp	3.0 Gbp	39578210	57.85%	3260736	5.21%
3702	Hinny	CG	D33	GA	8	58667110	44bp	2.5 Gbp	33225856	56.63%	2554250	4.84%
3710	Mule	fetus	D33	GAIIx	1	32643560	40bp	1.3 Gbp	22111265	67.74%	2048222	6.51%
4108	Hinny	fetus	D33	GAIIx	1	30411669	40bp	1.2 Gbp	20301210	66.75%	1786222	6.09%
3713	Mule	fetus	D33	GAIIx	1	25383996	40bp	1.0 Gbp	13640226	53.74%	1233320	5.19%

Table S2. Candidate imprinted genes identified in mule and hinny day-33 chorionic girdle samples.

Rank	Gene name	Chr	# of sig SNPs	Exp. Allele	mule3703 ref count	mule3703 alter count	hinny3702 ref count	hinny3702 alter count	p_1	p_2	*P*-value	*q*-value	Status
1	*H19*	chr12	10	M	94136	225	1176	130562.5	99.76%	0.89%	0	0	known_M
2	*IGF2*	chr12	18	P	224.5	33105	27580	15	0.67%	99.95%	0	0	known_P
3	*INS-IGF2*	chr12	17	P	222.5	32287	27287.5	15	0.68%	99.95%	0	0	known_P
4	*PHLDA2*	chr12	8	M	1236.5	1	9.5	897.5	99.92%	1.05%	0	0	known_M
5	*ENSECAG00000016536*	chr27	3	M	4732.5	0.5	9.5	2909.5	99.99%	0.33%	0	0	mtDNA insertion
6	*ENSECAG00000016730*	chr27	8	M	7148	10	29	23331.5	99.86%	0.12%	0	0	mtDNA insertion
7	*IGF2R*	chr31	43	M	2233	801.5	669	2109	73.59%	24.08%	0	0	known_M
8	*PEG10*	chr4	17	P	3	574.5	298	0	0.52%	100.00%	0	0	known_P
9	*MEST*	chr4	6	P	114.5	4141.5	1293.5	0	2.69%	100.00%	0	0	known_P
10	*CSMD1*	chr27	2	M	178	0.5	2.5	1336	99.72%	0.19%	3.4E-233	1.5E-230	mtDNA insertion
11	*PEG3*	chr10	17	P	0	632.5	278	27	0.00%	91.15%	4.4E-208	1.7E-205	known_P
12	*MAGED2*	chrUn	2	M	342.5	0	0	283	100.00%	0.00%	3.6E-186	1.1E-183	X-linked scaffold
13	*HAT1*	chr18	3	P	8.5	292.5	273.5	0	2.82%	100.00%	7.8E-157	5.2E-111	candidate_P
14	*INSR*	chr7	17	P	66.5	358	362.5	0	15.67%	100.00%	4.4E-156	4E-146	candidate_P
15	*TSR2*	chrUn	7	M	286.5	0	0	208.5	100.00%	0.00%	2.6E-145	6.4E-143	X-linked scaffold
16	*MAGED1*	chrUn	3	M	236	0	0	151.5	100.00%	0.00%	3.7E-112	7.1E-110	X-linked scaffold
17	*GNL3L*	chrUn	2	M	226.5	0	1	73	100.00%	1.35%	6.98E-70	1.16E-67	X-linked scaffold
18	*LY6E*	chr9	4	P	124	531	451.5	252.5	19.74%	63.74%	1.91E-66	3.41E-43	
19	*NINJ1*	chr23	9	P	107	257.5	595.5	139.5	29.36%	81.02%	2.93E-63	4.02E-61	
20	*NU1M*	chr10	1	M	49.5	1	0	311.5	98.02%	0.00%	5.4E-61	7E-59	mtDNA insertion
21	*NAP1L4*	chr12	10	M	375.5	141	46.5	213	72.70%	17.92%	1.38E-49	1.46E-47	known_M
22	*KIAA1161*	chr23	9	M	424.5	195.5	336.5	696.5	68.47%	32.58%	4.65E-46	4.43E-44	too many nonsig SNPs
23	*PTGR1*	chr25	2	P	1	51.5	78	17.5	1.90%	81.68%	2.05E-23	1.11E-21	
24	*SNRPN*	chr1	2	P	0	73	20	0	0.00%	100.00%	9.42E-22	5.03E-12	known_P
25	*LY6G6C*	chr20	3	M	91	42	9	93	68.42%	9.00%	9.68E-21	4.64E-19	candidate_M
26	*SSBP1*	chr4	2	P	12.5	72	35	0	14.79%	100.00%	3.21E-20	1.58E-18	candidate_P
27	*D7ERTD715E*	chr1	2	P	0	57	22	0	0.00%	100.00%	5.09E-20	2.42E-18	candidate_P
28	*SLC1A5*	chr10	4	P	23	97	67	19	19.17%	77.91%	1.55E-17	4.9E-13	
29	*PHF8*	chrUn	1	M	32	0	0	22.5	100.00%	0.00%	9.02E-17	3.51E-15	X-linked scaffold
30	*VAT1*	chr11	4	M	150.5	69	60.5	148	68.56%	29.02%	1.76E-16	6.73E-15	inconsistent SNPs
31	*CPPED1*	chr13	2	M	181.5	80.5	101.5	192.5	69.27%	34.52%	1.85E-16	7.01E-15	too many nonsig SNPs
32	*STON1*	chr15	1	P	1	19.5	21	0	4.88%	100.00%	1.34E-11	3.43E-10	candidate_P

