

UNIVERSITA' DEGLI STUDI DI MILANO



PhD Course in Veterinary and Animal Science
Class XXXII

Department of Veterinary Medicine (DIMEVET)

Next Generation Genomic and Proteomic Analyses in Feline Amyloidosis

Francesca GENOVA

R11547

Supervisor: Prof. Maria LONGERI, PhD

Co-supervisor: Prof. Leslie LYONS, PhD

PhD Course Coordinator: Prof. Valeria GRIECO

Academic Year 2018-2019

“Science has discovered that, like any work of literature, the human genome is a text in need of commentary, for what Eliot said of poetry is also true of DNA: ‘all meanings depend on the key of interpretation’. What makes us human, and what makes each of us his or her own human, is not simply the genes that we have buried into our base pairs, but how our cells, in dialogue with our environment, feed back to our DNA, changing the way we read ourselves. Life is a dialectic”.

Jonah Lehrer

Proust Was a Neuroscientist

ABSTRACT

The advent of genomics in human health-care management has led to the introduction of a new powerful tool for diagnosis, allowing for genetic variants potentially associated to a disease phenotype to be identified by using the genome sequence of a single patient. The development of improved technologies together with the cost reduction, made the whole genome sequencing (WGS) feasible for domestic animals as well. Cats are among the most common pets and share different diseases with humans, representing therefore a valuable biomedical model for studies focused on human disorders. The 99 Lives Cat Genome Sequencing Initiative was launched in 2015, with the main aim of collecting 99 whole genome cat sequences through the collaboration of different researchers and institutes worldwide. Up to date, almost 200 cats of different breeds and random bred cats have joined the project and have their whole genome sequences as part of the database. Some disease-causing variants have already been identified thanks to the sequencing data available from the database, while many others are currently under investigation. The present PhD project shows how WGS is suitable for different kinds of research, including the detection of disease-associated variants and genomic breed descriptive studies.

The main part of the present PhD project has been focused on the investigation of the pathogenic pathways characterizing feline amyloidosis, a common disease in both the Abyssinian and Siamese breeds. The whole genome sequences of two affected Siamese and two affected Abyssinians were used to detect amyloidosis associated variants and were compared to the WGS data of the other cats present in the 99 Lives database for variants exclusion. Each breed was analyzed separately. At first, all the analyses were carried out referring to the version 6.2 of the cat genome assembly and using the WGS of 113 cats. Amyloidosis resulted to be a polygenic disease and variants underlying the affected phenotype were not shared between Abyssinian and Siamese cats, leading to the hypothesis of different mechanisms involved in the disease onset in the two breeds.

With the advent of the version 9.0 of the *Felis catus* genome in November 2017, and the increased number of cats that joined the 99 Lives Project, the analysis was repeated using the new genome assembly and 195 cats as controls for variants filtering. Moreover, as in the first part of the work feline amyloidosis turned out to be a multifactorial complex disorder, the new genomic analyses were further integrated with both proteomic and miRNAomic approach in the Abyssinians. When the variants detection with the version 9.0 of the cat genome was repeated, most of the variants related to the affected cats mapped into intronic or intergenic regions, and the variants mapping within exonic regions did not match those found with version 6.2. However, through the multi-omics approach conducted in the Abyssinian breed, different miRNAs resulted to be differentially expressed between healthy and affected Abyssinians. In particular, almost all the downregulated miRNAs were already reported in the Alzheimer Disease (AD) literature, a disorder having many features in common with amyloidosis. Additionally, these miRNAs targeted proteins expressed exclusively in the affected cats and one gene resulted mutated in the binding site of one of the most significantly downregulated miRNA.

Finally, the sequencing data of the 99 Lives Project were used to detect Copy Number Variants (CNVs) in 41 cats belonging to 14 different breeds. This genomic structure characterization, which represents the first study mapping CNVs in domestic cats, showed that individuals belonging to different breeds can cluster together based on their CNVs similarities.

Overall, the present PhD project demonstrated how WGS data available through the 99 Lives Initiative represent an important source for genomic analyses in cats, providing useful data for research and bridging the gap between the new frontiers of genomics and cats health-care.

Table of Contents

ABSTRACT V

TABLE OF CONTENTS VII

LIST OF ABBREVIATIONS IX

CHAPTER 1. INTRODUCTION 3

- 1.1. Genomics and Precision Medicine
- 1.2. Genomics in Animals
- 1.3. *Felis catus* Genome
- 1.4. 99Lives Cat Genome Sequencing Initiative
- 1.5. Study Aims and Outline
- 1.6. References

CHAPTER 2. FELINE AMYLOIDOSIS 21

- 2.1. Introduction
- 2.2. Genomic localization and genetic characterization of amyloidosis in the Abyssinian cat
International Conferences Presentation
- 2.3. Genomic localization and genetic characterization of amyloidosis in the Siamese cat
International Conferences Presentation

2.4. Characterization of Abyssinian Amyloidosis Through a Multi-OMICS Approach

Manuscript in preparation for submission

2.5. Localizing the Regions of Causative Variants in Siamese Amyloidosis Through an NGS Approach

Ongoing Project

2.6. References

CHAPTER 3. FIRST GENOME – WIDE CNV MAPPING IN FELIS CATUS USING NEX GENERATION SEQUENCING DATA 151

F. Genova, M. Longeri, L. A. Lyons, A. Bagnato, the 99Lives Consortium and M. G. Strillacci.

BMC Genomics (2018) 19:895

CHAPTER 4. CONCLUSIONS 167

4.1. Summary of Results

4.2. Concluding Remarks

ACKNOWLEDGEMENTS XI

List of Abbreviations

3'UTR	3' Untranslated Region
AA	Amyloid A
aCGH	array-based Comparative Genomic Hybridization
AD	Alzheimer Disease
AL	Amyloid Light Chain
APP	Amyloid Precursor Protein
AU-P	Approximately Unbiased P-value
BP-P	Bootstrap Probability Value
CNV	Copy Number Variant
CNVR	Copy Number Variant Region
DP or RD	Read Depth
ECM	Extracellular Matrix
FDR	False Discovery Rate
FFPE	Formalin Fixed Paraffin Embedded
GO	Gene Ontology
GQ	Genotype Quality
GWAS	Genome Wide Association Analysis
HGP	Human Genome Project
HWE	Hardy-Weinberg Equilibrium
IBD	Identity By Descendent
KEGG	Kyoto Encyclopedia of Genes and Genomes
LC-MS	Liquid Chromatography Mass Spectrometry
LCV	Large Scale Copy Number Variant
MALR	Multivariate Additive Logistic Regression
MDS	Multidimensional Scaling
miRNA	micro RNA
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NGS	Next Generation Sequencing
NIH	National Institute of Health
OR	Odd Ratio
OSH	Oriental Shorthair
PCA	Principal Component Analysis
SAA	Serum Amyloid A
SNP	Single Nucleotide Polymorphism
TMM	Trimmed Mean of M-values
TTR	Transthyretin
UPGMA	Unweighted Pair-Group Average
VCF	Variant Call Format

WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
WL	Window Length

CHAPTER 1.

INTRODUCTION

CHAPTER 1. INTRODUCTION

1.1. Genomics and Precision Medicine

The advent and the development of the Human Genome Project (HGP) in the 1990's brought new insights in the field of the molecular medicine. The HGP was an international research program which led to the complete mapping of all the human genes and revolutionized the concept of genetics, introducing the concept of genomics as a new approach for disease detection. Until that moment, genetics had taught to study the inheritance and the genes as the main contributors in this process. Genomics introduced the possibility to look at all the genes belonging to an organism at the same time, analyzing therefore its genome. Through the HGP it was possible to sequence all the bases of the human DNA and annotate the location of the genes along these sequences. One of the major finding of the HGP was that the number of genes was significantly lower than previous estimates, with about 20,500 human genes in total (<https://www.genome.gov/human-genome-project/What>).

Applications of the HGP particularly involved medicine technologies and research, providing new approaches to diagnose, prevent and cure different diseases. It was the starting point for what is known now as “precision medicine”. Precision medicine is a brand-new approach which aims to consider the patient in his individuality, taking into account all the aspects that could specifically contribute in the development of a particular phenotype. It allows to diagnose, treat and prevent diseases, considering the individual genetic variability, the environmental incidence and the lifestyle, being therefore very useful in the prediction of the most suitable treatment or in the determination of drug dosages.

Whole Exome Sequencing (WES) has represented the main tool for the identification of rare coding variants associated to diseases, but it is a limited approach when the disease is caused by the combination of coding and non-coding variants. Polygenic diseases are characterized by several contributing variants of small effects that could also be located in non-coding, regulatory regions of the genome. That explains why the WES is not always sufficient to find all causative the

variants [Bamshad *et al.* 2011; Rabbani *et al.* 2012]. The Whole Genome Sequencing (WGS) approach improved the identification of the causal variants considering all the variants along the genome. WGS allowed to identify de novo individual variants associated to a phenotype, building a genome variants database to collect and to compare all the individuals. WGS is already a complementary powerful tool for the diagnostic process, and it was already shown that the genome medicine can be applied also in short time for patients with acute medical conditions [Westerink *et al.* 2014; Saunders *et al.* 2012; Miller *et al.* 2015].

The HGP has led to the development of new tools to obtain the whole genome sequence of an individual in less than one month. The cost to sequence the whole genome of a single individual has reached the \$1,000 and there are now several worldwide centers which use WGS to detect congenital disorders. A further step in this direction will be represented by the application and the adaptation of this technology in other species.

1.2. Genomics in Animals

Several efforts followed the HGP to obtain and characterize the whole genome sequences of other organisms. It is common that different species share homologous genes with similar functions and could therefore represent model organisms to better study and characterize diseases affecting human as well. Within a decade after the HGP, the genomes of mouse and rat were completed [Mouse Genome Sequencing Consortium *et al.* 2002; Gibbs *et al.* 2004]. In fact, mouse and rat are well known for their easy experimental manipulation, their high proliferation and their high rate of spontaneous mutations, which made them excellent models for studies concerning human diseases [Rossant & McKerlie 2001; Liu *et al.* 2003]. Moreover, these two species are characterized by a high homology with human: “the divergence rate is low enough that one can still align orthologous sequences, but high enough so that one can recognize many functionally important elements by their greater degree of conservation” [Mouse Genome Sequencing Consortium *et al.* 2002]. This meant that all the missing information in the human genome could be retrieved from the studies in these other species.

After the accomplishment of mouse and rat's WGS, the initial sequence of the chimpanzee genome and its comparison with the human one were carried out [The Chimpanzee Sequencing and Analysis Consortium 2005]. By the way, the four mammalian genomes obtained until that moment represented only one clade out of the four clades belonging to placental mammals.

Among the domestic animals, the *Gallus gallus* was the first to be whole genome sequenced in 2004. The chicken represented an important model organism for new insights on vertebrate genome evolution, but it also gave an important contribution in the mammalian genome annotation [International Chicken Genome Sequencing Consortium, 2004].

It was in 2005 that the first whole genome sequence of the *Canis familiaris* was published. The dog was considered by the National Institutes of Health (NIH) due to its breeding history, its behavioral traits and its position within the mammalian evolutionary tree, that made the dog a valuable source for comparative analysis of the human genome [Lindblad-Toh *et al.* 2005]. Moreover, the close relationship

between humans and dogs, made this domestic animal the most studied in the medical practice thanks to the huge amount of information available through the veterinary surveillance [Patterson D. 2000]. The dog genome sequence signed a further step in comparative genomics, helping in the improvement of the mammalian genome analysis and particularly in the human one. After this achievement, a light was shed on further applications in domestic animals to characterize, complete and increase the genomic studies in mammalian genome evolution. Shortly, the genomes of the cattle [The Bovine Genome Sequencing and Analysis Consortium *et al.* 2009], sheep [The International Sheep Genomics Consortium *et al.* 2010] and pig [Groenen *et al.* 2012] were released, determining the advent of the genomics era in animals too.

1.3. *Felis catus* Genome

Cats, like dogs, have become pets par excellence, with 47.1 million of cat owners only in the United States, according to the American Pet Products Association's 2017-2018 National Pet Owners Survey. With an annual payroll of ~ \$16 billions for pet general maintenance (U.S. Census Bureau – U.S Department of Commerce, Economics and Statistics Administration, https://www.census.gov/library/visualizations/2017/comm/national_pet_day), cats benefit of an extensive veterinary medical surveillance which has allowed to identify about 200 genetic diseases analogous to human disorders [O'Brien 2004].

The importance of sequencing the whole genome of the domestic cat was highlighted in 2002, when the proposal of a Feline Genome Project came forward [O'brien *et al.* 2002]. The costs to whole genome sequence an individual with a low coverage (5x) was still high and selecting a species to be whole genome sequenced required considerable biological reasons. In this context, the cat represented a perfect candidate, with a karyotype of 18 autosomal chromosomes and the XY sex chromosome pair, for a total of 38 chromosomes, in which gene arrangement was very similar to the human one. Cat represented a valuable model for human hereditary diseases, since many of the disorders detected in cats were found to have a homology with human genetic defects [O'brien *et al.* 2002]. Moreover, *Felis catus* provided a powerful model for infectious diseases [Hardy *et al.* 1969; Hardy W. 1993].

