Whole-genome sequencing of CRFK and PG-4 cells to infer the phenotype of the original donor cats

Ganma Tanaka[1]*, Rikuto Goto[1]*, Tetsushi Komoto[1,2], Akiko Kubota[1], Reo Hayashi[1], Takeshi Igawa[1,2], Naoaki Sakamoto[1,3], Akinori Awazu[1,3]

[1]Graduate School of Integrated Sciences for Life, Hiroshima University, Higashihiroshima, Hiroshima, Japan

[2]Amphibian Research Center, Hiroshima University, Higashihiroshima, Hiroshima, Japan

[3]Research Center for the Mathematics on Chromatin Live Dynamics, Hiroshima University, Higashihiroshima, Hiroshima, Japan

* Ganma Tanaka and Rikuto Goto contributed equally to this work and share first authorship.

Corresponding author:

Akinori Awazu

Graduate School of Integrated Sciences for Life, Hiroshima University, 1-3-1, Kagamiyama, Higashi-Hiroshima 739-8526, JAPAN,

email address: awa@hiroshima-u.ac.jp

Abstract

Background Crandell-Rees Feline Kidney (CRFK; kidney-derived) cells and PG-4 cells (astrocyte-derived) have been in use and have been passaged for decades in laboratories worldwide; however, no detailed information on the genetic background of the donor individuals is available, particularly regarding phenotype characteristics such as coat and iris color.

Results We performed whole-genome sequencing of CRFK and PG-4 cells. We analyzed the resulting data to infer the phenotype of the individual from which the cells were derived, specifically for the coat color, coat length, coat pattern, and iris color. Our data suggested that CRFK cells originated from a cat with long black fur lacking stripes and with non-blue irises; PG-4 cells originated from a cat with long bicolored white and black fur without stripes, and with non-blue irises. Analysis of publicly available RNA-seq data confirmed that genes associated with coat phenotype and iris color are expressed in the skin and eyes, as well as in various other organs.

Conclusions Variants of the genes affecting coat phenotype and iris color may influence physiological functions throughout the body. These results shed light on the previously unknown genetic background of commonly used feline cultured cells and the phenotype of the donor individuals. These findings may facilitate more accurate interpretation of data obtained from cultured feline cells and provide guidelines for developing cell lines with domestic cat genotypes exhibiting diverse phenotypes through genome editing. This will help to provide clarity regarding deafness in cats with white fur and blue irises and elucidate the effects of melanocyte destruction caused by KIT gene mutations on the nervous system and the influence of other coat-related factors in organs other than the skin or functions independent of coat formation.

Keywords: genomics; coat pattern; iris color; RNA sequencing;

Background

Domestic cats (*Felis catus*) are widely kept as companion animals around the world, and the phenotypic diversity of their coat color, coat pattern, coat length, and iris color has been a topic of interest to breeders and scientists alike. Recent advances in genomics have markedly improved our understanding of the genetic basis controlling these diverse phenotypes, and the respective genotype-phenotype relationships are being increasingly elucidated.[1] Examples include the *TYR* gene, which is the color locus associated with albinism and a color point phenotype[2–4]; the *TYRP1* gene, conveying brown coat color[2,5]; the *MC1R* genes, associated with in red coat color[6]; the *KIT* gene, determining white coat and white spotting[7–11]; the *ARHGAP36* gene, involved in orange coat color[12,13]; the *MLPH* gene, causing dilute pigmentation[14]; the *ASIP* gene, controlling the expression of the agouti pattern[15]; the *LVRN* gene, determining the type of tabby pattern[16]; the *DKK4* gene, associated with the ticked tabby pattern[17]; the *FGF5* gene, involved in long-fur variants[18,19]; the *KRT71* and *LPAR6* genes, involved in hairless and rexing (curly hair) phenotypes[20–22]; and the *PAX3* gene, which is associated with blue irises[23,24]. Furthermore, technological advancements such as long-read sequencing facilitated the construction of a more continuous and accurate draft genome of the domestic cat[25–27], and a plethora of genetic polymorphisms associated with disease have been identified [17,25,28–32].

Immortalized cell lines are an indispensable tool in current biomedical and veterinary science. Crandell-Rees Feline Kidney (CRFK) cells, derived from the kidneys of a domestic cat[33–34], have been used for numerous virological studies, including feline calicivirus,[35–37] feline immunodeficiency virus[38–41], feline coronavirus[42–45], and various feline endogenous retroviruses such as RD114[46-48], as well as for vaccine development[49–51]. PG-4 cells, derived from feline fetal astrocytes, have also played an important role in retrovirus research[52–53].

Additionally, Crandell et al.[33] reported that CRFK cells exhibited aneuploidy (2n = 37) compared with the normal feline autosome count of 2n = 38, and demonstrated a characteristic karyotype with three marker chromosomes. Lasfargues et al.[54] also observed that the chromosome count fluctuated between 36 and 38 in each cell. By contrast, no such findings have been reported for PG-4 cells.

These cell lines have been passaged for decades in laboratories worldwide and have facilitated numerous scientific breakthroughs. However, despite their widespread use, no detailed information on the genetic background of the donor individuals, particularly with regard to their phenotype, such as coat and iris color, is available. Genomic information on such cell lines is

crucial for a deeper understanding of their biological properties, and elucidating the phenotypes of their donors may provide new insights into the influence of specific genotypes on cell properties.

In this study, we performed whole-genome sequencing of CRFK and PG-4 cell lines. The sequencing data were used for in-depth analysis of the nucleotide sequences of genes known to be associated with coat and iris phenotypes (i.e., *TYR*, *TYRP1*, *MC1R*, *KIT*, *ARHGAP36*, *MLPH*, *ASIP*, *LVRN*, *DKK4*, *FGF5*, *KRT71*, *LPAR6*, and *PAX3*; Table 1). Based on the genotyping results for each gene, we inferred the previously unknown phenotypes of the donor individuals. Furthermore, analysis of publicly available RNA-seq data confirmed that these genes were expressed in skin and eyes as well as in various other tissues throughout the body. The findings and genome sequence data are expected to provide fundamental insights into the genetic background of extensively used feline cell lines. This information will support the interpretation of future research outcomes involving these cell lines and contribute to the development of guidelines for establishing cell lines with various genotypes through genome editing.

Methods

Cell lines and culture conditions

CRFK cells (JCRB9035, RRID: CVCL_2426) and a feline sarcoma-positive leukemia-negative (S+L−) astrocyte cells termed PG-4 (S+L-) (JCRB9125, RRID: CVCL_3322) were obtained from the JCRB Cell Bank (National Institutes of Biomedical Innovation, Health and Nutrition, Osaka, Japan). CRFK cells were cultured in Dulbecco's Modified Eagle Medium (Gibco, Thermo Fisher Scientific, Waltham, MA, USA) supplemented with 10% fetal bovine serum (FBS) (Gibco), penicillin (100 IU/mL), streptomycin (100 ng/mL) (Sigma-Aldrich, St. Louis, MO, USA), and 1% non-essential amino acids (FUJIFILM Wako Pure Chemical Corporation, Osaka, Japan) at 37 °C and 5% CO_2. PG-4 cells were cultured in McCoy's 5A Modified Medium (Gibco) supplemented with 10% heat-inactivated FBS (Gibco), penicillin (100 IU/mL), and streptomycin (100 ng/mL) (Sigma-Aldrich) at 37 °C and 5% CO_2. For cell passaging, cells were washed using BASIC DPBS (no calcium or magnesium; Gibco) and then treated with TrypLE™ Express Enzyme (1X; Gibco) for 5 min at 37 °C and 5% CO_2 conditions.

Paired-end short-read whole-genome sequencing

Genomic DNA was extracted from logarithmically growing cells (approximately 1×10^6 cells) using the DNeasy Blood & Tissue Kit (QIAGEN, Hilden, Germany) according to the manufacturer's protocol. The concentration and purity (A260/A280 ratio) of the extracted DNA were measured using a NanoDrop 2000c Spectrophotometer (Thermo Fisher Scientific). Genomic DNA extracted from CRFK cells and PG-4 cells was used for library preparation and paired-end sequencing (150 bp read length) at Novogene (Beijing, China) using a Rapid Plus DNA Lib Prep Kit for Illumina V2 (Illumina) and an Illumina NovaSeq X Plus platform (Illumina). Library quality control was performed by Novogene using qPCR and fragment size analysis according to Novogene's standard protocols. The raw sequence data were obtained as .fastq files.

Mapping of paired-end reads and quantification

Adapter sequences were trimmed, and low-quality reads were removed from the original .fastq files using fastp software (ver. 0.21.0) with default parameters. Reads in processed .fastq files were mapped to the reference genome sequence of *Felis catus*, F.catus_Fca126_mat1.0 (RefSeq assembly GCF_018350175.1) or AnAms1.0 (GenBank assembly GCA_013340865.2)[27] using bwa mem software (ver. 0.7.19) with the setting "-R "@RG¥tID:sample1¥tSM:sample1¥tLB:lib1¥tPL:ILLUMINA"" and default parameters otherwise. Mapping results were obtained as .sam files. Text data in .sam files were transformed to binary data in sorted .bam files using Samtools (ver. 1.7) with default parameters. Index files were created for the obtained .bam files in the form of .bam.bai files.

Small variations in coat morphology-related genes

The variations of coat morphology-related phenotypes of the donor cats from which CRFK and PG-4 cell lines originated were inferred based on the comparisons between the wild-type sequences and mapped sequences on variant positions of each coat morphology-related gene (Table 1). Coat phenotype-related polymorphisms in *TYR*, *TYRP1*, *MC1R*, *KIT*, *MLPH*, *ASIP*, *LVRN*, *DKK4*, *FGF5*, *KRT71*, and *LPAR6* genes, which relate to phenotypes such as the color, brown, red, white glove, dilution, agouti, tabby, ticked, long-fur, hairless, and rexing (Table 1), were examined using alignments of paired-end sequence reads with wild-type sequences.

Large deletions in intron 1 of ARHGAP36 and the intergenic region between *KIT* and *KDR* genes

The deletion of a part of intron 1 of the *ARHGAP36* gene (X: 109,186,183_109,191,258 in

143 AnAms1.0; Table 1) was reported to produce orange coat color[12,13], and that of the intergenic
144 region between the *KIT* and *KDR* genes (B1: 161142880_161237957 in F.catus_Fca126_mat1.0,
145 Table 1) produces salmiak coat color[11]. The occurrence of these deletions in CRFK and PG-4
146 was estimated by comparing the mapped read number distributions between these regions and
147 their neighboring regions with the same region length. It should be noted that
148 F.catus_Fca126_mat1.0 has a deletion of the genomic region corresponding to X:
149 109186183_109191258 in AnAms1.0 (Table 1). Thus, AnAms1.0 (or felCat9[25]) was used as the
150 reference genome for analysis of the orange locus.