Rank	Gene name	Chr	# of sig SNPs	Exp. Allele	mule3703 ref count	mule3703 alter count	hinny3702 ref count	hinny3702 alter count	p_1	p_2	*P*-value	*q*-value	Status
33	*GALM*	chr15	2	M	148.5	44.5	12	37	76.94%	24.49%	1.48E-11	3.75E-10	inconsistent SNPs
34	*IRF2BP1*	chr10	2	P	6	56.5	41.5	21.5	9.60%	65.87%	3.52E-11	8.46E-10	inconsistent SNPs
35	*C12ORF47*	chr8	1	M	23	0	0	13	100.00%	0.00%	5.96E-11	1.42E-09	inconsistent SNPs
36	*DLK1*	chr24	1	P	0	8	53.5	0	0.00%	100.00%	2.96E-10	6.97E-09	known_P
37	*NDN*	chr1	2	P	0	10	25	0	0.00%	100.00%	8.06E-10	1.68E-08	known_P
38	*USP43*	chr11	2	P	16	46	77.5	29	25.81%	72.77%	3.37E-09	6.72E-08	inconsistent SNPs
39	*ENSECAG00000015880*	chr31	1	M	11	0	0	17	100.00%	0.00%	7.64E-09	1.44E-07	X retrotransposed
40	*Gsta4*	chr20	1	P	48.5	111	61.5	29	30.41%	67.96%	7.82E-09	1.47E-07	inconsistent SNPs
41	*MTFR1*	chr9	2	P	28.5	75.5	55	23.5	27.40%	70.06%	8.86E-09	1.65E-07	
42	*MPP5*	chr24	3	P	44	95.5	73	34.5	31.54%	67.91%	1.09E-08	1.98E-07	
43	*MIIP*	chr2	3	P	13	47.5	37	12	21.49%	75.51%	1.24E-08	2.22E-07	
44	*SLK*	chr1	1	P	28.5	71.5	78.5	42	28.50%	65.15%	4.64E-08	7.62E-07	
45	*RCN1*	chr7	2	P	12	37	24.5	4	24.49%	85.96%	5.83E-08	9.34E-07	
46	*PAR-SN*	chr1	1	P	0	14	10	0	0.00%	100.00%	1.05E-07	1.61E-06	known_P
47	*CPNE2*	chr3	1	M	48	23	3	26	67.61%	10.34%	1.26E-07	1.95E-06	
48	*C13orf15*	chr17	2	P	4	36	22	10	10.00%	68.75%	1.45E-07	2.22E-06	
49	*PWWP2B*	chr1	1	P	5	19.5	25.5	4	20.41%	86.44%	4.29E-07	5.92E-06	
50	*NOL8*	chr23	0	M	67.5	33	27	61	67.16%	30.68%	5.81E-07	7.83E-06	
51	*MUC1*	chr5	1	M	11	0	0	10	100.00%	0.00%	9.32E-07	1.21E-05	
52	*KLHL18*	chr16	1	M	10	2	0	18	83.33%	0.00%	9.81E-07	1.27E-05	
53	*COMMD4*	chr1	1	P	6	32	20	7	15.79%	74.07%	1.37E-06	1.73E-05	
54	*CFH*	chr30	1	M	8	0	0	11	100.00%	0.00%	3.35E-06	3.88E-05	candidate_M
55	*MTRR*	chr21	1	P	12	23	28	3.5	34.29%	88.89%	3.41E-06	4.04E-05	
56	*OBSL1*	chr6	0	P	15	32	28	6	31.91%	82.35%	4.91E-06	5.52E-05	
57	*STYX*	chr24	1	M	15	7	3.5	30.5	68.18%	10.29%	5.01E-06	5.58E-05	
58	*MEG3*	chr24	1	M	8	0	0	10	100.00%	0.00%	6.99E-06	7.39E-05	known_M
59	*FKBP5*	chr20	1	M	19.5	6.5	0	10	75.00%	0.00%	9.99E-06	0.000102	
60	*GPSM2*	chr5	1	P	24	59	26.5	9	28.92%	74.65%	1.08E-05	0.000107	
61	*ZBED4*	chr28	1	P	11	39	36	19	22.00%	65.45%	1.15E-05	0.000113	
62	*ITPKC*	chr10	1	M	16	5.5	0	12	74.42%	0.00%	1.27E-05	0.000123	
63	*PTPRA*	chr22	1	P	12	24	13	0	33.33%	100.00%	1.27E-05	0.000123	
64	*YEATS2*	chr19	1	P	1	18	14	5	5.26%	73.68%	1.62E-05	0.000152	