In addition to all these important aspects for both human and veterinary medicine, the whole genome sequence of the domestic cat could also offer new insights into comparative genomics, conservation genetics and evolutionary studies.

In 2005, the NIH commissioned a 2x coverage WGS of 26 mammals and the cat was included in the project. The genome of Cinnamon, a female Abyssinian cat, was sequenced with a whole genome shotgun approach at the Broad Institute and Agencourt. Despite the light coverage, this first attempt to sequence the cat genome led to the identification of highly conserved sequence elements, revealing patterns of conservation and divergence across the mammalian radiation [Pontius *et al.*

2007]. The light 2x genome coverage together with a comparative approach involving references to six other mammalian genomes, allowed to assemble, map and annotate the first *Felis catus* genome, with the identification of ~20,285 feline orthologous genes [Pontius *et al.* 2007]. Several applications followed the information that came from this first cat WGS attempt. Genomic data were used to discover genes associated to different diseases and to study cat domestication and species conservation, increasing also the knowledge of the major lineages of mammals [O'brien *et al.* 2008].

Later, the *Felis catus* assembly Fca-6.2 with annotated genome features was displayed. The Genome Institute at Washington University School of Medicine, St Louis, MO sequenced Cinnamon genome with an additional 12x coverage of 454 reads to close the gaps of the previous sequence. A comparative approach led to the identification of 21,865 protein-coding genes, 217 loci of endogenous retrovirus-like elements, 99,494 new SNVs, 8,355 new indels, 743,326 evolutionary constrained elements, and 3,182 microRNA homologues [Tamazian *et al.* 2014]. This work brought a huge amount of new information into domestic cat genome studies and the sequencing data were used to better investigate the aspects concerning the adaptation, the evolution and the behavior of this species [Montague *et al.* 2014].

After the release of the version 6.2 of the cat genome assembly, additional data were obtained using the Illumina platform with a 20X genome coverage. These sequencing data were combined with the Sanger and 454 data produced with the 6.2 assembly. The *de novo* assembly version 8.0 was obtained and merged with the previous version 6.2 to create the final assembly Fca-8.0. With the alignment and the assemble of Illumina reads, also the small gaps (<1kb) were filled (https://www.ncbi.nlm.nih.gov/assembly/GCF_000181335.2).

Recently, the brand-new version of the cat genome assembly was released. Cinnamon DNA was isolated from a fibroblast cell line and a 72x genome coverage was generated at the McDonnell Genome Institute at Washington University School of Medicine, St Louis, MO, and the Texas A&M Agrilife Genomics core facility, using

the PacBio long reads method. The new assembly was aligned to the previous one to refine order and orientation information and the final version Fca-9.0 was obtained (https://www.ncbi.nlm.nih.gov/assembly/GCF_000181335.3).

In the version 9.0 of the cat genome, the sequence length is 2,521,863,845 bp, for a total size of 2.5Gb. The protein coding genes are 19,735 and the non-coding ones are 11,669. A comparison among the three versions of the cat assembly is reported in **Table 1**.

Table 1. Features and annotation of the three cat genome assembly versions (<https://www.ncbi.nlm.nih.gov>).

	Fcat-6.2	Fcat-8.0	Fcat-9.0
Date of release on NCBI	December 2011	November 2014	November 2017
Total sequence length	2,455,541,136	2,641,342,258	2,521,863,845
Total ungapped length	2,364,296,207	2,599,716,822	2,476,453,204
Gaps between scaffolds	900	303	19
Number of scaffolds	6,400	267,928	4,525
Scaffold N50	4,658,941	18,072,971	83,967,707
Scaffold L50	154	45	11
Number of contigs	203,083	367,672	4,909
Contig N50	20,621	45,189	41,915,695
Contig L50	33,87	16,252	19
Total number of chromosomes and plasmids	20	20	20
Number of component sequences (WGS or clone)	203,254	367,856	4,998

1.4. 99Lives Cat Genome Sequencing Initiative

The 99 Lives Cat Genome Sequencing Initiative (<http://felinegenetics.missouri.edu/ninety-nine-lives>) is a public effort aimed to improve coverage and assemblies of the cat genome. The 99 Lives project has a centralized resource with genome sequences produced of similar quality and similar techniques to facilitate the identification of normal and abnormal genetic variations in cats, bringing new insights into health concerns in cats as in humans.

The 99 Lives Cat Genome Sequencing Initiative has been promoted to create a collaborative and supportive group for people interested in cat health and genetics and to provide individual cat genome sequences that could be used for the developing of precision medicine in cats. Just as in human medicine, mapping the cat genome allows to detect and analyze the variants that make a single cat unique, helping with the identification of rare diseases. More than 100 genetic tests are now available for disease screening in cats and dogs but, if a variant associated to a disease is still unknown, the WGS could be the best approach to identify the variant, even if different factors, such as the mode of inheritance, need to be considered. This is even more important when polygenic diseases are considered, and more than one variant are involved in the determination of the pathology.

The advances in human medicine have led to a consistent decrease of the WGS costs, and that is one of the reasons that made possible the application of these approaches in dog and cats as well.

When the 99 Lives Project was launched, the main aim was to bring together universities, researchers and institutes worldwide, to collect 99 whole genome cat sequences with 30x genome coverage. Up to date, the number of sequenced cats in the 99 Lives database is close to 200 (**Table 2**). These cats represent different genetic traits that could involve diseases, or they could simply represent specific breed characters. Together with cats of different breeds and random bred cats, also several wild felids joined the project, allowing studies on comparative genomics and species conservation.

Table 2. List of the number of cats per breed in the 99 Lives Database.

Breed	Cat/breed	Breed	Cat/breed	Breed	Cat/breed
Abyssinian	4	Himalayan	1	Peterbald	1
American Shorthair	1	Korat	2	Ragdoll	4
American Wirehair	1	LaPerm	1	Random bred	55
Australian Mist	1	Lykoi	1	Savannah	1
Bengal	5	Maine Coon	11	Selkirk Rex	1
Bengal x Japanese Bobtail	3	Maine Coon Cross	20	Siamese	7
Bengal x Somali	1	Napoleon	3	Siamese cross	1
Birman	10	Ocicat	1	Siberian	2
British Shorthair	2	Oriental	4	Sphynx	3
Burmese	13	Oriental Shorthair	3	Tennessee Rex	1
Cross-breed colony	4	Oriental Shorthair cross	3	Tonkinese	1
Devon Rex	5	Persian	3	Toybob	2
Domestic Longhair	1	Persian cross	1	Toyger	1
Don Sphynx	1	Persian x Bengal	2	Turkish Angora	1
Egyptian Mau	3	Persian x Japanese Bobtail	1	Unknown	4
				Total	197

The public dataset available thanks to the 99 Lives Initiative has already led to the discovery of different disease associated variants, such as AIPL1 variant, causing progressive retinal atrophy in Persian cats, and HES7 variant causing the bobbed tail of the Japanese Bobtail breed [Lyons *et al.* 2016]. Another goal of the 99 Lives project is the detection of an insertion in a gene involved in the FAS pathway in British Shorthair cats with autoimmune lymphoproliferative syndrome [Aberdein *et al.* 2017] and the application of WGS in wild felid species for health care management [Mauler *et al.* 2017]. This is very important as the introduction of genetic tests in zoo populations may be essential to prevent diseases in endangered species.

Sequencing the genome of just one cat and joining the 99 Lives project, allows the open access to all the data of the cats in the 99 Lives database, even prior to

publication. This means that a feline patient with a particular unknown condition could be easily genome sequenced and the sequence can be compared to all the other whole genome sequences in the database (almost 200 cats), allowing the identification of the specific variant that can be related to the cat's condition. This leads also to the development of a specific treatment, bringing precision medicine into cat health management too.

More than 40 genes with approximately 70 DNA variants were already known to be associated to phenotypic traits or specific diseases in the domestic cat [Lyons L. 2015]. Up to date, there are 108 likely causal variants in *Felis catus*, according to OMIA database (<http://omia.org/>). The screening of these variants together with novel identified variants, could help with the veterinary surveillance in cats, leading to more effective treatments and to better diagnosis also for rare diseases.

Even if the tools and the bioinformatic techniques available in human and veterinary medicine cannot be compared, the resources now available for cat's genetics and genomics studies can be considered strong and cost efficient. In this context, the 99 Lives project represents a valuable database which mainly aim to bring personalized medicine into companion animals healthcare.

1.5. Study Aims and Outline

The 99 Lives Initiative has led to the resolution of different genetic diseases but many projects, concerning complex diseases, are still unsolved. Feline amyloidosis is among these projects, with four representative individuals included in the 99 Lives database: two affected Siamese and two affected Abyssinians.

The main aim of the study conducted during the PhD, was to use the cat whole genome sequences available through the 99 Lives database, together with other genomics approaches, to identify the causative variants of feline amyloidosis in both Abyssinian and Siamese breed.

For both the breeds, the work was planned to be developed in different steps:

1. Classify all the available samples (blood, buccal swabs, frozen tissues and FFPE) and information (necropsy report and pedigrees) collected in the recent past.
2. Validate protocols for a good DNA extraction from the specimens, especially for Formalin – Fixed, Paraffin – Embedded samples (FFPE) which enriched the case samples.
3. Perform a case-control genome wide association analysis (GWAS) on a private Feline DNA chip array to identify regions of the genome potentially associated to amyloidosis.
4. Screen for amyloidosis variants through the 99 Lives database, using the whole genome sequences of two affected Abyssinians and two affected Siamese, against the whole genome sequences of all the other cats present in the database.
5. Genotype the identified variants in a population of cats of the same breed, and in a control population of ~200 cats of different breeds and random bred cats, to confirm the association of the variants to the disease.
6. Analyze the total data.

For steps 3, 4 and 5, the project was conducted at the University of Missouri (USA), in collaboration with Dr. Lyons' Feline Genetics and Comparative Medicine Laboratory.

The samples for the present study came from Abyssinian/Somali and Siamese/Oriental cats, both affected and healthy, with unquestionable diagnosis (histological report and Congo Red staining). A great part of these samples was already available for the Abyssinian breed while, especially for the Siamese breed, samples were collected during these years from different pathology units (mainly University of Milan, Columbia-Missouri, Bristol, Sidney, Utrecht, London).

The Abyssinian breed was the first breed to be investigated. The first WGS analyses were conducted using the version 6.2 of the *Felis catus* genome assembly as the reference genome, and the whole genome sequences of the two affected Abyssinians were compared to other 113 whole genome cat sequences available in the database to detect the potential variants associated to amyloidosis. The identified variants were then genotyped in an Abyssinian population and through a statistical analysis, the associated variants were detected. Siamese underwent the same procedure, but the variants identified through the WGS analysis could not be genotyped due to the very low number of cats composing the Siamese population. Therefore, while the Abyssinian study was under development, a worldwide call was made to collect more Siamese samples. Simultaneously, the NGS data available through the 99 Lives database were also used to perform a structural variants analysis and to detect the Copy Number Variants (CNVs) that could possibly be related to amyloidosis. This analysis brought to a parallel study in which CNVs were used to characterize 14 different cat breeds.

The advent of the new version of the cat genome assembly (Fcat-9.0), together with the increased number of cats joining the 99 Lives Initiative, led to reconsider the results obtained with this first study in the Abyssinian breed. Therefore, the WGS analysis was repeated using the new version of the assembly as the reference genome and the additional cats sequenced with vs9.0 to increase the control population for variants exclusion.

This time, given the previous inconclusive results obtained with the genomic analyses, it was decided to carry out a multi-omics approach in the Abyssinians, and the genomics analysis was integrated with both proteomics and micro RNA analyses (miRNA). The Abyssinian cats used in this study were selected among the ones already used in the previous work. The two whole genome sequenced affected cats remained the same and six Abyssinians were chosen for the proteomics and miRNA analyses based on their availability as FFPE specimens. The main aims of this project can be summarized as it follows:

1. Multi-omics approach:

- Amyloidosis related variants detection through the comparison of the whole genome sequences of two affected Abyssinians with the WGS of a control population of 195 cats, using the version 9.0 of the *Felis catus* genome assembly as the reference genome.
- Identification of differentially expressed proteins with a shotgun mass spectrometry, using FFPE samples of three affected and three healthy Abyssinians. Proteomics analysis was carried out by the Biochemistry group of the Department of Veterinary Medicine at the University of Milan.
- Isolation, sequencing and identification of differentially expressed miRNAs using the same six Abyssinians FFPE samples used for the proteomics analysis. MiRNA study was conducted in collaboration with the Bioinformatics group of the Institute of Biomedical Technologies at the National Research Council (CNR) in Segrate – Milan.

2. Matching all the resulting data to reconstruct the amyloidosis pathway.

Finally, also the Siamese breed was investigated using the version 9.0 of the cat genome assembly. The study is still ongoing but the detection of the amyloidosis related variants through WGS analysis has already been performed and ~100 Siamese samples have been collected and are now ready to be genotyped on a new high-density DNA array.