151 **Insertions of FERV1-LTR and RD114-LTR into *PAX3***

152 The insertions of homologous sequences of the long terminal repeat (LTR) of the feline
153 endogenous retrovirus 1 (FERV1), termed FERV1-LTR, into a specific region in intron 4 of
154 *PAX3* (C1: g.205833101_205833102 in F.catus_Fca126_mat1.0, Table 1) or the LTR of RD114
155 retrovirus, termed RD114-LTR, into another specific region in the same intron (C1:
156 g.205834854_205834855 in F.catus_Fca126_mat1.0, Table 1) are associated with blue iris color
157 in cats.[23,24] The occurrence of such insertions was estimated by extracting the genome reads that
158 were mapped in both FERV1-LTR and intron 4 of *PAX3* or both RD114-LTR sequences and
159 intron 4 of *PAX3*. The genome sequence of RD114-LTR was obtained from the NCBI
160 database[55].

161 **Long-read whole-genome sequencing**

162 Genomic DNA of CRFK cells and PG-4 cells was extracted for long-read genome sequencing
163 using the NucleoBond HMW DNA kit (TaKaRa, Tokyo, Japan), and libraries for the Oxford
164 Nanopore Technologies (ONT, UK) sequencer were constructed using a ligation library
165 preparation kit (SQK-LSK114, ONT). All procedures were performed according to the
166 manufacturer's instructions, and the quality and molecular weight of the genomic DNA were
167 measured using the Qubit fluorometric quantification system (Thermo Fisher Scientific).
168 Sequencing was conducted using a PromethION sequencer (PromethION 2 Solo, ONT) and
169 R10.4.1 flow cells. POD5 files were basecalled to .fastq files using MinKNOW software
170 v25.05.14.

171 **Insertions of FERV1 homolog into *KIT* examined by long-read sequencing**

172 The insertion of FERV1 into a specific region of intron 1 of the *KIT* gene (B1:
173 161388687_161388688 in F.catus_Fca126_mat1.0, Table 1) and that of FERV1-LTR make the

174 coat color of cats white-spotted and white throughout, respectively [8]. The occurrence of such
175 insertions in CRFK and PG-4 cells was estimated by mapping the reads of long-read sequencing
176 to the genome sequence named KIT-FERV1 sequence using minimap2 (ver. 2.30), where the
177 KIT-FERV1 sequence is a part of the *KIT* intron 1 sequence with the insertion of the FERV1
178 sequence, as shown in Supplementary Figure 1 of the study by David et al[8].

179 **Transcriptome data analysis of coat phenotype- and iris color-related genes in various**
180 **tissues of domestic cats**

181 Mapped RNA-seq data on the F.catus_Fca126_mat1.0 were obtained in .bam and .bam.bai
182 formats for various tissues from the Ensembl public database[56], specifically, the file transfer
183 protocol (ftp) site[57]. The annotation file of the F.catus_Fca126_mat1.0 in gene transfer format
184 (GTF) was obtained from the NCBI database (Bethesda, MD, USA). In the file, the
185 chromosome numbers were converted from NC_058368.1, NC_058369.1, etc. to A1, A2, etc.

186 Using FeatureCounts with the options --M-O--fraction,[58] read counts data for each gene
187 described in the GTF file were obtained from the .bam and .bam.bai formatted files. This count
188 data could also be obtained from a file
189 "F.catus_Fca126_mat1.0.ENA_gene_exp_mat_by_FeatureCounts.csv" deposited in F.catlas[59].
190 From these counts, the transcript level of each gene was calculated by dividing the read count
191 by the total exon length. The total transcript level was defined as the sum of all the transcript
192 levels of the genes. Based on this result, the transcripts per million (TPM) of each gene were
193 calculated as 1,000,000*(transcript level)/(total transcript level).

194

195 **Results**

196 **Short- and long-read whole-genome sequencing**

197 For CRFK and PG-4 cells, the total output of raw sequence data of 150 bp paired-end
198 sequencing was 60.6 Gb and 60.1 Gb, and the [average] $\pm$ [standard deviation] of mapping
199 depth of the reads for F.catus_Fca126_mat1.0 were 23.59 ± 9.79 and 23.45 ± 8.17,
200 respectively; the total output of raw sequence data of long-read sequencing was 9.6 Gb and 11.8
201 Gb, N50 values were 23.5 kb and 38.2 kb, and the [average] $\pm$ [standard deviation] of mapping
202 depth of reads for F.catus_Fca126_mat1.0 were 3.91 ± 3.11 and 4.79 ± 3.34, respectively.
203 The average and standard deviations specified above were estimated through the entire genome
204 region without the regions with singular sequences where the read coverage exhibited >100.

205 **Predicted phenotype of the CRFK cell donor**

206 The mapping of paired-end reads from the whole-genome sequences of the CRFK cell revealed
207 that most coat morphology-related genes were conserved (Table 1). However, the following
208 variations were identified in mapped genome reads from CRFK cells (Table 1): a homozygous
209 mutation of c.123_124del of the *ASIP* gene (*nonagouti* allele), a homozygous mutation of
210 c.475A>C of the *FGF5* gene (*long coat* allele; Fig. 1), a homozygous mutation of c.2522G>A of
211 the *LVRN* gene (*blotched-tabby* allele), and a heterozygous mutation of c.83delT of the *MLPH*
212 gene (*dilute* allele). No 95 kbp deletion was detected in the intergenic region between the *KIT*
213 and *KDR* genes, nor a 5 kbp deletion was observed in intron 1 of the *ARHGAP36* gene (Fig. 2).
214 Additionally, no genome reads were identified that could be simultaneously mapped to both the
215 *PAX3* intron 4 and FERV1-LTR or to both the *PAX3* intron 4 gene and RD114-LTR sequences.
216 Furthermore, four reads from the long-read sequencing mapped to the KIT-FERV1 sequence;
217 however, these reads contained no insertions of FERV1 and FERV1-LTR (Fig. 3).

218 The abovementioned homozygous mutation of the *FGF5* gene was known to exhibit long fur[18].
219 However, the mutation of the *LVRN* did not influence coat morphology in this case because of
220 the mutation of the *ASIP* gene that suppresses the stripe pattern on the cat's skin coat, and the
221 mutant genotype of the *MLPH* gene was recessive.[1] Therefore, the cat from which the CRFK
222 cells originated is expected to have had long, black, unstriped fur and non-blue eyes (Fig. 4).

223 **Predicted phenotype of the PG-4 cell donor**

224 Mapped paired-end reads from the whole-genome sequences of the PG-4 cell indicated that
225 most coat morphology-related genes were conserved but revealed the following variations
226 (Table 1): a homozygous mutation of c.123_124del of the *ASIP* gene, a compound heterozygous
227 mutation of c.474delT and c.475A>C of the *FGF5* gene (Fig. 1) where one allele exhibits
228 c.474delT and the other allele exhibits c.475A>C as mentioned below, the heterozygous
229 mutations of c.416C>A and c.2522G>A of *LVRN* gene, and a heterozygous mutation of
230 c.904G>A of the *TYR* gene (*Siamese-color* allele). No 95 kbp deletion was identified in the
231 intergenic region between the *KIT* and *KDR* genes, nor a 5 kbp deletion was found in intron 1 of
232 the *ARHGAP36* gene (Fig. 2), as was similarly observed with CRFK. Additionally, we did not
233 identify any genome reads that mapped concurrently to both the *PAX3* intron 4 and FERV1-LTR
234 or to both the *PAX3* intron 4 gene and RD114-LTR sequences.

235 Notably, the genome sequence reads obtained from PG-4 cells containing bases 474 and 475 of

236 the CDS of the FGF5 gene revealed that seven reads exhibited only a deletion at base 474, while
237 the remaining 13 reads showed only a substitution at base 475 (Fig. 1). Accordingly, we
238 expected a compound heterozygous mutation in the FGF5 gene of PG-4 cells. Similarly, PG-4
239 cells exhibited heterozygous substitutions at bases 416 and 2522 of the CDS of the LVRN gene.
240 However, owing to the considerable distance between the genomic locations of these two bases,
241 none of our sequencing data, including the long-read sequencing data, produced reads that
242 contained both bases together. Therefore, we were unable to conclude whether the LVRN gene
243 in PG-4 cells harbors a compound heterozygous mutation.

244 Six long-read sequences mapped to the KIT-FERV1 sequence, with four of them containing the
245 insertions of the FERV1 homolog sequence exceeding 7 kbp (Fig. 3). This indicates that the PG-
246 4 genome contains the heterozygous insertion of the entire FERV1 homolog sequence.

247 The mutant genotype of the *TYR* gene was recessive, and in this case, the two heterozygous
248 mutations of the *LVRN* did not influence coat morphology because of the *ASIP* gene mutation,
249 the same as in CRFK[1]. However, the effect of *FERV1* is known to be dominant[8]. Additionally,
250 the compound heterozygous mutation of the *FGF5* gene produces long fur[18]. Therefore, the PG-
251 4 cells likely originated from a cat with white and black (white spotting), unstriped, long fur
252 (Fig. 4).

253 Notably, certain cats with white spotting may have blue eyes or atypical eye coloration;
254 however, this occurrence is less common than that in cats exhibiting the Dominant White allele
255 in the KIT gene[8]. Additionally, no insertions of FERV1-LTR and RD114-LTR in intron 4 of the
256 PAX3 gene were observed. Therefore, the PG-4 cells likely originated from a cat with non-blue
257 irises (Fig. 4).

258 **Expression features of coat phenotype- and iris color-related genes in various tissues**

259 The analysis of publicly available RNA-seq data from various organs of domestic cats in the
260 Ensembl database
261 (https://ftp.ensembl.org/pub/data_files/felis_catus/F.catus_Fca126_mat1.0/rnaseq/) revealed that
262 coat phenotype-related genes are expressed in various organs at levels (TPM) comparable to or
263 greater than those in the skin (Fig. 5, Table S2). For example, *TYR*, *TYRP1*, and *MLPH* were
264 highly expressed in the ear tip, retina, and optic nerve, and *MC1R*, *KIT*, *DKK4*, and *FGF5* were
265 highly expressed in the brain. *KIT*, *DKK4*, and *LPAR6* tend to be expressed throughout the body,
266 with prominent expression in the reproductive organs. Additionally, *ARHGAP36* was highly

expressed in the spinal cord, and *ASIP*, along with *MLPH*, was highly expressed in the lungs. The iris color-related gene *PAX3* was also highly expressed in the cerebellum, in addition to visual organs, and, along with *DKK4*, in the kidney and embryonal tissue.

Discussion

We performed whole-genome sequencing of CRFK, a cultured feline kidney cell line, and PG-4, a feline fetal astrocyte-derived cell line, and analyzed the genome data to determine the coat color, hair length, coat pattern, and iris color of the donor individuals. The results suggested that the CRFK cell line originated from a cat with long black hair without stripes and non-blue irises, and the PG-4 cell line came from a cat with long fur, black and white bicolor coat without stripes, and non-blue irises.

Coat morphology and iris color are phenotypes that can be readily identified and used to predict genotypes. However, the expressions of these genes are not limited to the skin and eyes; rather, they are expressed in various tissues throughout the body (Fig. 5). Therefore, differences in these genotypes may contribute to disease susceptibility in specific organs. The effects of these genetic variations are likely weaker than those of the rearing environment and may not be significant, especially in young, healthy individuals. However, as domestic cats have become longer-lived in recent years, individual-specific risks may become more apparent, and the underlying genotypes may exert increasing effects as the cats age. Therefore, elucidating the relationships between genotypes underlying readily recognizable phenotypes and disease risk is crucial from the perspectives of both medical care and welfare, such as prevention and symptom suppression/alleviation.