Rank	Gene name	Chr	# of sig SNPs	Exp. Allele	mule3703 ref count	mule3703 alter count	hinny3702 ref count	hinny3702 alter count	p_1	p_2	*P*-value	*q*-value	Status
65	*RNF139*	chr9	1	P	0	13.5	15	7	0.00%	68.18%	2.74E-05	0.000245	
66	*MIPEP*	chr17	1	M	50	25.5	16	40	66.23%	28.57%	4.2E-05	0.000356	
67	*WDR37*	chr29	2	P	36	68.5	51	26	34.45%	66.23%	4.22E-05	0.000357	
68	*ZNF295*	chr26	1	P	0	11	10	3	0.00%	76.92%	5.33E-05	0.000439	
69	*POLR3H*	chr28	1	P	5.5	24	27.5	14	18.64%	66.27%	5.51E-05	0.000451	
70	*SGCE*	chr4	1	P	0	8	7	0	0.00%	100.00%	5.92E-05	0.000479	known_P
71	*PPP2R5C*	chr24	1	M	23.5	10	8	27	70.15%	22.86%	7.76E-05	0.000606	
72	*ZNF189*	chr25	1	M	45	21.5	10	27	67.67%	27.03%	9.96E-05	0.00075	
73	*SGK1*	chr10	0	P	21.5	43	28.5	10.5	33.33%	73.08%	0.000107	0.000798	
74	*ZNF12*	chr13	1	P	1	17	9	4	5.56%	69.23%	0.000158	0.001107	
75	*TRIM13*	chr17	1	P	3	8	11.5	0	27.27%	100.00%	0.000203	0.001371	
76	*P4HA1*	chr1	1	P	7	18	35.5	13.5	28.00%	72.45%	0.000203	0.001371	
77	*COMMD1*	chr15	1	P	17	40	34	18	29.82%	65.38%	0.000255	0.001657	known_M
78	*C17orf68*	chr11	1	M	20.5	5	7	17	80.39%	29.17%	0.000261	0.001684	
79	*CECR2*	chr6	0	P	25	51	29.5	15	32.89%	66.29%	0.000361	0.002212	
80	*RAP1A*	chr3	1	M	49	25	19	37	66.22%	33.93%	0.000364	0.002229	
81	*KRR1*	chr28	1	M	21	7	2	10	75.00%	16.67%	0.000475	0.002796	
82	*LRCH3*	chr19	0	P	15	38.5	19.5	9	28.04%	68.42%	0.000487	0.002849	
83	*ASH1L*	chr5	1	P	7	23	8.5	2	23.33%	80.95%	0.000518	0.002997	
84	*ZNF777*	chr4	1	P	11	21	40.5	15	34.38%	72.97%	0.000667	0.003645	
85	*BMP4*	chr24	1	M	14.5	5	2	11	74.36%	15.38%	0.000733	0.003949	
86	*PHPT1*	chr25	0	M	50.5	25	16	30	66.89%	34.78%	0.00075	0.004014	
87	*NEK4*	chr16	0	P	3	13	11	3	18.75%	78.57%	0.001098	0.005538	
88	*C22ORF39*	chrUn	0	P	5	14	11	2	26.32%	84.62%	0.001133	0.005676	
89	*PRIM1*	chr6	0	M	28.5	11	2	8	72.15%	20.00%	0.001642	0.007664	
90	*EXTL2*	chr5	0	P	15	36	19	9	29.41%	67.86%	0.001762	0.008127	
91	*C6orf203*	chr10	0	P	9	22.5	18	8	28.57%	69.23%	0.001956	0.008857	
92	*GDAP2*	chr5	0	M	41	14	3	9	74.55%	25.00%	0.001995	0.008981	
93	*SIKE1*	chr5	0	M	17	5	2	8	77.27%	20.00%	0.002087	0.009279	