1.6 References

1. Aberdein, D., Munday, J.S., Gandolfi, B., Dittmer, K.E., Malik, R., Garrick, D.J., Lyons, L.A., and the 99 Lives Consortium, 2017. A FAS-ligand variant associated with autoimmune lymphoproliferative syndrome in cats. *Mamm Genome* 28, 47–55. <https://doi.org/10.1007/s00335-016-9668-1>
2. Bamshad, M.J., Ng, S.B., Bigham, A.W., Tabor, H.K., Emond, M.J., Nickerson, D.A., Shendure, J., 2011. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 12, 745–755. <https://doi.org/10.1038/nrg3031>
3. Chimpanzee Sequencing and Analysis Consortium, 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437, 69–87. <https://doi.org/10.1038/nature04072>
4. Chowdhary, B.P., Raudsepp, T., 2006. The horse genome. *Genome Dyn* 2, 97–110. <https://doi.org/10.1159/000095098>
5. Gibbs, R.A., Weinstock, G.M., Metzker, M.L., Muzny, D.M., Sodergren, E.J., Scherer, S., Scott, G., Steffen, D., Worley, *et al.*, 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428, 493–521. <https://doi.org/10.1038/nature02426>
6. Groenen, M.A.M., Archibald, A.L., Uenishi, H., Tuggle, C.K., Takeuchi, Y., Rothschild, M.F., Rogel-Gaillard, C., *et al.*, 2012. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491, 393–398. <https://doi.org/10.1038/nature11622>
7. Hardy, W.D., 1993. Feline Oncoretroviruses. In: Levy J.A. (eds) *The Retroviridae. The Viruses*. Springer, Boston, MA
8. Hardy, W.D., Geering, G., Old, L.J., De Harven, E., Brodey, R.S., McDonough, S., 1969. Feline Leukemia Virus: Occurrence of Viral Antigen in the Tissues of Cats with Lymphosarcoma and Other Diseases. *Science* 166, 1019–1021. <https://doi.org/10.1126/science.166.3908.1019>
9. International Chicken Genome Sequencing Consortium, 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432, 695–716. <https://doi.org/10.1038/nature03154>
10. International Sheep Genomics Consortium, Archibald, A.L., Cockett, N.E., Dalrymple, B.P., Faraut, T., Kijas, J.W., Maddox, J.F., *et al.*, 2010. The sheep genome reference sequence: a work in progress. *Anim. Genet.* 41, 449–453. <https://doi.org/10.1111/j.1365-2052.2010.02100.x>

11. Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., Clamp, M., *et al.*, 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438, 803–819.
<https://doi.org/10.1038/nature04338>
12. Liu, M.Y., Poellinger, L., Walker, C.L., 2003. Up-regulation of hypoxia-inducible factor 2alpha in renal cell carcinoma associated with loss of Tsc-2 tumor suppressor gene. *Cancer Res.* 63, 2675–2680.
13. Lyons, L.A., 2015. DNA mutations of the cat: The good, the bad and the ugly. *Journal of Feline Medicine and Surgery* 17, 203–219.
<https://doi.org/10.1177/1098612X15571878>
14. Lyons, L.A., Creighton, E.K., Alhaddad, H., Beale, H.C., Grahn, R.A., Rah, H., Maggs, D.J., *et al.*, 2016. Whole genome sequencing in cats, identifies new models for blindness in AIPL1 and somite segmentation in HES7. *BMC Genomics* 17, 265.
<https://doi.org/10.1186/s12864-016-2595-4>
15. Mauler, D.A., Gandolfi, B., Reiner, C.R., O'Brien, D.P., Spooner, J.L., Lyons, L.A., and 99 Lives Consortium, 2017. Precision Medicine in Cats: Novel Niemann-Pick Type C1 Diagnosed by Whole-Genome Sequencing. *J Vet Intern Med* 31, 539–544.
<https://doi.org/10.1111/jvim.14599>
16. Miller, N.A., Farrow, E.G., Gibson, M., Willig, L.K., Twist, G., Yoo, B., Marrs, T., *et al.*, 2015. A 26-hour system of highly sensitive whole genome sequencing for emergency management of genetic diseases. *Genome Med* 7, 100.
<https://doi.org/10.1186/s13073-015-0221-8>
17. Montague, M.J., Li, G., Gandolfi, B., Khan, R., Aken, B.L., Searle, S.M.J., Minx, P., *et al.*, 2014. Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. *Proc Natl Acad Sci USA* 111, 17230–17235. <https://doi.org/10.1073/pnas.1410083111>
18. Mouse Genome Sequencing Consortium, Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., *et al.*, 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420, 520–562.
<https://doi.org/10.1038/nature01262>
19. O'Brien, S.J., Johnson, W., Driscoll, C., Pontius, J., Pecon-Slattery, J., Menotti-Raymond, M., 2008. State of cat genomics. *Trends Genet.* 24, 268–279.
<https://doi.org/10.1016/j.tig.2008.03.004>
20. O'Brien, S.J., Menotti-Raymond, M., Murphy, W.J., Yuhki, N., 2002. The Feline Genome Project. *Annu. Rev. Genet.* 36, 657–686.
<https://doi.org/10.1146/annurev.genet.36.060602.145553>

21. O'Brien, S.J., 2004. Cats. *Curr. Biol.* 14: R988–R989.
22. Online Mendelian Inheritance in Animals, OMIA. Sydney School of Veterinary Science, {date of download}. World Wide Web URL: <http://omia.org/>
23. Patterson, D.F., 2000. Companion animal medicine in the age of medical genetics. *J. Vet. Intern. Med.* 14, 1–9.
24. Pontius, J.U., Mullikin, J.C., Smith, D.R., Agencourt Sequencing Team, Lindblad-Toh, K., Gnerre, S., Clamp, M., *et al.*, 2007. Initial sequence and comparative analysis of the cat genome. *Genome Res.* 17, 1675–1689. <https://doi.org/10.1101/gr.6380007>
25. Rabbani, B., Mahdieh, N., Hosomichi, K., Nakaoka, H., Inoue, I., 2012. Next-generation sequencing: impact of exome sequencing in characterizing Mendelian disorders. *J Hum Genet* 57, 621–632. <https://doi.org/10.1038/jhg.2012.91>
26. Rossant, J., McKerlie, C., 2001. Mouse-based phenogenomics for modelling human disease. *Trends Mol Med* 7, 502–507.
27. Saunders, C.J., Miller, N.A., Soden, S.E., Dinwiddie, D.L., Noll, A., Alnadi, N.A., Andrews, N., *et al.*, 2012. Rapid Whole-Genome Sequencing for Genetic Disease Diagnosis in Neonatal Intensive Care Units. *Science Translational Medicine* 4, 154ra135-154ra135. <https://doi.org/10.1126/scitranslmed.3004041>
28. Tamazian, G., Simonov, S., Dobrynin, P., Makunin, A., Logachev, A., Komissarov, A., Shevchenko, A., *et al.*, 2014. Annotated features of domestic cat – *Felis catus* genome. *GigaSci* 3, 13. <https://doi.org/10.1186/2047-217X-3-13>
29. The Bovine Genome Sequencing and Analysis Consortium, Elsik, C.G., Tellam, R.L., Worley, K.C., Gibbs, R.A., Muzny, D.M., Weinstock, G.M., Adelson, D.L., *et al.*, 2009. The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution. *Science* 324, 522–528. <https://doi.org/10.1126/science.1169588>
30. Westernik *et al.*, 2014. Diagnostic Clinical Genome and Exome Sequencing. *N Engl J Med* 371, 1169–1170. <https://doi.org/10.1056/NEJMc1408914>

CHAPTER 2.

FELINE AMYLOIDOSIS

CHAPTER 2. FELINE AMYLOIDOSIS

2.1. Introduction

The term amyloidosis refers to a heterogeneous group of metabolic diseases widespread in humans and animals and characterized by the abnormal deposition of amyloid proteins mainly in the extracellular space of various organs and tissues. These proteins typically have a high degree of β -pleated sheet secondary structure, forming insoluble fibrillary masses that are resistant to proteolysis and destroy structural integrity of the tissues [Holub *et al.* 2019, Sipe *et al.* 2016]. The amyloid deposits are characterized and officially categorized according to the “specific” amyloid fibril proteins they are composed of [Sipe *et al.* 2016]. In humans, 36 different extracellular fibril proteins types have been identified and the most frequent type of amyloidosis is the one characterized by the deposition of the immunoglobulin light chains (AL amyloidosis). Another common type of amyloidosis derived from the misfolded wild-type or mutated transthyretin protein (TTR) [Merlini *et al.* 2013]. Chronic infections and autoimmune inflammations with increased levels of serum amyloid A protein (SAA) may result in inflammatory amyloidosis or AA amyloidosis. Moreover, mutations in proteins such as fibrinogen α , apolipoprotein A-I, apolipoprotein A-II, apolipoprotein A-IV and lysozyme can lead to the hereditary systemic form of amyloidosis [Sipe *et al.* 2016, Sipe *et al.* 2014, Sipe *et al.* 2012]. The guidelines recommend the use of the terms “hereditary amyloidosis” when the mutation is “in the fibril protein gene itself”. Alternatively, the term “familial amyloidosis” should refer to a syndrome occurring “in a familial setting when the causal mutations occur in genes expressing non-amyloid proteins”, such as for the AA Amyloidosis [Sipe *et al.* 2016].

Amyloidosis in animals is rare. Cases are mainly characterized by a preponderance of protein AA in the deposits and are associated with chronic inflammatory, infectious or neoplastic diseases in domestic animals, such as bovine [Gruys E. 1977; Elitok *et al.* 2008] and ovine [Ménsua *et al.* 2003], wild mammals (Gazella dorcas [Rideout *et al.* 1989], Panthera tigris altaica [Schulze *et al.* 1998], Panthera leo [Williams *et al.* 2005], Mustela nigripes [Garner *et al.* 2007]), and birds [Guo *et al.* 1996; Ovelgönne *et al.* 2001]. Extremely rare cases of AL have been

recorded in horse [Andel *et al.* 1988; Kim *et al.* 2005] but they were inconclusive [Østevik *et al.* 2014], and in dogs and cats associated with plasmacytomas [Platz *et al.* 1997]. The only evidence of a genetic predisposition in developing amyloidosis in non-humans was found in the Shar-Pei dog breed (identified as the AA form) [Benson *et al.* 1985; Di Bartola *et al.* 1990; Segev *et al.* 2012] and it was also suggested in both captive cheetahs (*Acinonyx jubatus*, [Papendick *et al.* 1997; Zhang *et al.* 2008]) and black footed cats (*Felis nigripes*, [Terio *et al.* 2008]).

The domestic cat is the only reported domestic species to spontaneously develop amyloidosis. This means that in this species, the disease could be also considered as familial or as primary amyloidosis, since it does not always occur as a consequence of a preexistent inflammatory condition, but it could rise spontaneously. The familiarity of the disease was firstly described more than three decades ago in the Abyssinian [Chew *et al.* 1982; Boyce *et al.* 1984] and Siamese breeds [Van der Linde-Sipman *et al.* 1997; Niewold *et al.* 1999]. Three features are consistent in Abyssinian renal amyloidosis, including occurrence in related cats, systemic deposition of AA protein with a high presence in kidneys, and death from renal failure at a young age. In Siamese/Oriental cats the deposition occurs mainly in the liver causing jaundice, anemia, hemorrhage and rupture, but other organs can also be involved, such as spleen, heart, intestine, pancreas, lymph nodes, bladder and the adrenal glands. No data are available on the current prevalence of the disease in any cat breed. The disease was reported as “familial” [Niewold *et al.* 1999], but the mode of inheritance remains undetermined [Niewold *et al.* 1999; van Rossum *et al.* 2004]. The preeminent peptide in the amyloid fibrils in the domestic cat was sequenced from an affected Abyssinian cat with AA amyloidosis [Di Bartola *et al.* 1985]. The complete primary structure of the AA protein showed the amino terminal residues were heterogeneous among different species of mammals, indicating a high tolerance for mutations in this region [Kluve-Beckerman *et al.* 1989]. Sequencing the cDNA of the N-terminal region in affected Abyssinians, Siamese and Domestic Shorthair cats [van Rossum *et al.* 2004; Di Bartola *et al.* 1985; Kluve-Beckerman *et al.* 1989; Johnson *et al.* 1989], identified some non-synonymous substitutions, mainly in the Domestic Shorthairs, but these substitutions did not exhibit any clear association with the disease.

To devise an appropriate therapy, the identification of the amyloid fibril protein composition is essential. The presence of amyloid deposits is firstly confirmed

through histochemical staining such as Congo and Sirius red or metachromatic staining of the tissue samples obtained from an affected organ.

In the last decade, the mass spectrometry (MS)-based methods for amyloid typing has been developed, allowing an unbiased identification of the proteins characterizing the deposits [Lavatelli and Vrana 2011]. Several proteomic approaches are now considered new clinical standard for amyloid typing, either on FFPEs or on fresh unfixed specimens. Macrodissection or laser microdissection techniques are used to identify amyloid fibril composition of the major amyloid types, such as ATTR, AL, AA [Murphy *et al.* 2006, Vrana *et al.* 2009], as well as more rare amyloid types, such as AGel [Klein *et al.* 2011], ApoAI [Murphy *et al.* 2006, Rowczenio *et al.* 2011] and ApoAIV [Sethi *et al.* 2012]. Unlike the antibodies approach, the MS-based methods allow to identify the majority of the proteins within the dissected amyloid deposit and to determine the predominant fibril type causing the disease. Although the tissue preparation and Liquid Chromatography (LC)-MS/MS protocols varied slightly between laboratories, the proteomic analyses successfully identified both common and rare amyloid types [Lavatelli and Vrana 2011].