Results from research on cultured cells to elucidate disease mechanisms and develop therapeutic drugs must be interpreted with caution, i.e., not only on the experimental results obtained using established, immortalized cells but also with regard to their genotype and expression profiles of the genes harboring polymorphisms in various tissues and organs. However, so far, such information has not been available, even for certain frequently used cell lines such as CRFK. The present study and genome sequence data provide such information and facilitate a more comprehensive interpretation of research on feline cultured cell lines.

Furthermore, with the availability of detailed genomic information, genome editing can be applied to these cell lines to establish various derivative lines that differ only in specific aspects,

298 e.g., regarding coat morphology or other traits, and compare their properties. This is expected to
299 advance research on the regulatory mechanisms of coat color, for example, the validation of the
300 role of the novel *KIT* isoform caused by FERV1-LTR insertion in the formation of white coat[60],
301 as well as the relationships between the visually identifiable phenotypes examined in the present
302 study and disease risk and treatment methods. Simultaneously, this comprehensive genomic
303 information can help to develop guidelines for establishing cell lines with various disease-
304 associated genotypes using genome editing. The development of these derivative cell lines may
305 provide valuable insights into the mechanism underlying deafness in cats with white fur and
306 blue irises, as well as elucidate the influence of other coat-related factors in organs beyond the
307 skin or functions unrelated to coat formation.

308 The limitations and future directions of this study are outlined below.

309 First, notably, the findings of this study are specific to the cell types used. The cells used in this
310 study were CRFK cells (JCRB9035, RRID: CVCL_2426), which were prepared from CRFK
311 (ATCC: CCL 94) and PG-4 (S+L-) cells (JCRB9125, RRID: CVCL_3322) obtained from PG-4
312 (S+L-) cells (ATCC: CRL-2032). Consequently, these cells were expected to have similar
313 characteristics to those of the CRFK and PG-4 cells used in numerous previous studies. For
314 example, however, the number of chromosomes has been reported to vary from 36 to 38 in
315 CRFK[33,54]. Additionally, the karyotype of PG-4 cells, which has not yet been reported, likely
316 differs from the normal domestic cat genome. Therefore, using the same cell line may still
317 produce results that differ slightly from those used by other researchers. Investigating additional
318 variation among cells in these lines remains an important future consideration.

319 As mentioned previously, the karyotype of CRFK differs from that of the normal domestic cat
320 genome[33,54]. Additionally, the specific chromosomes that cause aneuploidy remain unidentified.
321 However, identifying large structural variations using only whole-genome analysis of the
322 present sequence read data, approximately 60 Gb paired-end sequencing data and 10 Gb long-
323 read data, was challenging. Similarly, accurately identifying sequences and locations of various
324 long repeat sequences, endogenous retroviral sequences, and their homologs was challenging
325 owing to limited read coverage, particularly for long-read coverage. However, these sequences
326 may also be associated with various diseases, as demonstrated by the involvement of FERV1-
327 LTR in deafness in cats[8,23,24] and the positive and negative roles of enFeLV-LTR and enFeLV-
328 env in resistance to exogenous feline leukemia virus[61-63]. The increased volume of ongoing
329 long-read sequencing, along with the draft genome assembly using these data, is expected to

330 address these challenges in the future.

331 This study focused on coat and iris phenotypes in domestic cats. However, major
332 histocompatibility complex (MHC) polymorphisms are strongly correlated with breed in
333 domestic dogs[64-68]. Although research on MHC polymorphisms in domestic cats is ongoing[69-71],
334 the relationship with breed remains unclear. In this study, we showed the results of the primary
335 analysis of the small indels and SNPs in MHC class I and II regions of CRFK and PG-4 cells
336 described by Variant Call format (VCF) [72] (Table S3). However, because the MHC gene cluster
337 comprises a complex and large genomic region, it is difficult to identify the donor breeds of
338 CRFK and PG-4 using only the sequence data and analytical methods presented in this study.
339 However, as mentioned above, further long-read sequencing and its data analysis may enable
340 breed estimation of these cell lines.

341 In addition, the present genome sequence data also included mitochondrial DNA data. In this
342 study, we presented the results of the primary analysis of the small indels and SNPs in the
343 mitochondrial DNA of CRFK and PG-4 cells described by VCF[72] (Table S4[72]). Although the
344 correlations are not substantially strong, an association may exist between the polymorphisms of
345 mitochondrial DNA and breeds of domestic cats[73-76]. Therefore, a comprehensive analysis of
346 this data may also contribute to breed estimation.

347 These analyses are also critical future challenges.

348

349 **Conclusions**

350 We performed whole-genome sequencing of CRFK and PG-4 cells and inferred the phenotype
351 of the donors of these cells. We suggested that CRFK cells originated from a cat with long,
352 black fur lacking stripes and non-blue irises; PG-4 cells originated from a cat with long,
353 bicolored white and black fur without stripes, and with non-blue irises. Additionally, analysis of
354 publicly available RNA-seq data confirmed that genes associated with coat phenotype and iris
355 color are expressed in the skin and eyes, as well as in various other organs, indicating that
356 variants of these genes, which affect coat phenotype and iris color, may influence physiological
357 functions throughout the body. These insights may facilitate a more accurate interpretation of
358 data derived from feline cultured cells and inform guidelines for developing cell lines with
359 domestic cat genotypes that exhibit various phenotypes using genome editing techniques. This
360 will help to elucidate the role of various coat phenotype-related and iris color-related factors in

organs other than the skin and functions independent of coat formation, including the effects of melanocyte destruction caused by KIT gene mutations on the nervous system.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

These raw read data of the short and long read sequencings were deposited in the DDBJ Sequence Read Archive (DRA) with BioProject accession PRJDB37530 (DRR748391-DRR748394, DRR786948, and DRR786949). All other data generated or analyzed during this study were included in this published article (and Supplementary Information files).

Conflicting interests

The authors have no conflicting interests to declare.

Funding

This work was supported by Japan Society for the Promotion of Science (JSPS) KAKENHI grants (award numbers 24K01783 (T. I.) and 24K09248 (A. A.)).

Author contributions

Conceptualization: A.A., Data curation: T.K., A.K., R.H., and T.I., Formal analysis: G.T. and R.G., Funding acquisition: A.A., Investigation: G.T., R.G., T.K., R.H., and A.A., Methodology:

G.T., R.G., T.K., T.I., and A.A., Project administration: N.S. and A.A., Supervision: N.S. and A.A., Validation: R.G. and A.A., Visualization: T.G., N.S., and A.A., Writing - original draft: A.A., Writing - review & editing: T.G., T.K., T.I., and N.S.

Acknowledgements

Not applicable.

References

1. Lyons LA. DNA mutations of the cat: The good, the bad and the ugly: The good, the bad and the ugly. *J Feline Med Surg*. 2015; 17:203-219. doi:10.1177/1098612X15571878
2. Schmidt-Kuntzel A, Eizirik E, O'Brien SJ, et al. Tyrosinase and tyrosinase related protein 1 alleles specify domestic cat coat color phenotypes of the Albino and Brown loci. *J Hered* 2005; 96: 289–301
3. Lyons LA, Imes DL, Rah HC, et al. Tyrosinase mutations associated with Siamese and Burmese patterns in the domestic cat (*Felis catus*). *Anim Genet* 2005; 36: 119–126.
4. Imes DL, Geary LA, Grahn RA, et al. Albinism in the domestic cat (*Felis catus*) is associated with a tyrosinase (TYR) mutation. *Anim Genet* 2006; 37: 175–178
5. Lyons LA, Foe IT, Rah HC, et al. Chocolate coated cats: TYRP1 mutations for brown color in domestic cats. *Mamm Genome* 2005; 16: 356–366
6. Peterschmitt M, Grain F, Arnaud B, et al. Mutation in the melanocortin 1 receptor is associated with amber colour in the Norwegian forest cat. *Anim Genet* 2009; 40: 547–552
7. Montague MJ, Li G, Gandolfi B, et al. Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. *Proc Natl Acad Sci USA* 2014; 111: 17230–17235
8. David VA, Menotti-Raymond M, Wallace AC, et al. Endogenous retrovirus insertion in the KIT oncogene determines white and white spotting in domestic cats. G3 (Bethesda) 2014; 4: 1881–1891
9. Frischknecht M, Jagannathan V, Leeb T. Whole genome sequencing confirms KIT insertions in a white cat. *Anim. Genet* 2015; 46: 98
10. Górska A, Drobik-Czwarno W, Bryś, J. Genetic determination of the amount of white spotting: a case study in Siberian cats. *Genes* (Basel) 2022; 13(6)

11. Anderson H, Salonen M, Toivola S, Blades M, Lyons L A, Forman OP, Hytönen MK, Lohi H. A new Finnish flavor of feline coat coloration, "salmiak," is associated with a 95-kb deletion downstream of the KIT gene. *Anim. Genet* 2024; 55: 676-680
12. Kaelin CB, McGowan KA, Trotman JC, Koroma DC, David VA, Menotti-Raymond M, Graff EC, Schmidt-Küntzel A, Oancea E, Barsh GS. Molecular and genetic characterization of sex-linked orange coat color in the domestic cat. *Curr Biol* 2025; 35:2826-2836.e5. Pubmed reference: 40378841. DOI: 10.1016/j.cub.2025.04.055
13. Toh H, Au Yeung WK, Unoki M, Matsumoto Y, Miki Y, Matsumura Y, Baba Y, Sado T, Nakamura Y, Matsuda M, Sasaki H. A deletion at the X-linked ARHGAP36 gene locus is associated with the orange coloration of tortoiseshell and calico cats. *Curr Biol* 2025; 35:2816-2825.e3, 2025. Pubmed reference: 40378840. DOI: 10.1016/j.cub.2025.03.075
14. Ishida Y, David VA, Eizirik E, et al. A homozygous single-base deletion in MLPH causes the dilute coat color phenotype in the domestic cat. *Genomics* 2006; 88: 698–705
15. Eizirik E, Yuhki N, Johnson WE, et al. Molecular genetics and evolution of melanism in the cat family. *Curr Biol* 2003; 13: 448–453
16. Kaelin CB, Xu X, Hong LZ, et al. Specifying and sustaining pigmentation patterns in domestic and wild cats. *Science* 2012; 337: 1536–1541
17. Lyons LA, Buckley RM, Harvey RJ, 99 Lives Cat Genome Consortium. Mining the 99 Lives Cat Genome Sequencing Consortium database implicates genes and variants for the Ticked locus in domestic cats (*Felis catus*). *Anim Genet* 2021; 52: 321–332 https://doi.org/10.1111/age.13059
18. Kehler JS, David VA, Schaffer AA, et al. Four independent mutations in the feline fibroblast growth factor 5 gene determine the long-haired phenotype in domestic cats. *J Hered* 2007; 98: 555–566
19. Drogemuller C, Rufenacht S, Wichert B, et al. Mutations within the FGF5 gene are associated with hair length in cats. *Anim Genet* 2007; 38: 218–221
20. Gandolfi B, Outerbridge C, Beresford L, et al. The naked truth: Sphynx and Devon Rex cat breed mutations in KRT71. *Mamm Genome* 2010; 21; 509–515
21. Gandolfi B, Alhaddad H, Joslin SE, et al. A splice variant in KRT71 is associated with curly coat phenotype of Selkirk Rex cats. *Sci Rep* 2013; 3: 2000
22. Gandolfi B, Alhaddad H, Affolter VK, et al. To the root of the curl: a signature of a recent selective sweep identifies a mutation that defines the Cornish Rex cat breed. *PloS One* 2013; 8: e67105
23. Abitbol M, Couronné A, Dufaure de Citres C, Gache V. A PAX3 insertion in the Celestial