Table S3. Comparison of known imprinted genes in mouse and equine placenta.

No.	horse gene name	Exp. Allele	Imprinted in equid CG?	p1:p2 in RNA-seq	Imprinted in equid fetus?	p1:p2** in fetus	mouse gene name	human gene name	Imprinted in mouse placenta?	p1:p2 in RNA-seq
1	*DLK1*	P	+	0:100	+	0:100	*Dlk1*	*DLK1*	+	0:100
2	*MEG3*	M	+	100:0	+	95:0	*Gtl2*	*MEG3*	+	100:4
3	*H19*	M	+	100:0	+	100:0	*H19*	*H19*	+	100:0
4	*IGF2*	P	+	0:100	+	0:100	*Igf2*	*IGF2*	+	0:100
5	*IGF2R*	M	+	74:24	+	70:15	*Igf2r*	*IGF2R*	+	100:3
6	*INS*	P	+	0:100	+	0:100	*Ins2*	*INS*	+	0:100
7	*MEST*	P	+	3:100	+	10:90	*Mest*	*MEST*	+	0:100
8	*NAP1L4*	M	+	73:18	-	45:40	*Nap1l4*	*NAP1L4*	?	$
9	*NDN*	P	+	0:100	+	10:100	*Ndn*	*NDN*	+	20:94
10	*PEG10*	P	+	0:100	+	0:100	*Peg10*	*PEG10*	+	0:100
11	*PEG3*	P	+	0:100	+	0:100	*Peg3*	*PEG3*	+	3:100
12	*PHLDA2*	M	+	100:1	+	100:5	*Phlda2*	*PHLDA2*	+	80:20$$
13	*SGCE*	P	+	0:100	+	35:63	*Sgce*	*SGCE*	+	7:91
14	*SNRPN*	P	+	0:100	+	15:85	*Snrpn*	*SNRPN*	+	2:91
15	*SNURF*	P	+	0:100	+	15:85	*Snurf*	*SNURF*	+	2:91
16	*PAR-SN*	P	+	0:100	+	10:100	*No ortholog*	*PAR-SN*	No ortholog	-
17	*AMPD3*	M	-	19:11	low expr.		*Ampd3*	*AMPD3*	-	18:63
18	*ANKRD11*	M	-	47:46	-	51:57	*Ankrd11*	*ANKRD11*	-	49:48
19	*BLCAP*	M	-	45:28	partial	58:31	*Blcap*	*BLCAP*	-	50:55
20	*CD81*	M	-	43:23	-	53:51	*Cd81*	*CD81*	No SNP	-
21	*COMMD1*	M	-	41:43*	-	61:73	*Commd1*	*COMMD1*	-	48:47
22	*COPG2*	P	-	11:42	low expr.		*Copg2*	*COPG2*	-	54:53
23	*DCN*	M	-	40:43	-	49:34	*Dcn*	*DCN*	-	53:53
24	*DHCR7*	M	-	51:64	-	73:55	*Dhcr7*	*DHCR7*	low expr.	-
25	*DLX5*	M	-	50:48	low expr.		*Dlx5*	*DLX5*	low expr.	-
26	*IMPACT*	P	-	50:59	-	43:52	*Impact*	*IMPACT*	+	8:99
27	*MCTS2*	P	-	46:53	-	48:56	*Mcts2*	*NA*	No SNP	-
28	*PDE4D*	P	-	40:52	-	85:83	*Pde4d*	*PDE4D*	-	56:41
29	*PON2*	M	-	38:52	low coverage		*Pon2*	*PON2*	-	66:33
30	*PPP1R9A*	M	-	53:70	partial	79:42	*Ppp1r9a*	*PPP1R9A*	+	82:15
31	*TH*	M	-	eQTL	-	40:61	*Th*	*TH*	low expr.	-
32	*TRAPPC9*	M	-	60:36	-	51:49	*Trappc9*	*TRAPPC9*	-	56:38
33	*TSSC4*	M	-	42:39	low expr.		*Tssc4*	*TSSC4*	+	100:0$$
34	*UBE3A*	M	-	44:50	-	45:55	*Ube3a*	*UBE3A*	-	54:54
35	*XIST*	P	-	biallelic	-	biallelic	*Xist*	*XIST*	+	0:100
36	*ASCL2*	M	No SNP				*Ascl2*	*ASCL2*	No SNP	-
37	*CDKN1C*	M	No SNP				*Cdkn1c*	*CDKN1C*	+	100:3
38	*IGF2AS*	P	No SNP				*Igf2as*	*IGF2AS*	+	4:100
39	*KCNQ1OT1*	P	No SNP				*Kcnq1ot1*	*KCNQ1OT1*	low expr.	-
40	*TCEB3C*	M	No SNP				*NA*	*TCEB3C*	No ortholog	-