In humans, more recent studies have been focusing also on disclosing the potential role of miRNAs in the accumulation of amyloid fibrils [Weng *et al.* 2011; Patel *et al.* 2008] and it was shown that miRNAs did not play a direct role in the onset and the development of the disease, but they could have a significant role in specific alterations characterizing the disease itself. In cats, the identification of miRNAs is a rather recent achievement. The profiling of miRNAs in this species was carried out to underlie their potential differential expression in specific diseases [Weber *et al.* 2015; Cong *et al.* 2017], or to characterize their normal distribution considering specific organs such as kidneys [Ichii *et al.* 2014]. In a recent study, miRNAs from a feline kidney cell line were evaluated before and after infection with mink enteritis virus (MEV), offering a more detailed characterization of the kidney miRNAome in cats [Sun *et al.* 2014]. It was only in 2017 that the cat miRNAome was finally defined using normal feline tissues [Laganà *et al.* 2017].

Correct diagnosis of amyloidosis can only be achieved via biopsy or necropsy and predictive and prognostic diagnostic tools *in vivo* are still missing. Particularly, the necropsy represents the best way to determine if the amyloidosis is primary or if it is secondary to a chronic inflammatory process, but another issue is represented by

the economic and logistic costs that a necropsy could involve. In veterinary practice the analysis of the chemical nature of the specific amyloid fibrils is never carried out, making the classification of amyloidosis often approximate in domestic cat. All these considerations together make difficult to find informative samples and to properly characterize the pathology.

In the following four studies on feline amyloidosis, carried out through the PhD, different approaches have been used to identify the causative variants underlying primary amyloidosis and to reconstruct the pathogenic pathways. The first and the second study (page 25 and 69 respectively) focused only on the genetics of the disease through an NGS approach and were performed using the vs 6.2 of the cat genome. The release of the brand-new version 9.0 of the cat genome assembly, led to reconsider the results obtained and to perform again the variants detection. Therefore, in the third study (page 73), the genomics analysis in the Abyssinian breed was repeated and integrated with both proteomics and miRNAomics approaches. Finally, in the fourth study (page 125), the Siamese cats were investigated using the vs 9.0 of the genome. The genomic analysis for variants detection in this breed is reported but the study is not completed as amyloidosis in the Siamese breed is still under exploration.

2.2. Genomic Localization and Genetic Characterization of Amyloidosis in the Abyssinian Cat

Abstract

This study represents the first large-scale investigation of the genetic basis for amyloidosis in domestic animals, specifically renal amyloidosis within the Abyssinian breed. The worldwide sample collection lasted several years and allowed to perform the study on 115 Abyssinian cats. An integrated genomic approach using a Genome Wide Association Study, WGS (99 Lives initiative) and Genotyping, was performed to identify the potential causative mutations for amyloidosis. A total of 281 additional samples from other breeds and random bred cats were used for population screening to validate the variants resulted associated to the disease. A multivariate additive logistic regression model (MALR) was applied to verify the contributions of these variants to the disease risk. Through the MALR and based on the different genetic profiles found in the Abyssinian population, the probability of developing the disease was calculated for each cat in the study. Results show that the disease is not inherited in a Mendelian manner but is probably multifactorial. Single gene mutations in six different genes underlie the onset of amyloidosis in Abyssinian cats. These variants do not affect specific amyloid proteins, making the disease classifiable as familial. Disease probabilities can be used to classify Abyssinians into different risk categories. The genes identified as causative in Abyssinian amyloidosis are found in Alzheimer β -amyloid gene networks. These finding should be helpful for reconstructing the amyloidosis pathological pathway.

Materials and Methods

Samples

Samples were donated by owners or were collected as part of the routine health care of the cat, including diagnostic or monitoring purposes, under owner informed consent and in accordance with ethical statements when required (University of Missouri institutional Animal Care and Use Committee protocol (ACUC protocol # 9065, 9178), University of Milan protocol OPBA_56_2016). Ascertainments

occurred over the past 20 years, with a majority of samples collected within the last 10 years.

Abyssinian cats were considered affected when positive for Congo red staining of renal tissues [Pucktler *et al.* 1962], less than 5 years of age and with a confirmatory necropsy. Cats were considered to be healthy controls when (after death) individuals were documented as negative by a histological report for absence of amyloidosis and older than 5 years of age. As the necropsy/histological reports were not always retrievable, aged cats with a breeder statement of their good health and absence of any amyloidosis-related disorder were also included in the healthy controls. Data for each cat are presented in **Table 1** and **Supplementary Table 1**.

Table 1. Summary of cat populations used in each analysis for amyloidosis.

Breed	Name	Health	No.	Analysis	Aim
Abyssinian	Cohort	Cases	21	GWAS	Map disease regions
		Controls	23		
	Cat_ID3 Cat_ID12	Affected	2 (99 Lives)	WGS	Identify & prioritize disease associated variants
	Abyssinian Population	Affected	25	MALDI-TOF Genotyping	Confirm associated variants
		Healthy	40		
Other breeds and random bred cats	Control Population	Unknown	36	MALDI-TOF Genotyping	Screen breed for confirmed variants
			112 (99 Lives)	WGS	Screen species for confirmed variants
			169 (MALDI-TOF)	MALDI-TOF Genotyping	

DNA was isolated from 500 µl of EDTA blood, 5g of frozen tissue, 2-4 buccal swabs or FFPE specimens (10 slices of 8–10 µm). The DNeasy Blood & Tissue Kit and the QIAamp DNA FFPE Tissue Kit (Qiagen, Venlo, Netherlands) were used for DNA isolation according to the manufacturer's instructions. DNA was isolated by organic methods for the two Abyssinians used for WGS [Sambrook 1989].

A total of 79 Abyssinian/Somali cats, 30 affected and 49 controls were examined (**Table 1**, **Supplementary Table 1**). Among these individuals, two affected cats were whole genome sequenced (one from Europe, one from USA) via the 99 Lives Cat Genome Sequencing Initiative (www.felinegenetics.missouri.edu/99Lives) [Mauler *et al.* 2017]. Of the 79 Abyssinians, different subgroups were used in the genomic analyses: the GWAS was performed using a subset of 21 affected cases

and 23 healthy controls (Abyssinian Cohort) that had sufficient high-quality DNA for array genotyping. Genotyping by MALDI-TOF and the single locus analyses were performed on 25 affected cases and 40 healthy controls (Abyssinian Population), including 12 DNA samples extracted from FFPE tissues. Sixteen cases and 14 controls were examined using both the GWAS and the genotyping of candidate Single Nucleotide Polymorphisms (SNPs) discovered from the WGS. Thirty-six additional Abyssinians with unknown health status (Abyssinian Unknown) were used for a preliminary random screening of variants within the breed (**Table 1**). Archived samples from other pedigreed and random bred cats ($n = 169$) were used together with the genotypes already available from the WGS of 113 cats included in the 99 Lives project (Control Population) for the overall domestic cat population screening (**Table 1**).

Genome Wide Association Study (GWAS)

The case-control GWAS analysis was conducted using a 340K private feline Illumina iSelect DNA array (Hill's Pet Nutrition, Inc., Topeka, Kansas) on the Abyssinian Cohort. Approximately 600 ng of DNA from each cat was submitted to the Translational Genomics Research Institute (Phoenix, AZ) for generation of the array data. The GWAS was conducted as previously described [Alhaddad *et al.* 2014] using PLINK v1.07 [Purcell *et al.* 2007] with a minor allele frequency threshold of 0.05 for SNP inclusion and excluding variants with more than 20% of missing genotypes, as well as individuals genotyped for <80% of SNPs. Results were visualized using a Manhattan plot produced with R Studio [<http://www.rstudio.com/>]. False discovery rate (FDR) threshold was fixed at 0.05 level. The identity by descent analysis (IBD) was performed on the GWAS cat population using PLINK [Purcell *et al.* 2007]. Considering the presence of high inbreeding within the Abyssinian cat breed, individuals were included with an IBD ≤ 0.3 , as previously described [Gandolfi *et al.* 2012].

WGS

WGS was performed as a part of the 99 Lives Cat Genome Sequencing Initiative (www.felinegenetics.missouri.edu/99Lives) [Oh *et al.* 2017] and all the analyses were carried out using the *Felis catus* vs 6.2 genome assembly

(<http://www.ncbi.nlm.nih.gov/assembly/320798>) [Pontius *et al.* 2007] as the reference genome. Sequences for 115 domestic cats were available, including the two amyloidosis affected individuals and the reference genome cat, “Cinnamon”, so that both alleles detected at variable loci were considered for this cat for the subsequent variants detection. Starting from a single ~350 bp fragment library, approximately 30x genome sequence coverage was generated using 150 bp paired-end reads on an Illumina XTen. Sequencing was carried out at The McDonnell Genome Institute at Washington University in St. Louis as previously described [Gandolfi *et al.* 2015].

Variants detection

Generated sequences were processed following demultiplexion, trimming and alignment to the *Felis catus* vs 6.2 genome [Gandolfi *et al.* 2015]. Variant calling was performed using PLATYPUS and SNPEFF was used to predict variant effects [Aberdein *et al.* 2017]. The identification of the variants was conducted through the Variant Explorer tool on the Maverix Analytic Platform (Maverix Biomics, Inc. platform - Santa Cruz, CA, maverixbio.com). The sequences of the two affected Abyssinians were compared to the sequences of the other 113 cats in the 99Lives database, which were used as normal controls for variant exclusion. Particularly, WGS detected variants were filtered, and only those variants present either as homozygous or heterozygous in the two affected Abyssinians and with an allele frequency <1% in the 99 Lives database, were considered as candidates for further analysis. Only high and moderate impact variants were considered, according to the Ensembl classification of variant severity, available at https://m.ensembl.org/info/genome/variation/prediction/predicted_data.html.

Variant Validation and Genotyping

The WGS identified 51 candidate variants, according to the filters previously described. These 51 variants were genotyped in the Abyssinian Population using mass spectroscopy (MALDI-TOF) genotyping and variants were confirmed in the two whole genome sequenced affected cats who belonged to the Abyssinian Population as well. Primers were designed within 200bp flanking regions of the target variants using the MassARRAY Assay Design included in the package Typer

4.0 (Agena Bioscience, San Diego, CA) according to manufacturer's instructions (**Supplementary Table 2**). Genotypes were produced using the Agena Biosciences iPLEX Gold Genotyping kit as previously described [Oh *et al.* 2017]. Genotyping was arranged in two multiplexes using a Nanodispenser RS1000 (Agena Bioscience) and the MassARRAY System.

Variants found to be associated with risk of amyloidosis ($n = 6$) were also validated through direct Sanger sequencing in each of the two whole genome sequenced affected Abyssinians. Briefly, PCR primers specific to the templates (**Supplementary Table 3**) were designed using Primer3 (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>) and PCR products were genotyped by DNA sequencing both strands, using the ABI Dye Terminator Sequencing Chemistry 3.1 and an ABI310 analyzer (Applied Biosystems, Foster City, CA), according to standard protocols. The six prioritized variants were genotyped by mass spectroscopy in additional Abyssinians (Abyssinian Unknown) and in the Control Population using a single multiplex or genotyping the SNPs individually.

Statistical Analysis

Single locus statistical analysis and multivariate analysis investigated the association between the presence of the variants and the disease and evaluated the additive effect of the variants on the phenotype in the Abyssinian Population. Odds Ratios (OR) and Bonferroni correction were calculated using PLINK v1.07 [Alhaddad *et al.* 2014] to generate single locus statistics. ORs were calculated as risk of amyloidosis in heterozygotes plus homozygotes for one allele against the homozygotes for the other allele, for each risk locus individually. Samples with $OR > 1$ and with Bonferroni value < 1 were considered as associated and used in the further statistical analyses. Hardy-Weinberg Equilibrium (HWE) was tested for the variants with significant OR in the Abyssinian Population and in the Abyssinians with unknown health status.

In the MALR, data were recoded so that individuals that were homozygous for the common allele in the controls were scored as "0", heterozygotes "1" and homozygotes for the rare allele "2". Two missing genotypes for SNP45 were imputed by calculating the mean genotype for the cases and the mean genotype for the

control cats for this SNP. An Additive Logistic Model was used to fit additive allele dosage effects (linear regressions) on the individual loci and then, the MALR was used to fit all loci simultaneously. The probability of disease was calculated through the MALR Model as $P(\text{disease}) = \exp(B_0 + B_1\text{SNP1} + B_2\text{SNP2} + B_3\text{SNP3} + B_4\text{SNP4} + B_5\text{SNP5} + B_6\text{SNP6}) / (1 + \exp(B_0 + B_1\text{SNP1} + B_2\text{SNP2} + B_3\text{SNP3} + B_4\text{SNP4} + B_5\text{SNP5} + B_6\text{SNP6}))$ and the sign of each estimated regression coefficient (B_i) determines if the rare or common allele increases the risk of disease at each locus.