452 breed and certain feline breeding lines with dominant blue eyes. *Anim Genet* 2024; 55: 670-
453 675
454 24. Abitbol M, Dufaure de Citres C, Rudd Garces G, Lühken G, Lyons LA, Gache V. Different
455 founding effects underlie dominant blue eyes (DBE) in the domestic cat. *Animals (Basel)*
456 2024; 14:1845. Pubmed reference: 38997957. DOI: 10.3390/ani14131845
457 25. Buckley RM, Davis BW, Brashear WA, Farias FHG, Kuroki K, Graves T, et al. A new
458 domestic cat genome assembly based on long sequence reads empowers feline genomic
459 medicine and identifies a novel gene for dwarfism. *PLoS Genet* 2020; 16(10): e1008926.
460 https://doi.org/10.1371/journal.pgen.1008926
461 26. Bredemeyer KR, Harris AJ, Li G, Zhao L, Foley NM, Roelke-Parker M, O'Brien SJ, Lyons
462 LA, Warren WC, Murphy WJ. Ultracontinuous single haplotype genome assemblies for the
463 domestic cat (*Felis catus*) and Asian leopard cat (*Prionailurus bengalensis*). *J Hered* 2021
464 112(2), 165–173. https://doi.org/10.1093/jhered/esaa057
465 27. Matsumoto Y, Chung CYL, Isobe S, Sakamoto M, Lin X, et al. Chromosome-scale
466 assembly with improved annotation provides insights into breed-wide genomic structure
467 and diversity in domestic cats. *J Adv Res* 2025; 75, 863-874
468 28. Lyons LA, Creighton EK, Alhaddad H. et al. Whole genome sequencing in cats, identifies
469 new models for blindness in AIPL1 and somite segmentation in HES7. *BMC Genomics*
470 2016; 17: 265. https://doi.org/10.1186/s12864-016-2595-4
471 29. Rodney AR, Buckley RM, Fulton RS et al. A domestic cat whole exome sequencing
472 resource for trait discovery. *Sci Rep* 2021; 11: 7159. https://doi.org/10.1038/s41598-021-
473 86200-7
474 30. Anderson H, Davison S, Lytle KM, Honkanen L, Freyer J, Mathlin J, et al. Genetic
475 epidemiology of blood type, disease and trait variants, and genome-wide genetic diversity in
476 over 11,000 domestic cats. *PLoS Genet* 2022; 18(6): e1009804.
477 https://doi.org/10.1371/journal.pgen.1009804
478 31. Hernandez I, Hayward JJ, Brockman JA, White ME et al. Complex feline disease mapping
479 using a dense genotyping array. *Front Vet Sci* 2022; 9, 862414
480 32. Raffle J, Novo Matos J, Wallace M, et al. Identification of novel genetic variants associated
481 with feline cardiomyopathy using targeted next-generation sequencing. *Sci Rep* 2025; 15:
482 3871
483 33. Crandell RA, Fabricant CG, Nelson-Rees WA. Development, characterization, and viral
484 susceptibility of a feline (*Felis catus*) renal cell line (CRFK). *In Vitro* 1973; 9: 176-185
485 34. Lawson JS, Syme HM, Wheeler-Jones CPD, Elliott J. Characterisation of Crandell-Rees

486 Feline Kidney (CRFK) cells as mesenchymal in phenotype. *Res Vet Sci* 2019; 127: 99-102
487 35. Sosnovtsev SV, Prikhod'ko EA, Belliot G, Cohen JI, Green KY. Feline calicivirus
488 replication induces apoptosis in cultured cells. *Virus Res* 2003; 94: 1-10
489 36. Bidawid S, Malik N, Adegbunrin O, Sattar SA, Farber JM. A feline kidney cell line-based
490 plaque assay for feline calicivirus, a surrogate for Norwalk virus. *J Virol Methods* 2003;
491 107: 163-167
492 37. Tian J, Zhang X, Wu H, Liu C, Liu J, Hu X, Qu L. Assessment of the IFN-β response to
493 four feline caliciviruses: Infection in CRFK cells. *Infect Genet Evol* 2015; 34: 352-360.
494 38. Siebelink KH, Karlas JA, Rimmelzwaan GF, Osterhaus AD, Bosch ML. A determinant of
495 feline immunodeficiency virus involved in Crandell feline kidney cell tropism. *Vet Immunol*
496 *Immunopathol* 1995; 46: 61-69
497 39. Verschoor EJ, Boven LA, Blaak H, Van Vliet AL, Horzinek MC, De Ronde A. A single
498 mutation within the V3 envelope neutralization domain of feline immunodeficiency virus
499 determines its tropism for CRFK cells. *J Virol* 1995; 69: 4752-4757
500 40. Baldinotti F, Matteucci D, Mazzetti P, Giannelli C, Bandecchi P, Tozzini F, Bendinelli M.
501 Serum neutralization of feline immunodeficiency virus is markedly dependent on passage
502 history of the virus and host system. *J Virol* 1994; 68: 4572-4579
503 41. Mizuno T, Goto Y, Baba K, Masuda K, Ohno K, Tsujimoto H. TNF-α-induced cell death in
504 feline immunodeficiency virus-infected cells is mediated by the caspase cascade. *Virology*
505 2001; 287: 446-455
506 42. Drechsler Y, Vasconcelos EJ, Griggs LM, Diniz PP. CRFK and primary macrophages
507 transcriptomes in response to feline coronavirus infection differ significantly. *Front Genet*
508 2020; 11: 584744
509 43. Drechsler Y, Vasconcelos EJ, Diniz PP. Host transcriptome studies in response to feline
510 coronavirus reveal differences in macrophages vs CRFK. *J Immunol*, 2020;
511 204(1_Supplement): 92-27
512 44. Drechsler Y, Vasconcelos EJ, Griggs LM, Diniz PP. Host responses to feline coronavirus are
513 significantly different in primary macrophages compared to CRFK cells. *J Immunol* 2021;
514 206(1_Supplement): 19-13
515 45. Camero M, Lanave G, Catella C, Lucente MS, et al. ERDRP-0519 inhibits feline
516 coronavirus in vitro. *BMC Vet Res* 2022; 18: 55
517 46. Baumann JG, Günzburg WH, Salmons B. CrFK feline kidney cells produce an RD114-like
518 endogenous virus that can package murine leukemia virus-based vectors. *J Virol* 1998; 72:
519 https://doi.org/10.1128/jvi.72.9.7685-7687.1998

47. Shimode S, Nakagawa S, Miyazawa T. Multiple invasions of an infectious retrovirus in cat genomes. *Sci Rep* 2015; 5, 8164
48. Shimode S, Sakuma T, Yamamoto T, Miyazawa T. Establishment of CRFK cells for vaccine production by inactivating endogenous retrovirus with TALEN technology. *Sci Rep* 2022; 12: 6641
49. Lappin MR, Jensen WA, Jensen TD, Basaraba RJ, Brown CA, Radecki SV, Hawley JR. Investigation of the induction of antibodies against Crandell-Rees feline kidney cell lysates and feline renal cell lysates after parenteral administration of vaccines against feline viral rhinotracheitis, calicivirus, and panleukopenia in cats. *Am J Vet Res* 2005; 66: 506-511.
50. Whittemore JC, Hawley JR, Jensen WA, Lappin MR. Antibodies against Crandell Rees Feline Kidney (CRFK) cell line antigens, α-enolase, and annexin A2 in vaccinated and CRFK hyperinoculated cats. *J Vet Intern Med* 2010; 24: 306-313
51. Yoshikawa R, Sato E, Igarashi T, Miyazawa T. Characterization of RD-114 virus isolated from a commercial canine vaccine manufactured using CRFK cells. *J Clin Microbiol* 2010; 48: 3366-3369
52. Haapala DK, Robey WG, Oroszlan SD, Tsai WP, Isolation from cats of an endogenous type C virus with a novel envelope glycoprotein. *J Virol* 1985; 53: 827-33
53. Bassin RH, Ruscetti S, Ali I,Haapala DK, Rein A, Normal DBA/2 mouse cells synthesize a glycoprotein which interferes with MCF virus infection. *Virol* 1982; 123: 139-51
54. Lasfargues EY, Lasfargues JC, Dion AS, Greene AE, Moore DH. Experimental infection of a cat kidney cell line with the mouse mammary tumor virus. Cancer Res. 1976; 36: 67-72
55. RD114 retrovirus, complete genome: https://www.ncbi.nlm.nih.gov/nuccore/NC_009889.1/. Accessed 4 August 2025
56. Cunningham F, Allen JE, Allen J, Amode MR, Armean IM, Azov AG, Barnes I, Bennett R, Berry A, Bhai J, Bignell A, Billis K, Boddu S, et al. Ensembl 2022. *Nucl Acids Res* 2022; 50(D1), D988–D995
57. Ensembl Complete datasets and databases: https://ftp.ensembl.org/pub/data_files/felis_catus/F.catus_Fca126_mat1.0/rnaseq/. Accessed 26 January 2023
58. Liao Y, Smyth GK, Shi W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014; 30: 923–930
59. F. catlas: https://sites.google.com/view/fcatlas/. Accessed 1 November 2025
60. Awazu A, Takemoto D, Watanabe K, Sakamoto N. Possibilities of skin coat color-dependent risks and risk factors of squamous cell carcinoma and deafness of domestic cats inferred via