The list of known imprinted genes is from [7]. The mouse placenta imprinting data is from [23]. * Allelic expression ratios from pyrosequencing verification. ** 1-16 are from pyrosequencing verifications, others are from RNA-seq data. $: The imprinting of *Nap1l4* may be partial or isoform specific (27). $$: These genes are annotated manually.

Table S4. Selection of mule and hinny CG and fetus samples for pyrosequencing verification.

ID	Sample ID	Tissue	Stage	Mother	Father	Mother ID	Father ID
horse1	horse3879CG	CG	D34	horse	horse	3558	3475
donkey1	donkey3689CG	CG	D33	donkey	donkey	3418	3489
mule1	mule3703CG	CG	D33	horse	donkey	3639	3611
hinny1	hinny3702CG	CG	D33	donkey	horse	3524	3475
mule2	mule3700CG	CG	D33	horse	donkey	3638	3611
hinny2	hinny4108CG	CG	D33	donkey	horse	2175	0834
mule3	mule3710CG	CG	D33	horse	donkey	3638	3611
mule4	mule3713CG	CG	D33	horse	donkey	2994	3611
horse1	horse3879fetus	fetus	D34	horse	horse	3558	3475
donkey1	donkey3689fetus	fetus	D33	donkey	donkey	3418	3489
mule1	mule3703fetus	fetus	D33	horse	donkey	3639	3611
mule2	mule3700fetus	fetus	D33	horse	donkey	3638	3611
hinny2	hinny4108fetus	fetus	D33	donkey	horse	2175	0834
mule3	mule3710fetus	fetus	D33	horse	donkey	3638	3611
mule4	mule3713fetus	fetus	D33	horse	donkey	2994	3611
mule5	mule3680fetus	fetus	D34	horse	donkey	3424	1992

Table S5. Pyrosequencing verification results for six novel imprinted genes in mule and hinny chorionic girdle and fetus samples.

Rank	Gene name	Exp. allele	RNA-Seq p_1	RNA-Seq p_2	q-value	RNA-Seq ratio	variable?	pyro p_1	pyro p_2	Pyro ratio	M1	H1	M2	H2	M3	M4	status in fetus	M1	H2	M2	M3	M4	M5
13	*HAT1*	P	2.8%	100.0%	5.24E-111	0:100	NO	9.30%	100.00%	10:100	I	I	I	I	I	I	not imprinted	B	B	B	B	B	B
14	*INSR*	P	15.7%	100.0%	3.96E-146	15:100	YES	14.30%	100.00%	15:100	I	I	B	I	B	I	not imprinted	B	B	B	B	B	B
25	*LY6G6C*	M	68.4%	9.0%	4.64E-19	70:10	NO	84.0%	11.0%	85:10	I	I	I	I	I	NA	not covered	-	-	-	-	-	-
27	*D7ERTD715E*	P	0.0%	100.0%	2.42E-18	0:100	NO	0.00%	100.00%	0:100	I	I	I	I	I	I	imprinted_P	I	I	I	I	I	I
32	*STON1*	P	4.9%	100.0%	3.43E-10	5:100	YES	18.90%	90.40%	20:90	I	I	B	I	B	I	not imprinted	B	B	B	B	B	B
54	*CFH*	M	100.0%	0.0%	3.88E-05	100:0	NO	100.00%	0.00%	100:0	I	I	I	I	I	I	not covered	-	-	-	-	-	-

I: consistent with genomic imprinting. B: biallelic expression (within 40:60).