Results

Genome Wide Association Study (GWAS)

Twenty-one cases and 23 controls (Abyssinian Cohort, **Table 1, Supplementary Table 1**) were genotyped on an Illumina DNA array with 297,034 SNPs. Two affected cats and four control cats were eliminated for low genotype call rates (MIND > 0.2), 20,716 SNPs failed the missingness test (GENO > 0.2) and 132,301 SNPs were removed for low minor allele frequency (MAF < 0.05). Finally, 159,707 SNPs were analyzed for their association with risk of disease. Cats were selected to have no grandparents in common, however, 23 pairs in the GWAS cohort had a $\hat{p} > 0.3$ (**Supplementary Table 4**). No significant SNP associations were detected at a FDR adjusted p-value of 0.05 (**Figure 1**). The most significant SNPs with a $-\log_{10}(p) \geq 4$ are presented in **Supplementary Table 5**. None of these array SNPs are located within the coding regions of genes.

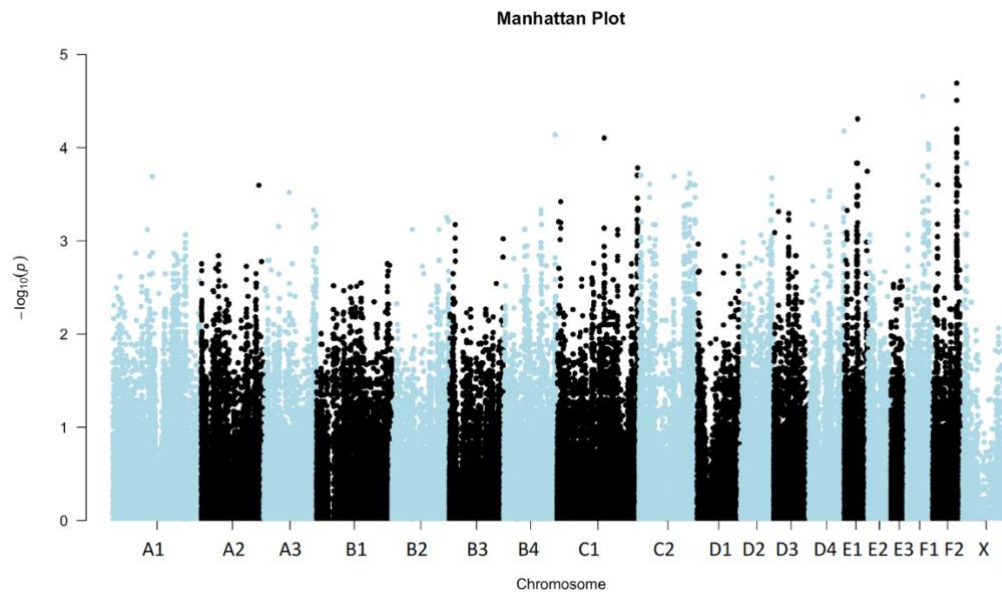


Figure 1. Manhattan plot for the Abyssinian GWAS Cohorts. The plot depicts Praw values for each SNP in the 21 case–23 control study using a 340K feline DNA array.

WGS

The whole genome sequences for the 115 domestic cats in the 99 Lives analysis have been submitted to the NCBI SRA under Bioproject PRJNA308208, except for Pennylane_18778 which is listed under Bioproject PRJNA288177. The two affected Abyssinian sequences are accessions SRR5373742 and SRR6997541 for the Dot_S16515 (Cat_ID 3) and the Pennylane_18778 (Cat_ID 12) bam files respectively. The whole genome sequences for the two affected Abyssinians were compared to the whole genome sequence for the reference cat, which is a healthy Abyssinian, and 112 cats generated in the 99 Lives project. Fifty-one SNPs were identified as tentative candidates for association with Abyssinian amyloidosis, based on both the assumptions of autosomal dominant and/or recessive inheritance. None of the genes containing the 51 SNPs are reported as amyloidosis-related. SNP locations and gene names are presented in **Supplementary Table 6**. All SNP variants were identified as having moderate effects on the protein and all were non-synonymous amino acid changes except SNP9, which was a synonymous alanine. No SNPs with high effects were discovered. SNPs with low and lowest effects were not examined (https://m.ensembl.org/info/genome/variation/prediction/predicted_data.html).

Variant Genotyping and Single Locus Statistics

The 51 WGS detected variants were validated in the WGS cats by MALDI-TOF genotyping (**Supplementary Table 7**) and then genotyped in 65 Abyssinians, including 25 affected and 40 healthy cats (Abyssinian Population, **Table 1**, **Supplementary Table 1**). Allele frequencies and ORs were calculated to determine each SNP's association to amyloidosis (**Supplementary Table 8**). Six variants were identified with significant ORs. Three SNPs were significant after Bonferroni correction (SNP26, SNP32 and SNP45) with p-values of 0.023, 0.021, and 0.032, respectively (**Table 2**).

Table 2. Disease-associated SNP annotation for Abyssinian amyloidosis.

SNP_ID	Chr. Position	Var.	Effect on coding	Aminoacid change	Gene	Odd Ratio	P value	CI 95%	Bonferroni
SNP9	B2:107714979	A	Synonymous	A112A	GPRC6A	2.271	0.082	0.701 6.734	0.227
SNP21	B3:6955760	C	Non-synonymous	R28K	AEN	2.478	0.002	0.731 5.612	0.088
SNP22	B4:138616316	T	Non-synonymous	A696T	EFCAB6	3.293	0.028	0.355 5.653	0.213
SNP26	B4:86030016	C	Non-synonymous	P59R	STAC3	2.912	0.001	0.591 8.522	0.024*
SNP32	C2:13760379	A	Non-synonymous	S1024T	TIAM1	3.712	0.006	0.791 8.613	0.021*
SNP45	E1:9641713	G	Non-synonymous	M1518V	NCOR1	3.014	0.036	0.735 7.274	0.032*

*Significant with $P < 0.05$

The six disease-associated variants were validated by direct Sanger sequencing and genotyped in the Control Cat Population and in the 36 Abyssinians of unknown health status (**Supplementary Tables 9 - 10**). The genotyping rate for the Abyssinians with unknown health status was less successful due to the poor quality of DNA. Nevertheless, allele frequencies were calculated (**Table 3**) and SNPs were tested for HWE in the Abyssinian populations (**Supplementary Table 11**). Considering only the 40 healthy Abyssinians, the genotype frequencies were in HWE. Two SNPs (SNP21 and SNP22) deviated from equilibrium ($p < 0.05$) in the

25 affected cats and the 36 Abyssinians of unknown health status. In the combined 65 Abyssinians, all loci except SNP9 ($p = 0.711$) significantly deviated from HWE.

Table 3. Allele frequencies in the Abyssinian amyloidosis analyses

		Abyssinian Population				Control		
		All	Affected	Healthy	Unknown	All	99 Lives	MALDI-TOF
SNP	$f()$	N = 65	n = 25	n = 40	n = 36	N = 281	n = 112	n = 169
9	A	0.27	0.38	0.21	0.36	0.87	0.98	0.78
21	C	0.36	0.50	0.28	0.40	0.96	0.00	0.94
22	T	0.10	0.18	0.06	0.40	0.23	1.00	0.45
26	C	0.50	0.66	0.40	0.20	0.00	0.00	0.00
32	A	0.19	0.32	0.11	0.15	0.02	0.00	0.03
45	G	0.29	0.44	0.20	0.15	0.01	0.00	0.03

Multivariate Additive Logistic Regression Model

MALR was performed on the Abyssinian Population data to estimate the disease risk attributed to the different genotype combinations at the six disease-associated SNPs. Under the additive model, the allele dosage effects for the rare allele at each of the individual loci were significant at $p < 0.05$, with no correction for multiple testing. In the MALR analysis, all of the signs for the slope coefficients were positive, indicating that disease risk increases with the number of copies of the rare allele at each SNP (**Supplementary Tables 12 - 13**). MALR was also used to estimate the probabilities of disease within the Abyssinian Population, based upon their genotype combinations at the six SNP loci. Of the 25 cases, 20 (80%) had a probability of disease that was greater than 50%. For the other five affected cats, their probabilities of disease ranged between 0.05 and 0.45. Of the 40 healthy Abyssinians, 35 (87.5%) had a probability of disease less than 37%. Ten cats (5 cases, 5 controls) were discordant considering their actual health status versus their estimated probability of disease (**Supplementary Table 14**).

Amyloidosis Risk Prediction

Disease probability was also estimated for all 729 potential genotype combinations at the six loci (data not shown). Only 71 multilocus genotypes were identified in the total sample set, including the Abyssinian Population, Abyssinian Unknown and Control Population, considering only the samples with no missing genotypes ($n =$

253, **Table 4**). Four multilocus genotypes were common in the Control Population, accounting for 166 of 172 cats (97%). Only one of the common multilocus genotypes was found in only one case, accounting for one of the discordant cats. For 25 of the cases, 23 unique multilocus genotypes were identified, one was shared with four healthy Abyssinians. In the healthy cats, 23 multilocus genotypes were unique, five were shared across 13 healthy cats. Four healthy cats had multilocus genotypes that were shared with a cat of unknown health status cat or a cat in the Control Population (unaffected). Eleven unique multilocus genotypes were identified in the cats with unknown health status. Overall, 14 multilocus genotypes were common between two or more cats and 57 were unique (**Table 4**).

Table 4. Multilocus genotypes for cats without any missing genotypes in the study. Genotypes at each locus are indicated with one allele when homozygous.

	Genotypes						Tot	Populations					Notes
	SNP9	SNP21	SNP22	SNP26	SNP32	SNP45		Abyssinian			Control		
	A	C	T	C	A	G	No.	Aff.	Heal.	Unkn.	99L	Plex	
Shared Profiles	A/G	C	C	G	T	A	20	1*	1	1	4	13	*Cat_ID-1
	A	C	C	G	T	A	107				104	3	
	A	C	T/C	G	T	A	28				1	27	
	G	C	C	C/G	T	A	2		1	1			
	G	T	C	C/G	T	G/A	5	1*	4				*Cat_ID-3
	G	T	C	C	T	G	2		2				
	G	T	C	C	T	A	2		1	1			
	G	T	C	G	T	A	3		2	1			
	A/G	C/T	C	G	A/T	G/A	2		1*	1			*Cat_ID-50
	A/G	C	T/C	G	T	A	15			1		14	
	A/G	C/T	T	G	A/T	G/A	2					2	
	A/G	T	T/C	G	T	A	2		2				
	A/G	T	C	G	T	G/A	2	1*		1			*Cat_ID-30
	A/G	T	C	G	T	A	3		3				
Unique Profiles	A	C	C	C/G	T	A	1				1		
	A	C	T/C	C/G	T	G/A	1			1			
	A	C	T/C	G	T	G/A	1					1	
	A	C/T	C	C	T	G	1	1					
	A	C/T	T/C	G	T	A	1		1*				*Cat_ID-59
	A	T	C	C/G	T	G	1	1					
	G	C	C	C	A	A	1		1*				*Cat_ID-51
	G	C	C	C	A/T	G	1	1					

	G	C	C	C	T	A	1		1				
	G	C	T	C	T	G	1	1					
	G	C	T	G	A	A	1	1					
	G	C/T	C	C	T	G	1	1					
	G	C/T	C	C/G	T	G/A	1		1				
	G	C/T	C	G	T	A	1			1			
	G	C/T	C	G	T	G/A	1		1				
	G	C/T	C	C	T	A	1		1				
	G	C/T	T	C/G	A/T	A	1	1					
	G	T	C	C/G	T	A	1		1				
	G	T	C	C	A	A	1	1					
	G	T	T/C	C/G	A	G/A	1	1					
	G	T	T/C	G	T	A	1		1				
	G	C	C	C	A/T	A	1			1			
	G	C	C	C/G	T	A	1	1*					*Cat_ID-18
	G	C	T/C	C	T	G	1	1					
	G	C/T	C	C/G	A/T	G/A	1		1				
	G	C/T	C	G	A/T	G/A	1			1			
	G	C/T	C	C/G	T	A	1		1				
	G	C/T	C	G	T	A	1		1				
	G	C/T	T/C	C/G	T	A	1		1				
	G	T	C	C/G	A/T	A	1		1				
	G	T	C	C/G	T	A	1		1				
	G	T	T	G	A	G/A	1	1					
	A/G	C	C	C	T	A	1		1* 1				*Cat_ID-67
	A/G	C	C	C	T	G	1	1					
	A/G	C	C	C/G	T	G/A	1	1					
	A/G	C	C	G	A/T	A	1					1	
	A/G	C	T	G	A/T	A	1					1	
	A/G	C/T	C	C	A/T	A	1	1					
	A/G	C/T	C	C/G	A/T	A	1	1					
	A/G	C/T	C	C/G	T	G/A	1			1			
	A/G	C/T	C	G	T	G/A	1		1				
	A/G	C/T	C	C	T	A	1			1			
	A/G	T	C	C	A/T	G/A	1	1					
	A/G	T	C	C	T	A	1		1				
	A/G	T	C	C	A/T	A	1	1					
	A/G	T	C	C	T	G/A	1	1					
	A/G	T	C	C/G	A/T	G/A	1		1*				*Cat_ID-57
	A/G	T	C	G	A	A	1	1*					*Cat_ID-24
	A/G	T	C	G	A/T	A	1		1				
	A/G	T	T/C	G	A/T	A	1			1			

	A/G	C	C	G	A/T	A	1			1			
	A/G	C/T	C	C/G	T	A	1			1			
	A/G	C/T	C	G	A/T	A	1		1				
	A/G	T	C	C/G	T	A	1			1			
	A/G	T	C	C/G	T	G/A	1		1				
	A/G	T	C	G	A/T	A	1			1			
	A/G	T	C	G	A/T	G/A	1		1				
						Total	252	24	39	18		172	

*Discordant cats considering actual health status versus probability of disease.