554 RNA-seq data. *Genes Cells* 2023; 28: 893-905
555 61. Chiu ES, VandeWoude S. Presence of endogenous viral elements negatively correlates with
556 feline leukemia virus susceptibility in puma and domestic cat Cells. *J Virol* 2020; 94:
557 10.1128/jvi.01274-20
558 62. Erbeck K, Gagne RB, Kraberger S,Chiu ES, Roelke-Parker M, VandeWoude S. Feline
559 leukemia virus (FeLV) endogenous and exogenous recombination events result in multiple
560 FeLV-B subtypes during natural infection. *J Virol* 2021; 95: 10.1128/jvi.00353-21
561 63. Pramono D, Takeuchi D, Katsuki M, AbuEed L, Abdillah D, Kimura T, Kawasaki J, Miyake
562 A, Nishigaki K. FeLIX is a restriction factor for mammalian retrovirus infection. *J Virol*
563 2024; 98:e01771-23
564 64. Kennedy LJ, Barnes A, Happ GM, Quinnell RJ, Bennett D, Angles JM, Day MJ,
565 Carmichael N, Innes JF, Isherwood D, Carter SD, Thomson W, Ollier WE. Extensive
566 interbreed, but minimal intrabreed, variation of DLA class II alleles and haplotypes in dogs.
567 Tissue Antigens. 2002; 59: 194-204
568 65. Angles JM, Kennedy LJ, Pedersen NC. Frequency and distribution of alleles of canine
569 MHC-II DLA-DQB1, DLA-DQA1 and DLA-DRB1 in 25 representative American Kennel
570 Club breeds. Tissue Antigens. 2005; 66:173-84
571 66. Miyamae J, Okano M, Katakura F, Kulski JK, Moritomo T, Shiina T. Large-Scale
572 Polymorphism Analysis of Dog Leukocyte Antigen Class I and Class II Genes (*DLA-*
573 *88*, *DLA-12/88L* and *DLA-DRB1*) and Comparison of the Haplotype Diversity between
574 Breeds in Japan. Cells. 2023;12: 809
575 67. Graumann MB, DeRose SA, Ostrander EA, Storb R. Polymorphism analysis of four canine
576 MHC class I genes. Tissue Antigens. 1998; 51(4 Pt 1): 374-81
577 68. Wagner JL, Burnett RC, Storb R. Organization of the canine major histocompatibility
578 complex: current perspectives. J Hered. 1999; 90 :35-8
579 69. Morris KM, Kirby K, Beatty JA, Barrs VR, Cattley S, David V, O'Brien SJ, Menotti-
580 Raymond M, Belov K. Development of MHC-Linked Microsatellite Markers in the
581 Domestic Cat and Their Use to Evaluate MHC Diversity in Domestic Cats, Cheetahs, and
582 Gir Lions. J Hered. 2014; 105: 493-505
583 70. Okano M, Miyamae J, Suzuki S, Nishiya K, Katakura F, Kulski JK, Moritomo T, Shiina T.
584 Identification of Novel Alleles and Structural Haplotypes of Major Histocompatibility
585 Complex Class I and DRB Genes in Domestic Cat (*Felis catus*) by a Newly Developed
586 NGS-Based Genotyping Method. Front Genet. 2020; 11: 750
587 71. Holmes JC, Holmer SG, Ross P, Buntzman AS, Frelinger JA, Hess PR. Polymorphisms and

588 tissue expression of the feline leukocyte antigen class I loci FLAI-E, FLAI-H, and FLAI-K.
589 Immunogenetics. 2013; 65: 675-89
590 72. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE,
591 Lunter G, Marth GT, Sherry ST, McVean G, Durbin R; 1000 Genomes Project Analysis
592 Group. The variant call format and VCFtools. Bioinformatics. 2011; 27: 2156-2158
593 73. Grahn RA, Kurushima JD, Billings NC, Grahn JC, Halverson JL, Hammer E, Ho CK, Kun
594 TJ, Levy JK, Lipinski MJ, Mwenda JM, Ozpinar H, Schuster RK, Shoorijeh SJ, Tarditi CR,
595 Waly NE, Wictum EJ, Lyons LA. Feline non-repetitive mitochondrial DNA control region
596 database for forensic evidence. Forensic Sci Int Genet. 2011;5: 33-42
597 74. Wesselink M, Bergwerff L, Hoogmoed D, Kloosterman AD, Kuiper I. Forensic utility of the
598 feline mitochondrial control region - A Dutch perspective. Forensic Sci Int Genet. 2015; 17:
599 25-32
600 75. Bilgen N, Güvendi M, Sezgin E, Can H, Köseoğlu AE, Erkunt Alak S, Döşkaya M, Ün C.
601 Mitotyping of random bred cats and pure breed cats (Turkish Angora and Turkish Van)
602 using non-repetitive mitochondrial DNA control region. Gene. 2024; 892: 147849
603 76. Jaito W, Singchat W, Patta C, Thatukan C, Kumnan N, Chalermwong P, Budi T, Panthum T,
604 Wongloet W, Wattanadilokchatkun P, Thong T, Muangmai N, Han K, Duengkae P,
605 Phatcharakullawarawat R, Srikulnath K. Shared alleles and genetic structures in different
606 Thai domestic cat breeds: the possible influence of common racial origins. Genomics
607 Inform. 2024; 22: 12

608

609 **Table 1: Phenotypic traits determined by DNA variants, and genotypes of CRFK and PG-4**
610 **cells.**

Locus Gene	Phenotype (Alleles)	Mutation in CDS (or related noncoding region): F.catus_Fca126_mat1.0 Genome: F.catus_Fca126_mat1.0 Genome: AnAms1.0	Depth of alleles, and Genotype in CRFK	Depth of alleles, and Genotype in PG-4
Color	Burmese	c.679G>T	+ : 18	+ : 16
TYR[2–4]	(c^b)	D1: g.44022819C>A	c^b : 0	c^b : 0
		D1: g.46081417C>A	+ / +	+ / +
	Siamese	c.904G>A	+ : 25	+ : 17
	(c^s)	D1: g.44013031C>T	c^s : 0	c^s : 20
		D1: g.46071661C>T	+ / +	+ / c^s
	Albino	c.939del	+ : 25	+ : 34
	(c^a)	D1: g.44012996del	c^a : 0	c^a : 0
		D1: g.46071626del	+ / +	+ / +
	Mocha	c.820_936delinsAATCTC	+ : 40	+ : 54
	(c^m)	D1: g.44012836_44012998delinsGAGATT	c^m : 0	c^m : 0
		D1: g.46071466_46071628delinsGAGATT	+ / +	+ / +
Brown	Chocolate	c.1261+5G>A	+ : 22	+ : 26
TYRP1[2,5]	(b)	D4: g.38142265G>A	b : 0	b : 0
		D4: g.39391308G>A	+ / +	+ / +
	Cinnamon	c.298C>T	+ : 20	+ : 19
	(b^l)	D4: g.38130163C>T	b^l : 0	b^l : 0
		D4: g.39379209C>T	+ / +	+ / +
Amber	Red	c.250G>A	+ : 35	+ : 34
MC1R[6]	(e)	E2: g.61570294G>A	e : 0	e : 0
		E2: g.62623146G>A	+ / +	+ / +

White	White	(Intron 1: FERV1-LTR insertion)	Fig. 3	Fig. 3
KIT[7–11]	(w^w)	B1: g.161388687_161388688insFERV1-LTR	+ / +	+ / +
		B1: g.163130991_163130992insFERV1-LTR		
	White	(Intron 1: FERV1 insertion)	Fig. 3	Fig. 3
	spotted	B1: g.161388687_161388688insFERV1	+ / +	+ / w^s
	(w^s)	B1: g.163130991_163130992insFERV1		
	White	c.1035_1036delinsCA	+ : 27	+ : 28
	glove	B1: g.161337385_161337386delinsTG	w^g : 0	w^g : 0
	(w^g)	B1: g.163079561_163079562delinsTG	+ / +	+ / +
	Salmiak	(KIT-KDR Intergenic region: 95 kb deletion)	Fig. 2a	Fig. 2a
	(w^a)	B1: g.161142880_161237957del	+ / +	+ / +
		B1: g.162885414_162980052del		
Orange	Orange	(Intron 1: 5 kb deletion)	Fig. 2b	Fig. 2b
ARHGAP36	(o)	(Lacking)	+ / +	+ / +
[12,13]		X: g.109186183_109191258del		
Dilution	Dilute	c.83del	+ : 15	+ : 38
MLPH[14]	(d)	C1: g.218197448del	d : 4	d : 0
		C1: g.219535339del	+ / d	+ / +
Agouti	Nonagouti	c.123_124del	+: 0	+: 0
ASIP[15]	(a)	A3: g.24831083-24831084del	a: 48	a: 25
		A3: g.25283750_25283751del	a / a	a / a
Tabby	Blotched	c.176C>A	+ : 16	+ : 17
LVRN[16]	tabby	A1: g.94477275C>A	t^b : 0	t^b : 0
	(t^b)	A1: g.97133509C>A	+ / +	+ / +
	Blotched	c.416C>A	+ : 34	+ : 10
	tabby	A1: g.94477515C>A	t^b : 0	t^b : 12
	(t^b)	A1: g.97133749C>A	+ / +	+ / t^b
	Blotched	c.682G>A	+ : 33	+ : 22
	tabby	A1: g.94477781G>A	t^b : 0	t^b : 0
	(t^b)	A1: g.97134015G>A	+ / +	+ / +

	Blotched	c.2522G>A	+ : 0	+ : 5
	tabby	A1: g.94536796G>A	t^b : 10	t^b : 3
	(t^b)	A1: g.97192983G>A	t^b / t^b	+ / t^b
Ticked	Ticked	c.188G>A	+ : 36	+ : 20
DKK4[17]	(ti)	B1: g.40429492G>A	ti : 0	ti : 0
		B1: g.41805975G>A	+ / +	+ / +
	Ticked	c.53C>T	+ : 33	+ : 30
	(ti)	B1: g.40428846C>T	ti : 0	ti : 0
		B1: g.41805329C>T	+ / +	+ / +
Longhair	Long fur	c.356_357insT	+ : 24	+ : 22
FGF5[18,19]	(l)	B1: g.139634675_139634676insA	l : 0	l : 0
		B1: g.141368532_141368533insA	+ / +	+ / +
	Long fur	c.406C > T	+ : 24	+ : 25
	(l)	B1: g.139645946G>A	l : 0	l : 0
		B1: g.141379803G>A	+ / +	+ / +
	Long fur	c.474del	+ : 26	+ : 0
	(l)	B1: g.139634261del	l : 0	l : 7
		B1: g.141368118del	+ / +	+ / l
	Long fur	c.475A > C	+ : 0	+ : 0
	(l)	B1: g.139634260T>G	l : 26	l : 13
		B1: g.141368117T>G	l / l	l / +
	Long fur	c.577G>A	+ : 37	+ : 27
	(l)	B1: g.139634158C>T	l : 0	l : 0
		B1: g.141368015C>T	+ / +	+ / +
Hairless	Hairless	c.816+1G>A	+ : 40	+ : 25
Rexing	(hr)	B4: g.78941699C>T	hr : 0	hr : 0
KRT71 [20,21]		B4: g.80471148C>T	+ / +	+ / +
	Rexing	c.1108-4_1184del,	+ : 51	+ : 32
	(re)	c.1184_1185insAGTTGGAG,	re : 0	re : 0
		and c.1196_1197insT	+ / +	+ / +
		B4: g.78939390_78939470del,		

		B4: g.78939389_78939390insCTCCAACT,		
		and B4: g.78939377_78939378insA		
		B4: g.80468919_80468919del,		
		B4: g.80468918_80468919insCTCCAACT,		
		and B4: g.80468826_80468827insA		
	Rexing	c.445-1G>C	+ : 37	+ : 29
	(re)	B4: g.78943283C>G	re : 0	re : 0
		B4: g.80472733C>G	+ / +	+ / +
Rexing	Rexing	c.250_253del	+ : 16	+ : 19
LPAR6[22]	(re)	A1: g.22865220_22865223del	re : 0	re : 0
		A1: g.23372953_23372956del	+ / +	+ / +
Blue iris	Blue iris	(Inrton4: FERV1-LTR homolog insertion)	+ :	+ :
PAX3[23,24]	(bi)	C1: g.205833101_205833102insN[395]	bi : 0	bi : 0
		C1: g.207180017_207180018insN[395]	+ / +	+ / +
	Blue iris	(Intron4: RD114-LTR homolog insertion)	+ :	+ :
	(bi)	C1: g.205834854_205834855insN[433]	bi : 0	bi : 0
		C1: g.207181773_207181774insN[433]	+ / +	+ / +

List of genes and phenotypes we focused on in this study, mutation position on CDS and genomic coordinates provided by Fca126_mat1.0 and AnAms1.0, coverage depth of alleles, and a summary of genotypes of CRFK and PG-4 cells as determined in the present study.

616 **Figure captions**

617 **Figure 1: Alignment results of whole genome sequencing reads of CRFK and PG-4 cells**
618 **with WT and mutant sequences of FGF5 genes.** Alignment results of genome sequencing
619 reads around FGF5: c.474 and c.475; CRFK (a) and PG-4 (b) cells. Left string indicates the
620 name of sequenced read, and red colored symbols indicate coat phenotype-related mutations

621 **Figure 2: Read coverage depth distributions of CRFK and PG-4 cells around the salmiak**
622 **and Orange loci.** Read coverage depth distributions from 95 kbp upper stream to 95 kbp lower
623 stream of the salmiak locus (a), and those from 5.1 kbp upper stream to 5.1 kbp lower stream of
624 the Orange locus (b) of CRFK and PG-4 cells. Bar graph colored orange indicates coverage
625 depth of reads mapped to the salmiak (a) and Orange loci (b), respectively. In all distributions,
626 no notable declines in read coverage depth were observed in both salmiak and Orange loci,
627 indicating that both cats from which the CRFK and PG-4 cells express no salmiak and orange
628 coat.

629 **Figure 3: Parts of upper and lower streams of alignment results of long-read sequencing**
630 **reads of CRFK and PG-4 cells with WT and KIT-FERV1 sequences of KIT intron 1.**
631 Alignment results of long-read genome sequencing reads from CRFK and PG-4 cells to the
632 KIT-FERV1 that contains sequence around the region B1: g.161388687_161388688
633 (F.catus_Fca126_mat1.0). Left string indicates [the name of cell (CRFK or PG-4): name of
634 sequenced read], and red colored sequences indicate FERV1 and aligned sequences with it. Four
635 long reads from CRFK could align, but no reads contained the FERV1 homologue, while 6
636 reads from PG-4 could align, and four of the six reads contained the FERV1 homologue. This
637 figure shows only a part of the results at the upper and lower streams of the KIT-FERV1
638 sequence. See Table S1 for the alignment result with the full-length KIT-FERV1 sequence

639 **Figure 4: Example illustrations of inferred original cats of CRFK and PG-4 cells.** An
640 example illustration of the inferred phenotypes of coat morphologies and iris colors of the
641 original cats of the CRFK cell (a) and the PG-4 cell (b). CRFK cells are expected to originate
642 from a cat with long, black fur, lacking stripes, and with non-blue irises. PG-4 cells were
643 expected to be derived from a cat with long, white and black bicolored fur without stripes and
644 non-blue irises

645 **Figure 5: Expression levels of coat phenotype-related and iris color-related genes in**
646 **various tissues.** The expression level of each gene in each tissue is quantified as $\log_2$(transcripts

per million+1). The length of the red bar for gene A for each tissue is proportional to the [expression levels of gene A in each tissue] / [the maximum expression level of gene A among all focused tissues]. Most coat phenotype-related genes were found to exhibit similar or higher expressions in various tissues other than skin, and the iris color-related gene (PAX3) was also found to exhibit similar or higher expressions in a variety of tissues other than the opt-functional organs

(a) CRFK

FGF5: c. 474, c. 475

```
                                                  ↓
WT                                                tcccccccaccaggccaaatt ta ccgatgactgcaagttcagggagcgattccaagaaaac
FGF5:c.474del                                     tcccccccaccaggccaaatt -a ccgatgactgcaagttcagggagcgattccaagaaaac
FGF5:c.475A>C                                     tcccccccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaac
Read name and sequence
_R_LH00504:186:22YKYJLT3:1:1136:23098:8018/2      -------------------- tc ccgatgactgcaagttcagggagcgattccaagaaaac
_R_LH00504:186:22YKYJLT3:1:1192:25992:16846/1     ----------------aatt tc ccgatgactgcaagttcagggagcgattccaagaaaac
_R_LH00504:186:22YKYJLT3:1:1187:1481:21444/1      ---------caggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaac
_R_LH00504:186:22YKYJLT3:1:1220:37134:16734/1     ----cccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaac
LH00504:186:22YKYJLT3:1:2219:7629:10021/2         --ccccccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaac
_R_LH00504:186:22YKYJLT3:1:2212:14989:24120/1     tcccccccaccaggccaaatt tc ccgatgactacaagttcagggagcgattccaagaaaac
LH00504:186:22YKYJLT3:1:2139:51206:15580/1        tcccccccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaac
LH00504:186:22YKYJLT3:1:1247:37291:5630/1         tcccccccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaac
_R_LH00504:186:22YKYJLT3:1:2246:43264:25113/2     tcccccccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaac
_R_LH00504:186:22YKYJLT3:1:2246:43255:25129/2     tcccccccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaac
LH00504:186:22YKYJLT3:1:2185:31669:12744/2        tcccccccaccaggccaaatt t- ccgatgactgcaagttcagggagcgattccaagaaaac
LH00504:186:22YKYJLT3:1:1132:13233:8867/2         tcccccccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaac
LH00504:186:22YKYJLT3:1:1106:21286:25834/1        tcccccccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaac
LH00504:186:22YKYJLT3:1:1172:1749:11559/2         tcccccccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaac
LH00504:186:22YKYJLT3:1:1172:1740:11575/2         tcccccccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaac
LH00504:186:22YKYJLT3:1:1181:35396:19906/1        tcccccccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaac
LH00504:186:22YKYJLT3:1:2274:13473:12616/2        tcccccccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaac
LH00504:186:22YKYJLT3:1:2274:18586:13209/2        tcccccccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaac
LH00504:186:22YKYJLT3:1:1267:31882:27341/2        tcccccccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaac
_R_LH00504:186:22YKYJLT3:1:1204:11014:26171/1     tcccccccaccgggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaac
LH00504:186:22YKYJLT3:1:1192:25992:16846/2        tcccccccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaac
LH00504:186:22YKYJLT3:1:1161:30662:6960/2         tcccccccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaac
LH00504:186:22YKYJLT3:1:1161:30671:6976/2         tcccccccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaac
LH00504:186:22YKYJLT3:1:2212:14989:24120/2        tcccccccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccacgaaaac
LH00504:186:22YKYJLT3:1:2157:26510:25049/2        tcccccccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaaa
LH00504:186:22YKYJLT3:1:2144:20306:10389/2        tcccccccacccggccaaatt tc ccgatgac------------------------------
```

(b) PG-4

FGF5: c. 474, c. 475

```
                                                  ↓
WT                                                cccccaccaggccaaatt ta ccgatgactgcaagttcagggagcgattccaagaaaaca-
FGF5:c.474del                                     cccccaccaggccaaatt -a ccgatgactgcaagttcagggagcgattccaagaaaaca-
FGF5:c.475A>C                                     cccccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaaca-
Read name and sequence
_R_LH00504:188:22Y5KLLT3:4:2128:16034:3147/2      -------------aaatt -a ccgatgactgcaagttcagggagcgattccaagaaaacag
_R_LH00504:186:22YKYJLT3:1:1209:31290:3628/1      --------aggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaacag
LH00504:186:22YKYJLT3:1:2102:7361:11959/1         -------caggccaaatt -a ccgatgactgcaagttcagggagcgattccaagaaaacag
_R_LH00504:186:22YKYJLT3:1:1153:39436:4957/2      --cccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaacag
LH00504:186:22YKYJLT3:1:2154:16238:26411/2        cccccaccaggccaaatt -a ccgatgactgcaagttcagggagcgattccaagaaaacag
LH00504:186:22YKYJLT3:1:2217:22645:11783/2        cccccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaacag
_R_LH00504:186:22YKYJLT3:1:2203:41359:11014/2     cccccaccaggccaaact -a ccgatgactgcaagttcagggagcgattccaagaaaacag
LH00504:188:22Y5KLLT3:4:1147:27379:6303/2         cccccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaacag
LH00504:186:22YKYJLT3:1:2249:9710:19490/2         cccccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaacag
LH00504:186:22YKYJLT3:1:1153:39436:4957/1         cccccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaacag
LH00504:186:22YKYJLT3:1:2149:37162:21492/2        cccccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaacag
LH00504:186:22YKYJLT3:1:1270:2452:27453/1         cccccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaacag
LH00504:186:22YKYJLT3:1:2276:50356:6095/2         cccccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaacag
LH00504:186:22YKYJLT3:1:2154:27490:12200/2        cccccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaacag
LH00504:186:22YKYJLT3:1:2136:34998:5022/1         cccccaccaggccaaatt -a ccgatgactgcaagttcagggagcgattccaagaaaacag
_R_LH00504:186:22YKYJLT3:1:1233:2174:8963/2       cccccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaacag
_R_LH00504:186:22YKYJLT3:1:1233:2165:8979/2       cccccaccaggccaaatt tc ccgatgactgcaagttcagggagcgattccaagaaaacag
LH00504:186:22YKYJLT3:1:2287:47203:29279/2        cccccaccaggccaaatt -a ccgatgactgcaagttcagggagcgattccaagaaaaca-
_R_LH00504:186:22YKYJLT3:1:1202:22719:14218/1     cccccaccaggccaaatt tc ccgatgactgcaagttc-----------------------
LH00504:186:22YKYJLT3:1:1222:6446:7233/1          cccccaccaggccaaatt -a ccgatgactgc-----------------------------
```