Gene Investigation and Comparative Analysis

The six disease-associated variants mapped to gene coding regions and five were predicted to be non-synonymous substitutions (**Table 2**). Generally, each SNP was located within a highly conserved domain when orthologous sequences were compared across human, dog, cow, and horse (**Supplementary Table 15**). SNP32 had the strongest association with the disease and is located within T-cell lymphoma invasion and metastasis 1 (TIAM1). The transition c.105031T>A causes an amino acid change from serine to threonine (p.S1024T). SNP26 was the second most strongly statistically disease-associated SNP. The transversion c.5625G>C is located within SH3 and Cysteine Rich Domain (STAC3), resulting in a proline to arginine amino acid change (p.P59R). SNP45 is a transition c.113165A>G in Nuclear Receptor Corepressor 1 (NCOR1) causing a methionine to valine (p.M1518V) substitution. SNP21 c.4459C>T is located in Apoptosis Enhancing Nuclease (AEN) and encodes a substitution between the two basic amino acids arginine and lysine (p.R28K), with humans having a lysine at this amino acid position. SNP22 c.88880C>T is located within EF-hand calcium binding domain 6 (EFCAB6) and causes an amino acid change from arginine to tryptophan (p.A696T). However, this SNP is poorly conserved across species as horse and cow have a proline at this position. Human does not have high homology. SNP9 was the least significant of the six amyloidosis associated variants. The nucleotide transition c.17899A>G occurs in G Protein-Coupled Receptor Class C Group 6 Member A (GPCR6A) and creates a synonymous alanine codon (p.A112A).

Discussion and Conclusions

In the 1980s and 1990s, amyloidosis was a highly prevalent and worldwide concern in the Abyssinian cat breed, and its long-haired variant the Somali. Although the disease prevalence has substantially reduced due to selective breeding, lineages positive for the propensity to develop amyloidosis persist. The disease is now re-emerging in some European countries where breeders are not familiar with at risk lines or the disease itself. In Abyssinians, the disease is due to abnormal protein deposition in the kidney interstitium and glomeruli (and occasionally in adrenal gland and thyroid tissue) and is characterized by the presence of proteinuria. It is usually diagnosed due to kidney failure with death at 4-5 years-of-age [Boyce *et al.* 1984]. Amyloidosis has also been documented in the Siamese and related breeds, such as Orientals (and occasionally in Devon Rex, Sphinx and even domestic crossbred cats [Beatty *et al.* 2002]), but the disease presentation and progression is different. In these breeds, the liver represents the target organ for the amyloid deposits with consequent rupture of the hepatic parenchyma resulting in hemorrhage, anemia and occasionally icterus. Death typically occurs at 2-3 years of age but can occur as late as 6 years of age [Van der Linde-Sipman *et al.* 1997]. Both forms of amyloidosis in the different cat breeds are highly characteristic for the respective breeds and are uncommon in the random bred and non-pedigree domestic cat population. Thus, breed predilections are apparent. The same situation is true for the large felids, with cheetahs far more likely to develop renal amyloidosis than other Felidae, consistent with the genetic bottleneck in cheetah phylogeny being akin to the formation of breeds within domestic cats [Menotti-Raymond & O'Brien 1993].

In humans, a total of 36 extracellular fibril proteins are known, nine of which have also been identified in animals. The different forms of amyloidosis are classified based on the biochemical nature of the proteins that compose the fibrils [Sipe *et al.* 2016]. In dogs, amyloidosis has mainly been noticed within the Shar-Pei breed [Di Bartola *et al.* 1990; Segev *et al.* 2012] where it occurs as a reaction to the molecular genetic defect associated with so-called Shar-Pei fever, a disease similar to Mediterranean fever of man and is characterized as a familial form with death occurring at approximately 4 years of age from hepatic rupture.

Amyloid deposits can occur as a secondary manifestation of chronic inflammation. Abyssinian cats are documented to have a higher prevalence of gingivitis and stomatitis [Pedersen N.C. 1992]. Chronic inflammation can trigger amyloid deposition and compromise the integrity of the kidneys. Generally, all cats defined as cases for this project were under the age of 5 years at death or diagnosis and had confirmatory necropsies of primary amyloidosis or a histological report from a kidney biopsy that was diagnostic of amyloidosis. However, misclassification of the cats is feasible due to variation in the disease presentations. The Abyssinian control population was composed by aged individuals and almost all of them older than 5 years of age, with sure diagnosis excluding amyloidosis or good information on their health. Anyway, four Abyssinians included in this group were 4 years old or less but were considered as controls basing on the recent information that we had from the breeder. Nevertheless, one of these cats resulted to have a discordant prediction versus the observed phenotype (Cat_ID 67) and this could be addressed to the possibility of a subsequent onset of the disease.

This study represents the first large-scale investigation of the genetic architecture of amyloidosis in domestic animals, specifically the Abyssinian cat breed. The initial analysis conducted was a case–control association study analyzing 21 affected Abyssinians and 23 controls using a private 297K DNA array. Because this array was developed with a focus on a particular non-Abyssinian population of cats, only ~50% of the variants were informative in the Abyssinian cohort. Several SNPs, including one on cat chromosome B4, approached significance, although no variants were significant after correction for multiple testing. The region identified on chromosome B4, however, was also suggested to have disease-associated SNPs from the WGS.

WGS was completed on two affected Abyssinian cats included in the GWAS and variants discovered in these animals were compared to those identified in 113 unaffected cats sequenced in the 99 Lives Cat Genome Sequencing Project. The reference Abyssinian is derived from a line with progressive retinal atrophy (PRA) but had no evidence of renal amyloidosis [Narfström K. 1983]. In addition, the reference cat was re-sequenced, thus heterozygous allele variants present in the reference Abyssinian were identified and are included in the 99 Lives variant dataset. Because filtering of variants allowed the inclusion of variants at very low frequency, potential disease-associated SNPs would have been included from the

reference sequence. From the variant analysis of the two amyloidosis genomes, 51 candidate coding variants were identified that were either heterozygous or homozygous in the two affected Abyssinian cats and not present in any other cats, including the reference Abyssinian. The genotypes of these variants were generated for a larger cohort of affected and unaffected Abyssinians and statistical analysis suggested that six variants were associated with disease.

The six variants had odds ratios for amyloidosis ranging from 2.71 - 3.71, individually. SNP26, SNP32, and SNP45 were the most highly significantly associated SNPs. The significant association of unlinked variants supports a multifactorial genetic influence for renal amyloidosis in Abyssinian cats. For all six loci, the allele that was rare in the controls increased the risk of disease, which would be expected if this allele was exposed to negative selection. In combination, the six identified variants can be used to develop risk profiles for amyloidosis in Abyssinian cats, with a higher risk of disease being related to the presence of increased numbers of associated alleles, particularly in the homozygous state. The mean disease risk was estimated to be 0.21 in the sampled Abyssinian population. Cats affected with amyloidosis had individual disease risks ranging between 0.99 and 0.51. Healthy cats had risks in the range from 0.005 to 0.372. However, a few cats were discordant between risk profile and disease status. This could be addressed to different factors. Particularly, the Abyssinian control population comprised aged individuals with almost all cats being older than 5 years, with accurate diagnoses or detailed information concerning their health, although four cats were 4 years of age or less. Five cats had estimated disease probabilities between 0.57 and 0.84, including Cat ID 67, which is one of those younger cats, hence this individual might develop the disease in the future. Conversely, five affected cats with necropsy report, were all < 5-year-old and positive for Congo red staining but had low estimated disease probabilities from 0.048 to 0.045. The cat with the lowest estimated disease probability was Cat ID 3, which was a 3-year-old cat with a necropsy report and was one of the two used for WGS. Therefore, other variants with effects too small to reach statistical significance in our association analyses are likely to influence disease development and progression and may not be shared in all affected Abyssinians. The problem of having 10 cats with a discordancy between the phenotype and the disease prediction could be addressed to the population sampling that is not random. The genotype frequencies and the overall disease risk

are biased by the sample ascertainment. This problem could be overcome by genotyping a much larger random sample of cats, in which disease prevalence is representative of the overall population. Certainly, a more extended analysis of the Abyssinians would support a better definition of the risk categories for the six-disease associated SNPs, but the insufficient availability of samples and the difficulty in their collection still represent a big issue to further develop this study.

The six SNPs suggest that 729 different multilocus genotypes are possible in cats, however, only 71 different genotypes were identified in the studied cats, including 57 unique profiles and 14 that were shared among cats. Interestingly, every affected cat had a different DNA profile. Four DNA profiles were shared for 170 cats, of which only one cat was affected with amyloidosis. Thus, the identified risk alleles and profiles are rare in the general cat population, as indeed is amyloidosis. Five DNA profiles were identified in more than one healthy Abyssinian, accounting for 13 of 40 cats. Overall, the Abyssinians had rare DNA profiles that were generally unique.

Of the six disease-associated variants, five are predicted *in vitro* to be associated with amino acid substitutions with moderate effects, and one is a synonymous amino acid alteration, an alanine in the GPRC6A. This SNP showed the least significant association with disease risk. SNP22 is located on cat chromosomes B4, supporting one of the regions suggested by the GWAS. SNP22 is located within EFCAB6 that is involved in Alzheimer Disease (AD) pathways [Feng *et al.* 2015] and plays an important role in regulating β -amyloid levels through the androgen receptor in humans [Drummond *et al.* 2009]. SNP32 is located within TIAM1, which has no clear association with amyloidosis at this time. However, there is evidence that TIAM1 encodes a guanine nucleotide exchange factor which activates the Rho family member Rac1. This in turn activates amyloid precursor protein (APP) transcription [Wang *et al.* 2009] and may contribute to the formation of toxic amyloid fragments in human AD [Boo *et al.* 2008]. SNP26 creates a purine to pyrimidine transversion in STAC3 which is involved in amyloid fibril formation in cattle [Guijarro *et al.* 1998]. SNP45 is located in NCOR1 which participates in both the ubiquitin-proteasome system and the retinoic acid pathway. This system and the associated pathways contribute to the relationship between β -amyloid accumulation in the brain and AD [Goodman & Pardee 2003]. SNP21 is in AEN with no evidence of direct links to amyloidosis, however, AEN is linked to apoptosis. No variants were identified in

genes demonstrated as amyloidogenic in other species, such as SAA, Ig Light chain (IGL) or TTR.

Overall, all the six variants map to genes that do not code for amyloid proteins. This evidence is in concordance with the definition of the term “familial amyloidosis” given by the nomenclature committee of the international society of amyloidosis.

The GWAS and WGS supported the identification of DNA variants that confer risk for amyloidosis in Abyssinian cats. The identified disease-associated DNA profiles are mainly unique to Abyssinians and rare within non-Abyssinian cat populations, where amyloidosis is also very rare. Cats with a disease estimated probabilities of 0.50 – 0.99 should be considered as having a high risk for disease while cats with DNA profiles suggesting a risk in the range from 0.005 to 0.372 have a lower risk for disease. It is also important to consider that other factors and genes may contribute to the onset of amyloidosis in the Abyssinian breed.

Supplementary Tables

Supplementary Table 1. Abyssinian samples used for amyloidosis genetic analyses.