Figure 1

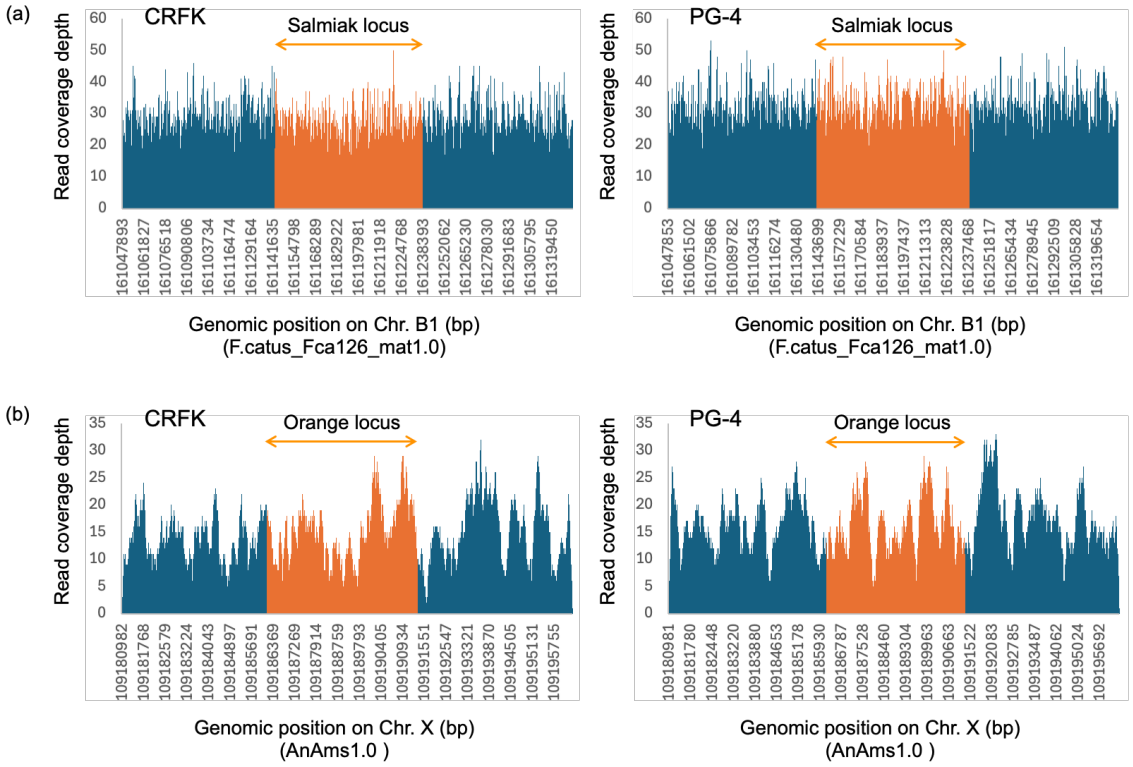
(a)
CRFK
Salmiak locus
Read coverage depth
Genomic position on Chr. B1 (bp)
(F.catus_Fca126_mat1.0)
PG-4
Salmiak locus
Read coverage depth
Genomic position on Chr. B1 (bp)
(F.catus_Fca126_mat1.0)
(b)
CRFK
Orange locus
Read coverage depth
Genomic position on Chr. X (bp)
(AnAms1.0)
PG-4
Orange locus
Read coverage depth
Genomic position on Chr. X (bp)
(AnAms1.0)

Figure 2

```
                                                                B1: g.161388687
                                                                        ↓
WT                                          ggggcgggggt-gggagatggagtggaacttt----------------------------
White spotted                               ggggcgggggt-gggagatggagtggaactttgttatgccc--aaaattcgtgatcccc
Cell: Read name and sequence
CRFK:b7c5b596-3c70-45bb-b00c-988c08291521   ggggcgggggt-gggagatggagtggaacttt----------------------------
CRFK:4f28cb2e-0b77-4276-927b-22cf1cd859a9   ggggcgggggt-gggagatggagtggaacttt----------------------------
CRFK:1a121a2d-cbd1-4954-953a-c59b730e8896   ggggcgggggt-gggagatggagtggaacttt----------------------------
CRFK:5cc67035-40b4-4180-8075-baa2bb4be3e7   ggggcgggggtggggagatggagtggaacttt----------------------------
PG-4:1d375068-d36d-463b-a23b-b79a65e7211e   ggggcgggggt-gggagatggagtggggtttttact-----------------------
PG-4:22777b42-f4c6-4e0d-994b-27b1798695bb   ggggcgggggt-gggagatggagtggaactttgttatgccc--aaaattcgtgatcccc
PG-4:befbbc54-737b-4de1-802e-a01ed3fea542   ggggcgggggt-gggagatggagtggaactttgttatgccc--aaaattcgtgatcccc
PG-4:9baf3e20-d0e4-4767-94ac-af0d2ef545da   ggggcgggggt-gggagatggagtggaacttt----------------------------
PG-4:8bc2a176-7ea8-4825-a2f8-488be22f49bf   ggggcgggt---gggagatggagtggaactttgttatgccccaaaaaatcgtgatcccc
PG-4:7ed64960-07a4-4574-8739-89a5ecb31ad5   ggggcgggggt-gggagatggagtggaactttgttatgccc--aaaattcgtgatcccc

WT                                          ------------------------------------------------------------
White spotted                               aaagaccaccagggagccgagtccgatgcaaaagcaaagagcctttattcgagctagctc
Cell: Read name and sequence
CRFK:b7c5b596-3c70-45bb-b00c-988c08291521   ------------------------------------------------------------
CRFK:4f28cb2e-0b77-4276-927b-22cf1cd859a9   ------------------------------------------------------------
CRFK:1a121a2d-cbd1-4954-953a-c59b730e8896   ------------------------------------------------------------
CRFK:5cc67035-40b4-4180-8075-baa2bb4be3e7   ------------------------------------------------------------
PG-4:1d375068-d36d-463b-a23b-b79a65e7211e   ------------------------------------------------------------
PG-4:22777b42-f4c6-4e0d-994b-27b1798695bb   aaagaccaccagggagccgagtccgatgcaaaagcaaagagcctttattcgagctagctc
PG-4:befbbc54-737b-4de1-802e-a01ed3fea542   aaagaccaccagggagccgagtccgatgcaaaagcaaagagcctttattcgagctagctc
PG-4:9baf3e20-d0e4-4767-94ac-af0d2ef545da   ------------------------------------------------------------
PG-4:8bc2a176-7ea8-4825-a2f8-488be22f49bf   aaagaccaccagggagccgagtccgatgcaaaagcaaagagcctttattc-agctagctc
PG-4:7ed64960-07a4-4574-8739-89a5ecb31ad5   aaagaccaccagggagccgagtccgatgcaaaagcaaagagcctttattcgagctagctc

                                            ---------------------- 7020 bp ----------------------
                                                                                B1: g.161388688
                                                                                        ↓
WT                                          ------------------------------------------------------------ctg
White spotted                               aggtgaaggtcacagaacaagatggcgacggctggcgtaggcccgccctttcactttctg
Cell: Read name and sequence
CRFK:b7c5b596-3c70-45bb-b00c-988c08291521   ------------------------------------------------------------ctg
CRFK:4f28cb2e-0b77-4276-927b-22cf1cd859a9   ------------------------------------------------------------ctg
CRFK:1a121a2d-cbd1-4954-953a-c59b730e8896   ------------------------------------------------------------ctg
CRFK:5cc67035-40b4-4180-8075-baa2bb4be3e7   ------------------------------------------------------------ctg
PG-4:1d375068-d36d-463b-a23b-b79a65e7211e   ------------------------------------------------------------
PG-4:22777b42-f4c6-4e0d-994b-27b1798695bb   aggtgaaggtcacagaacaagatggcgacggctggcgtaggcccgccctttcactttctg
PG-4:befbbc54-737b-4de1-802e-a01ed3fea542   aggtgaaggtcacagaacaagatggcgacggctggcgtaggcccgccctttcactttctg
PG-4:9baf3e20-d0e4-4767-94ac-af0d2ef545da   ------------------------------------------------------------ctg
PG-4:8bc2a176-7ea8-4825-a2f8-488be22f49bf   aggtgaaggtcacagaacaagatggcgacggctggcgtaggcctgccctttcac----tt
PG-4:7ed64960-07a4-4574-8739-89a5ecb31ad5   aggtgaaggtcacagaacaagatggcgacggctggcgtaggcccgccctttcactttctg

WT                                          caaaatctta---cttggatcctaaagctgtagtgaaaatcc-ggttttatctgcgcggg
White spotted                               caaaatctta---cttggatcctaaagctgtagtgaaaatcc-ggttttatctgcgcggg
Cell: Read name and sequence
CRFK:b7c5b596-3c70-45bb-b00c-988c08291521   caaaatctta---cttggatcctaaagctgtagtgaaaatcc-ggttttatctgcgcggg
CRFK:4f28cb2e-0b77-4276-927b-22cf1cd859a9   caaaatctta---cttggatcctaaagctgtagtgaaaatcc-ggttttatctgcgcggg
CRFK:1a121a2d-cbd1-4954-953a-c59b730e8896   caaaatctta---cttggatcctaaagctgtagtgaaaatcc-ggttttatctgcgcggg
CRFK:5cc67035-40b4-4180-8075-baa2bb4be3e7   caaaatctta---cttggatcctaaagctgtagtgaaaatcc-ggttttatctgcgcggg
PG-4:1d375068-d36d-463b-a23b-b79a65e7211e   caaaatctta---cttggatcctaaagctgtagtgaaaatcc-ggttttatctgcgcggg
PG-4:22777b42-f4c6-4e0d-994b-27b1798695bb   caaaatctta---cttggatcctaaagctgtagtgaaaatcc-ggttttatctgcgcggg
PG-4:befbbc54-737b-4de1-802e-a01ed3fea542   caaaatctta---cttggatcctaaagctgtagtgaaaatcc-ggttttatctgcgcggg
PG-4:9baf3e20-d0e4-4767-94ac-af0d2ef545da   caaaatctta---cttggatcctaaagctgtagtgaaaatcc-ggttttatctgcgcggg
PG-4:8bc2a176-7ea8-4825-a2f8-488be22f49bf   tgaaatcttaattcttggatctgaaagctgtagtgaaaatccaagttttatctgcgcggg
PG-4:7ed64960-07a4-4574-8739-89a5ecb31ad5   caaaatctta---cttggatcctaaagctgtagtgaaaatcc-agttttatctgcgcggg
```

Figure 3

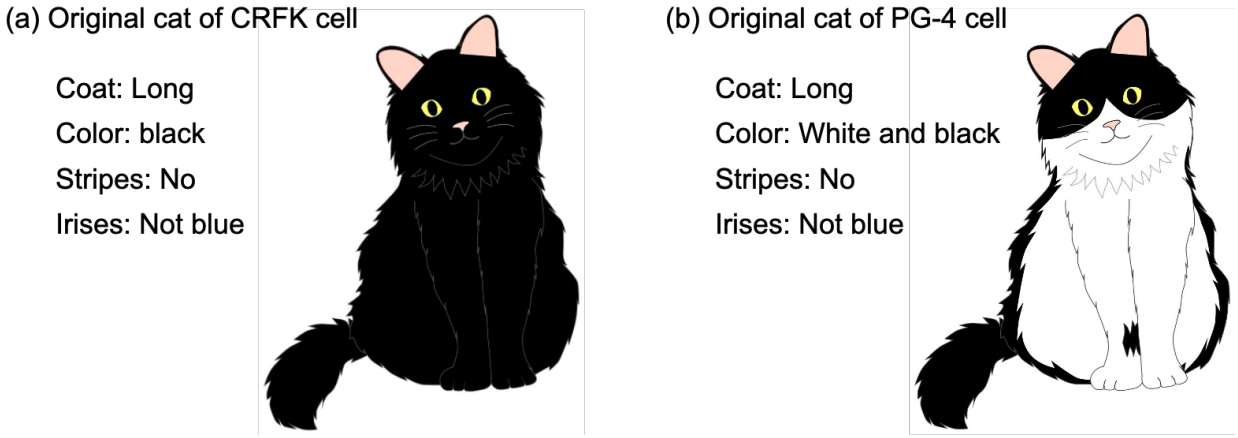
(a) Original cat of CRFK cell
Coat: Long
Color: black
Stripes: No
Irises: Not blue
(b) Original cat of PG-4 cell
Coat: Long
Color: White and black
Stripes: No
Irises: Not blue