Laboratory ID	Cat No.	Specimen analyzed	Sex	Age at death/at analysis	Sample origin	Sample used for diagnosis	Amyloidosis	GWAS	Geno typing
							Diagnosis		
Fcat 819	1	White Blood Cells	M	3	USA	tissue	Renal		X
Fcat 16343	2	Buccal swab	M	5	USA	whole body	Renal	X	X
Fcat 16515*	3	Muscle	F	6	USA	tissue	Renal	X	X
Fcat 18244	4	Buccal swab	F	7	USA	whole body	Renal		X
Fcat 18252	5	Buccal swab	F	5	USA	whole body	Renal	X	X
Fcat 18260	6	Buccal swab	M	3	USA	whole body	Systemic		X
Fcat 19340	7	FFPE	F	5	ITALY	whole body	Renal	X	X
MIRec 1246	8	Blood & Muscle	F	5	ITALY	whole body	Systemic		X
Fcat 18781	9	Blood & FFPE	F	4.5	ITALY	whole body	Systemic	X	X
Fcat 18779	10	Blood	F	4	ITALY	whole body	Systemic	X	X
Fcat 19338	11	Blood & muscles	M	4.5	ITALY	whole body	Systemic		X
Fcat 18778*	12	Blood & liver	F	4	ITALY	whole body	Systemic	X	X
MIRec 2865	13	Blood	M	2.5	ITALY	whole body	Systemic		X
Fcat 18780	14	Blood & FFPE	M	3	ITALY	whole body	Systemic	X	X
MI PN 233/02	15	FFPE	F	5	ITALY	whole body	Renal		X
Fcat 18631	16	tissue & FFPE	F	2.5	FRANCE	Tissue	Renal	X	X
Fcat 18629	17	Liver & FFPE	F	3.5	FRANCE	Tissue	Renal	X	X
Fcat 18782	18	Blood	M	5	FRANCE	Tissue	Renal	X	X
Fcat 18628	19	Blood & FFPE	F	5.5	FRANCE	Tissue	Renal		X
Fcat 18783	20	Blood	M	4	FRANCE	Tissue	Renal	X	X
Fcat 18784	21	Blood	F	9	FRANCE	Tissue	Renal	X	X
Fcat 20044	22	Kidney	F	2	USA	Tissue	Renal	X	X
Fcat 20043	23	Skin	F	2	USA	Tissue	Renal	X	X
Fcat 18619	24	Blood	M	4.5	FRANCE	Tissue	Renal	X	X
Fcat 18620	25	Kidney	F	3.5	USA	Tissue	Renal	X	
Fcat 18627	26	Liver & FFPE	F	5	FRANCE	Tissue	Renal	X	
Fcat 18632	27	Kidney	M	1.5	FRANCE	Tissue	Renal	X	
Fcat 20040	28	Blood	F	6	USA	Tissue	Renal	X	
Fcat 20045	29	Ear	F	6	USA	Tissue	Renal	X	
Fcat 19903	30	Muscle	M	4	USA	Tissue	Renal		X

Fcat 567	31	White Blood Cells	M	>15	USA	n.a.	No health issues	X	X
Fcat 18240	32	Buccal swab	M	7	USA	n.a.	alive healthy		X
Fcat 18241	33	Buccal swab	M	10	USA	Tissue	Hystology excluded		X
Fcat 18242	34	Buccal swab	F	8	USA	n.a.	alive and healthy		X
Fcat 18243	35	Buccal swab	F	8	USA	n.a.	alive healthy		X
Fcat 18245	36	Buccal swab	F	9	USA	n.a.	alive healthy		X
Fcat 18247	37	Buccal swab	M	7	USA	n.a.	alive healthy		X
Fcat 18248	38	Buccal swab	M	9	USA	Tissue	Hystology excluded		X
Fcat 18249	39	Buccal swab	?	7	USA	n.a.	alive healthy		X
Fcat 18253	40	Buccal swab	?	6	USA	n.a.	alive healthy		X
Fcat 18265	41	Buccal swab	F	7	USA	n.a.	alive healthy		X
Fcat 6935	42	Buccal swab	F	10	USA	n.a.	alive healthy		X
Fcat 18785	43	Blood & FFPE	F	12	ITALY	whole body	Hystology excluded	X	X
Fcat 18786	44	Blood	F	14	CH	n.a.	alive healthy	X	X
Fcat 18790	45	Blood	F	8	ITALY	n.a.	alive healthy	X	X
Fcat 18788	46	Blood	M	8	ITALY	n.a.	alive healthy		X
Fcat 18792	47	Blood	F	5	ITALY	whole body	Hystology excluded	X	X
Fcat 18789	48	Blood	F	8	ITALY	n.a.	alive healthy	X	X
Fcat 18791	49	Blood	F	6	ITALY	whole body	Hystology excluded	X	X
Fcat 18787	50	Blood	F	7	ITALY	n.a.	alive healthy	X	X
Fcat 19335	51	Blood	M	6	ITALY	n.a.	alive healthy	X	X
Fcat 19337	52	Blood	F	5	ITALY	n.a.	alive healthy	X	X
Fcat 19336	53	Blood	F	5	ITALY	n.a.	alive healthy	X	X
MIRec 1460	54	Blood	M	8	ITALY	n.a.	alive healthy		X
MIRec 2417	55	Blood	F	7	ITALY	n.a.	alive healthy		X
MIRec 2416	56	Blood	F	6	ITALY	n.a.	alive healthy		X
MIRec 2748	57	Blood	F	6	ITALY	n.a.	alive healthy		X
MIRec 2625	58	Blood	F	6	ITALY	n.a.	alive healthy		X
Fcat 19334	59	Blood	M	8	ITALY	n.a.	alive healthy	X	X
Fcat 18626	60	Liver& Spleen&FFPE	F	8	USA	whole body	Hystology excluded		X
Fcat 18793	61	Blood	?	6	ITALY	n.a.	alive healthy		X
Fcat 20041	62	Blood	M	11	ITALY	n.a.	alive healthy	X	X
Fcat 20054	63	Blood	F	12	FRANCE	Tissue	Hystology excluded	X	X
Fcat 20047	64	Liver	M	7	FRANCE	Tissue	Hystology excluded		X
Fcat 21976	65	Blood	F	4	FRANCE	n.a.	alive healthy		X

Fcat 21977	66	Blood	M	5	FRANCE	n.a.	alive healthy		X
Fcat 21978	67	Blood	F	4	FRANCE	n.a.	alive healthy		X
Fcat 21979	68	Blood	F	4	FRANCE	n.a.	alive healthy		X
Fcat 18617	69	Ear	F	8,5	USA	whole body	Hystology excluded	X	
Fcat 18625	70	Tongue & FFPE	F	15	USA	whole body	Hystology excluded	X	
Fcat 18630	71	Tongue	M	7	USA	whole body	Hystology excluded	X	
Name unknown	72	Blood	?	12	USA	n.a.	No health issues	X	
Fcat 20031	73	Buccal swab	F	19	USA	tissue	Hystology excluded	X	
UM_294	74	Blood	F	13	ITALY	whole body	Hystology excluded	X	
UM_293	75	Blood	F	13	ITALY	whole body	Hystology excluded	X	
Fcat 20053	76	Blood	M	11	FRANCE	tissue	Hystology excluded	X	
Fcat 20032	77	Buccal swab	F	15	CANADA	n.a.	No health issues	X	
Fcat 19904	78	Muscle	F	6	USA	whole body	Hystology excluded		X
Mi PN 1/08B	79	FFPE	M	1	ITALY	whole body	Hystology excluded		X

*Two affected cats used for whole genome sequencing

Supplementary Table 2. Primers used to validate the 51 WGS detected variants by MALDI-TOF.

PLEX	SNP	Chr	Position	Primary Amplification Primer (1st-PCR)	Secondary Amplification Primer 2nd-PCR	Extend Primer (UEP_SEQ)
PLEX 1	1	A1	195910142	ACGTTGGATGTGCTCACCAACCTGTGTTTC	ACGTTGGATGTGATCCTGCAGGTACTTGAC	GCAGGTACTTGACCTTCTGG
PLEX 1	2	A2	156629654	ACGTTGGATGACCCGTGCTTCTCTATCTC	ACGTTGGATGTTGACGCTGGGTTGATAAAC	GGGTTGATAAACTCCCAGT
PLEX 2	3	A2	162350489	ACGTTGGATGGGGAGCTTCAGCTTCTTCAG	ACGTTGGATGTGATGCTCTGGAGTCCTTG	GGCACCAGACCAGAGGAAGGT
PLEX 1	4	A2	4633642	ACGTTGGATGTAGGTGGCAGAGCTCCCAAG	ACGTTGGATGTGCTTCGAAGGCCTCGTC	TGCCTCGTCGTCGGACCCC
PLEX 1	5	A2	8928090	ACGTTGGATGAGCCAGTAGAGCATGTAAG	ACGTTGGATGTAGGAAGCATTGAACTGGG	TCAGGTCCCAACTGGCT
PLEX 1	6	A3	116208492	ACGTTGGATGCTTCTGTAGAGCAACCATC	ACGTTGGATGCTTGGGAGAGTGAAGTGAAG	GAGGGTGAATCTGCAAAACCAAC
PLEX 1	7	A3	116213029	ACGTTGGATGAGAGACTGCAAGCTACCTG	ACGTTGGATGAGTAGGTGGACTGTCATCGG	TGACAGTTGGGGAAGCTTCG
PLEX 1	8	B1	1225749	ACGTTGGATGATGGATGCCCTGGACCTTC	ACGTTGGATGGGAATACATTCTGCTCCTG	TACATTCTGCTCTGGGAGTAACCA
PLEX 2	9	B2	107714979	ACGTTGGATGTGTACAGAAGTCAGTGGC	ACGTTGGATGCTCCATGATTCTCTGGAGC	CTTAGAAAGAACTCAGAGC
PLEX 2	10	B2	125289688	ACGTTGGATGGTTGATATTACGCCTCAGAG	ACGTTGGATGTCCAAGTCTTCTTGATCCGC	CCTCGGTATCACTCTCAGATG
PLEX 2	11	B2	138345851	ACGTTGGATGGTCTTCTATGCTGGCACAC	ACGTTGGATGTAAGTACTCTCCAGTCACC	GCCAGTCACCACAGAGGAAG
PLEX 2	12	B2	146019210	ACGTTGGATGACAAACAGCACTTGCTTGCC	ACGTTGGATGCATGGGAAGCTGTTGATACC	ACAAAATAACATGCTTTATAAGATCA
PLEX 2	13	B2	146200670	ACGTTGGATGTGAGTCTGCACTCACTCAC	ACGTTGGATGCAACAGGAGCCACCTGCCG	TCCTCTGGGCCCTGTCTGGCAC
PLEX 2	14	B2	42064382	ACGTTGGATGTCAAGCTCTTCTGTCTCTG	ACGTTGGATGTGAGTCTGCAACTGGACAGC	GGGCACTGGACAGCAGTGACA
PLEX 2	15	B2	42129510	ACGTTGGATGCCTGTCACTGTGAGGAAATC	ACGTTGGATGAAACGCACATGTGGCTTCTC	CCGTGAGGGGTGAGAGCAC
PLEX 2	16	B2	42224102	ACGTTGGATGTGGACTCGGAATCAAAACCAG	ACGTTGGATGCTGGCTTTGGAGTCAACATC	GGAGTCAACATCACAAAT
PLEX 2	17	B2	46682907	ACGTTGGATGCGTGACAGGCTTGATGAAG	ACGTTGGATGGGATCCTTGTCATCAAGGTC	GAGAGAGGTCAGAAAAATGTTTCTTC
PLEX 2	18	B2	82912165	ACGTTGGATGGAATTTGATGCCGACTCTGG	ACGTTGGATGAAGCCTGTATCGATTGACC	TATCGATTTGACCTCAGAAC
PLEX 2	19	B2	88829020	ACGTTGGATGGTTACTTTAACTGGAGGAAG	ACGTTGGATGCACATGATCACAAGCCAAGC	TCGTGGAAGGTTTACTG
PLEX 1	20	B3	112107370	ACGTTGGATGACACAGAGTCCATCTGTTTG	ACGTTGGATGCACCTGCTAACTTTTCTCTCC	CCTCCAAAAGGCTGCTTTTC
PLEX 1	21	B3	6955760	ACGTTGGATGCCCGAATGCCAAGGATGTG	ACGTTGGATGCATGAATCGCTGGTGCTGTC	TCCTCGCTTGTGCTTT
PLEX 2	22	B4	138616316	ACGTTGGATGAAGAACCAGGAGTCTGCGATG	ACGTTGGATGACGTGGATGCGCCACTAAC	CCTCAGGTTTGTGATTGTTTCAGAC
PLEX 2	23	B4	138662322	ACGTTGGATGATCCATGAAACATCATAGG	ACGTTGGATGGACTCTAGCGGTGGTTTTA	GGCGATAGCGGTGGTTTTAAGTCCA
PLEX 2	24	B4	138902276	ACGTTGGATGATAAAGTCACCGGCCAGTTTC	ACGTTGGATGAGGTCACTACGCTGTACTC	TGAAACCCCTCTGCTCGCTCCTCC
PLEX 2	25	B4	85091126	ACGTTGGATGTCAGACGCTCCAGACGCAAG	ACGTTGGATGTTAACCCTGCCTACTTCTCTC	ACCTCTGACTCTGGAGGAGCTGCCAT
PLEX 2	26	B4	86030016	ACGTTGGATGAAGGAGATGGAGCTTCCCC	ACGTTGGATGCCTCTTCTCATAGATGTAG	CGTCCTCATAGATGTAGTAGATG
PLEX 2	27	B4	91914831	ACGTTGGATGTGCCATTGTGTCTTCTTCC	ACGTTGGATGAACTGTCTTCAAGAAGGGC	CCCTGCTGCTGTATCAC
PLEX 1	28	C1	182160304	ACGTTGGATGTCCATGTTTGCATCTTCTC	ACGTTGGATGTCCAAAAGACTTTATACCCC	ACCAGATCTTTAATGCCAGG
PLEX 1	29	C1	210586285	ACGTTGGATGTACAGGCAAAATTGTGTGTC	ACGTTGGATGAGGAGACTTTTAAAGGACAG	GACTTTTAAAGGACAGTCAGAAAAC
PLEX 2	30	C2	109557239	ACGTTGGATGGCTCCACATTCTGACTGTTG	ACGTTGGATGGGGAAGCAGCAAAATCATT	TCCAAGAACCTGGCCT
PLEX 2	31	C2	124223482	ACGTTGGATGAAAGAAAAAGAGCGGGCCAC	ACGTTGGATGAGAGTTGAGTTATGTTGGC	CGAGGTATGTTGGCAACACTTTT
PLEX 2	32	C2	13760379	ACGTTGGATGAGCTTGTGAGCTGCGTGAG	ACGTTGGATGTGAGATGAACCTCTCTGACC	TCCGACCCGAGCCCATCGCCTCAGGAC
PLEX 2	33	C2	141036180	ACGTTGGATGGAGACTGTAACCTCTGACTC	ACGTTGGATGCATAATCCTCACCTTTTGGG	GTAAGTCTTCTTGGAAATGC
PLEX 2	34	C2	59362987	ACGTTGGATGCACCATCCTTTCCAGAAGAG	ACGTTGGATGCTGAAATGGAACAGAGACGG	TAGAGACGGTCTACAGGC
PLEX 2	35	C2	82574100	ACGTTGGATGTGAGTCCAGCTCATTTTC	ACGTTGGATGAACCAGACAGAGAACGATGC	CGAAGACTTGTCCAGG
PLEX 2	36	C2	84454399	ACGTTGGATGAGAGAGACACAAGCGGAAC	ACGTTGGATGTAGTCTTCCAGAAACCCAC	TCCTTACCCGACCTGCTGAA
PLEX 2	37	D2	33354542	ACGTTGGATGTCTCTAGAGTCTGGGTTAGC	ACGTTGGATGCAGGCAAGAAATGGTGGAC	AATGGGGACTACAATGGGAACATC