Figure 4

	TYR	TYRP1	MC1R	KIT	ARHGAP36	MLPH	ASIP	LVRN	Dkk4	FGF5	KRT71	LPAR6	PAX3
skin	0.97949123	4.979143826	0.907122448	3.586949825	0.595540759	1.890367316	0.617256057	0.790593088	2.788113859	1.322693065	10.420824	4.264054274	1.420870477
skin (orange color)	1.287074139	4.681136779	1.514178733	4.728150746	0.92351867	1.992335312	0.560564858	0.730570842	2.385149963	2.506351052	11.57185462	4.780163285	1.519920078
skin (white color)	0.106508109	0	0.153761119	4.416760098	0.281844346	1.38368236	0.375739542	0.623645798	1.860833766	1.628767281	10.27795109	4.513211603	0.498482548
fibroblast adult	0.201311442	0.021819551	0.730664277	0.704854625	0.003685788	0.186873412	0.100232329	0.124979228	1.500135676	0	0.019069387	1.488602888	0.019355019
muscle	0.056483647	0	0.092118876	1.414603576	0.067172023	0.052994709	0.077034761	0.278525184	2.672768585	0.040640712	0	2.118117862	0.303226425
tooth	0.266573907	2.524633257	0.217491846	4.150102427	0.297112095	1.773640024	0.567123047	0.264163973	2.596387987	0.317583819	1.358416774	6.087545257	0.549355469
brain	0.074138775	0.244408995	6.154690407	6.125305599	0.488360957	0.106920683	0.445307287	0.482900784	2.997739706	1.214526236	0.082337234	2.495234234	0.920727113
hippocampus brain	0.104049158	0.986809378	7.141695431	6.633483047	1.263545739	0.381895847	0.462627145	0.267753817	3.196347533	2.064892	0.362603007	3.018231011	0
occipital brain	0.071192089	0.039376982	7.448888471	3.963105509	0.312113692	0.193440716	0.170021122	0.101684789	4.119749695	1.540684343	0.011456374	1.605354354	0
parietal lobe brain	0	0	7.186145561	4.875402302	0.660400161	0.64629313	0.441065494	0.042402384	3.887808995	1.329709621	0.132720662	2.664134262	0
temperal lobe brain	0	0	7.599838074	5.105537251	0.42981211	0.928785547	0.501174959	0.149430158	4.035299228	0.314304896	0.031937969	2.246188765	0
cerebellum brain	0.022500547	0.331065581	5.309532593	7.45398614	0.251584181	0.41274569	0.149029272	0.025240316	4.380271309	2.432348486	0.042684587	1.665399115	3.984523055
spinal cord	0	0.374623015	6.555970102	3.816470814	4.054688656	0	0.529844749	0.020240874	3.532849165	0.36837159	0	2.899959482	1.204872472
ear cartilage	0.103888854	0	1.723159734	5.477792674	0.473531193	1.938874602	0.644717239	2.07648309	2.615927848	0.123064897	0.133560503	3.788115166	0.40246664
ear tip	2.697859441	5.199579694	2.58708282	4.675513343	1.711850594	3.075465823	0.814532665	0.396620186	1.503211535	1.193018933	11.47542623	4.398175564	1.791957956
optic nerve head glial	4.130895559	8.012365116	1.665723362	3.485028822	0.400480608	1.551928804	0.247193444	0.058079205	4.050267469	0.184663749	0.013562176	4.736100557	2.687600024
retina	4.87612993	8.045708036	4.178402823	4.258174233	1.512735409	4.36109265	0.314914546	0.039583889	3.898728626	0.436863555	0.047022265	2.713403406	4.612167496
iridiocorneal angle	0	0	0	3.398316494	1.340508557	0	0.832811799	0.294549849	3.890824087	0.021837031	0.167713357	4.251573057	0
thymus	0	0	0	2.877720336	0	0.053402831	0.193039961	0.157927938	2.951650685	0	0	5.682280737	0.275646391
lung	0.46317526	1.08688907	1.00873704	4.4408657	0	3.890849264	2.986203115	0.073203084	2.716721885	0	0.036791184	3.877821731	0.02499963
blood	0	0	0	0.006915628	0.004805397	0.001327058	0	0.001050701	0	0	0	0.507191811	0
pbmc adult	0.074872749	0.063952512	0.502852639	1.648542894	0.074860346	0.082377222	0.074963936	0.028480146	1.806553716	0.036201871	0.260477102	2.751137087	0.043494327
pbmc kitten	0.028385287	0	0.12101391	0.275256019	0.026084545	0.034435013	0.006286765	0.002297281	1.584521711	0.020367796	0	2.368045487	0.018412154
peripheral blood	0	0	0.515689299	0.170624486	0.000698519	0.006921187	0.009584428	0.061271917	2.006770673	0	0	1.438811972	0.003684287
heart	0.029440623	0.063653346	0.446935034	3.011956496	0.021684551	0.60214092	0.712725987	0.406792622	2.349859914	0	0.109389212	3.178257348	0.056555256
liver	0.002312635	0	0.020241341	1.302396352	0.000849826	1.797793796	0.01719273	0.102107549	0.818617394	0.001655185	0	3.134084928	0.00448134
pancreas	0	0.013125298	0	0.496673638	0	1.264438441	0.021037088	0.006750463	0.281244484	0	0.017165925	1.504350567	0
spleen	0.012814297	0.095348535	0.25874521	2.535923479	0.028089502	0.571690571	0.093113907	0.028619065	1.858826078	0.089280815	0.024382448	2.679015895	0.016545063
kidney	0.096950442	0.362604563	3.114634737	0.733737527	0.043426715	0.056746401	0.215789184	0.42457845	2.804342234	0.117904886	0.028226921	1.230818933	2.817516663
small instestine	0.02306041	0.022430093	0.919040713	3.996618633	0.024456245	2.577793702	0.093414428	0.128730281	1.772402464	0.199878752	0.009835045	4.093872667	0.005000054
uterus	0	0	0.205531703	4.085978971	0	1.742410888	0.315615181	4.607945819	2.968203859	0.035454435	0	4.719142051	0.032066243
ovarian cortex	0.04283244	0.276187361	0.582670387	3.53684483	0.843720594	0.982349469	0.448130552	1.012993997	2.856184178	0.172768503	0.096062791	3.784122401	2.362116768
ovarian follicles	0.00495286	0.230281623	0	1.531380368	0.345167523	1.761071281	0.266114162	0.171579633	0.76826987	1.00405354	0.336193392	2.984532996	3.591937717
ovarian granulosa cell	0.008991358	0	1.641464675	1.34165827	0.003308901	0.007291401	0.008893111	1.374887713	1.83266614	0.06930108	0	1.809366183	0.017385505
sorted germ cells	0.642845699	0.171782823	1.686028601	4.851533678	0.507402211	1.575586521	1.613303246	0.984520286	2.500689812	0.319754165	0.724100131	3.873686288	0.73733179
testes	0.299475317	0	0.465669398	3.948613829	0.451191728	2.258861052	0.93877363	0.454226974	2.136381202	0.185946121	0.413609009	1.801861257	0.968554674
testis	0.24053022	0.125749477	1.0236267	3.966700341	0.83726628	1.840459266	0.733582575	0.640244481	2.440294652	0.169048244	0.723056047	2.635496549	0.961670128
seminiferous tubules	0.384927802	0.035315508	1.033753649	5.559293771	0.280038998	1.628919203	0.768256782	0.664462007	2.059442182	0.107003835	0.753137427	2.890890396	0.720638172
whole testes	0.242898475	0.107476514	0.504191868	3.382755818	0.445615224	2.131378771	0.599954091	0.501315507	2.128284656	0.053722971	0.727927623	3.216319425	0.293530387
embryo fetus	0.191399234	0.084838436	1.583090233	3.563052317	0.6927721	0.201677545	0.789584671	0.1446442	2.490336787	0.429363083	0.141768181	4.068385141	2.890874164
head embryo	0.277447948	0.107457441	1.063088406	4.490549175	1.944556238	0.246320218	0.978889696	0.151267039	3.035272045	1.711925434	0	4.256039654	4.241113543
body embryo	0.29294407	0.127453247	3.118263135	4.178248004	2.802053829	0.227815419	1.051822179	0.217242241	3.044565518	1.572135718	0.009663073	4.408907278	3.509347983

Figure 5

Table S1: Alignment results of long-read sequencing reads of CRFK and PG-4 cells

(File: Table_S1)

Alignment results of long-read sequencing reads of CRFK and PG-4 cells with WT and KIT-FERV1 sequences of KIT intron 1.

Table S2: Gene expression levels in various tissues

(File: Table_S2)

Gene expression levels (transcripts per million) in various tissues estimated from domestic cats (Felis catus) in the Ensembl database (https://ftp.ensembl.org/pub/data_files/felis_catus/F.catus_Fca126_mat1.0/rnaseq/)

Table S3: Lists of polymorphisms in MHC class I and II regions of CRFK and PG-4 cells.

(File: Table_S3.zip)

Variant Call Format (VCF)[72] files, text files providing lists of SNPs and small indels, for MHC class I region in CRFK (a) and PG-4 (b) cells, and those for MHC class II regions in CRFK (c) and PG-4 (d) cells. These VCF files were obtained as follows: 1) Similar to "Mapping of paired-end reads and quantification" in the Methods section, .bam and .bam.bai files of the mapping results of the original paired-end short read data of the whole genome sequencing to the reference genomes were obtained. Here, genome sequencing data were mapped to the reference genomes "Felis catus FLA distal class I region genomic sequence" (GenBank: EU153402.1) and "Felis catus FLA extended class II, class II, class III, proximal and central class I region genomic sequence" (GenBank: EU153401.1) for MHC class I and II, respectively. 2) Using the "MarkDuplicates" program of Picard (ver. 2.27.4) to obtain .bam files with the identification of duplicate reads. 3) By applying the HaplotypeCaller program of GATK (ver. 4.3.0.0) to previously obtained .bam file and .fasta file of reference genomes with default parameter values, VCF files of SNPs and small indels lists were obtained.

Table S4: Lists of polymorphisms in mitochondrial DNA of CRFK and PG-4 cells.

682 (File: Table_S4.zip)

683 VCF[72] files for the mitochondrial DNA in CRFK (a) and PG-4 (b) cells. VCF files were
684 obtained similarly to those presented in Table S3 using “Felis catus mitochondrion, complete
685 genome” (GenBank: NC_001700.1) as the reference genome.