PLEX 2	38	D2	63873265	ACGTTGGATGTGTGCTGCCTGTTGACCATC	ACGTTGGATGAGTGTCCAAGAGACGCAAGG	GGCCCGAGACGCAAGGCCCATC
PLEX 1	39	D3	45637034	ACGTTGGATGGGTACAGATGCTGCTAATGG	ACGTTGGATGATGAGTAATTCCTGGGCGG	TAGGCACTGCTCTTCCCTCCA
PLEX 1	40	D4	23980293	ACGTTGGATGATGTTGGTGGGAAACAAGGC	ACGTTGGATGTCTCCAAAGTATCCTGGGAC	GTCCCTCTGTCTCGACAGCGTCA
PLEX 1	41	D4	24908573	ACGTTGGATGAACTGTCACATGCTGTGGTC	ACGTTGGATGAACATGACCCAGGATGTCAG	GTCCAGAGGGAGCCACGGCGTCGGCA
PLEX 2	42	E1	15509974	ACGTTGGATGGAAATAACGGACAGTCCCTG	ACGTTGGATGTGAACACTACCCAGTGTGC	AGTGTGCGGAGAAATAGA
PLEX 2	43	E1	33512129	ACGTTGGATGCCTGCCITTCAGATGATCAC	ACGTTGGATGCCTTCGTTCCGGTTAATTCC	CCCCAGGATCTTCAGACCAAGGCC
PLEX 2	44	E1	34398695	ACGTTGGATGTGATGATGTCCAGGTCCTTC	ACGTTGGATGTGGAGCCCAGGAAGACCAT	AGGAAGACCATCGGATCTGC
PLEX 2	45	E1	9641713	ACGTTGGATGCTGTGAGTTATCAGAGCACC	ACGTTGGATGAGGACAGAAGAGTAACGCAC	TCCGAAGTTCTGTTTCATCA
PLEX 2	46	E2	6261917	ACGTTGGATGTGAGCTTGAACCTTCTGGAG	ACGTTGGATGCGGCTCAGAACATAAGAG	CCAGGGGTGTAGTGTATGTGGGA
PLEX 2	47	E2	6262273	ACGTTGGATGCCTGTATGAGTTTCTGATG	ACGTTGGATGATGAAAGAGGAGAAAGGCC	TCCACGGTTTTGCCAGAAATCA
PLEX 1	48	E3	26646466	ACGTTGGATGTCCCGGGTCAAGAAAAGGAG	ACGTTGGATGGAGGTTAGGGAAAATGGTGG	GGTGTTTTAGAATGCTTGGACT
PLEX 1	49	F1	40368085	ACGTTGGATGTCTGTGAGTCGTGTGGTCA	ACGTTGGATGATGTTGGAGGAAGGCTAAGG	GGGAGACCGTGAAAGGGCTTAC
PLEX 1	50	F1	67173383	ACGTTGGATGACAAACGGAGCCACGGATAG	ACGTTGGATGAAAAGAGAGGCAGACCGAAG	GAGGCAGATGAGGTCAC
PLEX 1	51	F2	56350219	ACGTTGGATGCACAAGACTTAGTTCACAATG	ACGTTGGATGCCATTACAATGAGCTCTGG	CAATGAGCTCTGGAATCCGCAGT

Supplementary Table 3. PCR primers used to validate the six variants associated with amyloidosis by Sanger sequencing.

SNP	SNP Position	Ref.	Alt	Primer Forward (5' – 3')	Primer Reverse (5' – 3')	Amplicon (bp)
SNP 9	B2:107714980	A	G	CTCTTGCCATGATTCACAGCA	CATGCCAAGGGTTAAGGCTG	209
SNP 21	B3:6955760	C	T	CCTCTGCCCCTTGTGTTCTT	GATGCAGCGTCCAAGGAATG	350
SNP 22	B4:138616316	C	T	TACCGCTGTTGAGTATCCCG	GCTTGGCTTATCTGGACACC	286
SNP26	B4:86030016	G	C	GTTTCCTCCCGTCCCCAC	CGGATGATTGTGCGTGAGTC	310
SNP 32	C2:13760379	T	A	AGTATGAACCCACAGAGCCC	TCTGTTCCGCCTCAAGATGT	347
SNP 45	E1:9641713	A	G	AGCTTTTCACCTGGTAGCCT	TGACCTAAGAAGTGAAGGGGC	306

Supplementary Table 4. Table showing the related Abyssinians ($\hat{p} > 0.3$) among the GWAS population.

Cat_ID 1	Cat_ID 2	$\hat{p}$	Cat_ID 1	Cat_ID 2	$\hat{p}$
27	10	0.624	26	9	0.508
48	49	0.586	63	69	0.500
44	9	0.581	48	47	0.484
48	45	0.563	62	9	0.464
27	52	0.546	62	26	0.452
45	47	0.543	26	52	0.397
44	62	0.540	10	52	0.341
44	52	0.533	45	7	0.310
45	49	0.529	49	7	0.308
44	26	0.527	52	9	0.308
12	14	0.516	53	14	0.304
49	47	0.513			

Cat_ID 1 = individual ID for the first individual

Cat_ID 2 = individual ID for the second individual

Supplementary Table 5. Ten most significant GWAS SNPs with $-\log_{10}(p) \geq 4$.

CHR	Position (bp)	P	FDR	log(P)
F2	65781832	2,03E-05	0,485	4,692
F1	42828353	2,80E-05	0,485	4,552
F2	65831486	3,11E-05	0,485	4,507
E1	36072017	4,92E-05	0,485	4,308
F2	66073392	6,32E-05	0,485	4,2
D4	94979928	6,64E-05	0,485	4,178
B4	138605559	7,29E-05	0,485	4,137
F2	65928083	7,68E-05	0,485	4,114
C1	127106996	7,88E-05	0,485	4,103
F2	65869984	8,31E-05	0,485	4,08

Supplementary Table 6. WGS candidate SNP locations in the *Felis catus* reference genome V6.2, impact and gene names.

SNP	Chr.	Position	Ref	Var	Amino acid Change	Gene
SNP1	A1	195910142	G	A	R393H	TNIP1
SNP2	A2	156629654	C	T	R70Q	ENSFCAG00000024016
SNP3	A2	162350489	C	T	A61T	ZNF467
SNP4	A2	4633642	T	C	K263R	GTF2F1
SNP5	A2	8928090	C	T	E370K	DNASE2
SNP6	A3	116208492	G	A	P1918L	BIRC6
SNP7	A3	116213029	T	C	T1613A	BIRC6
SNP8	B1	1225749	C	T	A422V	DLGAP2
SNP9	B2	107714980	A	G	A112A	GPRC6A
SNP10	B2	125289688	C	T	A858T	ENSFCAG00000008186
SNP11	B2	138345851	C	T	A912V	AKAP12
SNP12	B2	146019210	G	A	S20N	IGF2R
SNP13	B2	146200670	C	T	G124S	SLC22A2
SNP14	B2	42064382	G	A	G28D	NFYA
SNP15	B2	42129510	A	C	L1059F	ENSFCAG00000022408
SNP16	B2	42224102	C	T	V196M	TREM1
SNP17	B2	46682907	C	T	T98M	ENPP4
SNP18	B2	82912165	C	A	P1788H	ENSFCAG00000027150
SNP19	B2	88829020	A	G	T324A	FUT9
SNP20	B3	112107370	C	T	R186Q	ENSFCAG00000021991
SNP21	B3	6955760	C	T	R28K	AEN
SNP22	B4	138616316	C	T	A696T	EFCAB6
SNP23	B4	138662322	A	C	F345C	EFCAB6
SNP24	B4	138902276	C	T	R205W	SAMM50
SNP25	B4	85091126	C	T	S758F	ESYT1
SNP26	B4	86030016	G	C	P59R	STAC3
SNP27	B4	91914831	C	T	R754C	SRGAP1
SNP28	C1	182160304	T	C	T2134A	DNAH7
SNP29	C1	210586285	C	T	V1497I	SPHKAP
SNP30	C2	109557239	G	A	S256L	ARHGEF26
SNP31	C2	124223482	T	G	F278L	DZIP1L
SNP32	C2	13760379	T	A	S1024T	TIAM1
SNP33	C2	141036180	C	T	G1446S	TOP2B
SNP34	C2	59362987	C	T	A1212T	WDR52
SNP35	C2	82574100	C	A	A76E	SST
SNP36	C2	84454399	C	A	Q124H	MAP3K13
SNP37	D2	33354542	A	G	N347D	COL13A1
SNP38	D2	63873265	A	G	L9P	ENSFCAG00000002390

SNP39	D3	45637034	A	G	H285R	GREB1L
SNP40	D4	23980293	C	T	R656H	RASEF
SNP41	D4	24908573	T	C	N465S	KIF27
SNP42	E1	15509974	C	T	V86I	NXN
SNP43	E1	33512129	G	A	T29M	STXBP4
SNP44	E1	34398695	C	T	V198I	KIF2B
SNP45	E1	9641713	A	G	M1518V	NCOR1
SNP46	E2	6261917	G	A	R202K	ZNF649
SNP47	E2	6262273	C	G	Q321E	ZNF649
SNP48	E3	26646466	A	G	L429S	ENSFCAG00000015160
SNP49	F1	40368085	G	T	G926C	ENSFCAG00000023584
SNP50	F1	67173383	C	G	P899R	ASH1L
SNP51	F2	56350219	G	C	H2163D	CSMD3

Supplementary Table 7. Genotyping of the 51 SNPs identified by WGS as candidates for influencing risk of amyloidosis.

[illegible]

Supplementary Table 8. Allele frequencies and Odds Ratios for 51 WGS variants in Abyssinian cats (PLINK vs 1.07).

SNP	A1 (minor allele)	A2 (major allele)	MAF*	OR	BONF**
SNP1	A	G	0.465	0.467	1
SNP4	C	T	0.233	1.071	1
SNP5	T	C	0.171	1.005	1
SNP2	T	C	0.418	0.629	1
SNP3	T	C	0.278	1.000	1
SNP6	A	G	0.208	0.840	1
SNP7	C	T	0.214	0.740	1
SNP8	T	C	0.417	1.771	1
SNP14	G	A	0.182	0.556	1
SNP15	A	C	0.183	0.333	1
SNP16	C	T	0.197	0.260	1
SNP17	C	T	0.137	1.162	1
SNP18	A	C	0.329	0.400	1
SNP19	G	A	0.389	0.829	1
SNP9	A	G	0.286	3.306	0.917
SNP10	T	C	0.238	1.506	1
SNP11	T	C	0.442	0.739	1
SNP12	G	A	0.427	0.898	1
SNP13	T	C	0.213	0.615	1
SNP21	C	T	0.415	3.800	0.109
SNP20	C	T	0.336	1.160	1
SNP22	T	C	0.078	4.231	0.619
SNP25	T	C	0.349	2.429	1
SNP26	C	G	0.485	4.598	0.023**
SNP27	T	C	0.421	2.895	1
SNP24	T	C	0.436	0.906	1
SNP28	T	C	0.500	1.012	1
SNP29	T	C	0.253	0.863	1
SNP32	A	T	0.227	4.355	0.021**
SNP34	T	C	0.247	1.526	1
SNP35	C	A	0.256	0.897	1
SNP36	C	A	0.294	0.923	1
SNP30	G	A	0.367	1.403	1
SNP31	T	G	0.421	2.200	1
SNP37	G	A	0.169	1.293	1
SNP38	G	A	0.131	1.842	1
SNP39	G	A	0.236	1.150	1

SNP40	T	C	0.410	0.733	1
SNP41	C	T	0.333	0.761	1
SNP45	G	A	0.316	3.014	0.032**
SNP42	C	T	0.452	1.997	1
SNP47	G	C	0.270	0.425	1
SNP48	G	A	0.360	1.431	1
SNP49	T	G	0.181	1.875	1
SNP50	C	G	0.452	0.774	1
SNP51	G	C	0.472	1.364	1

*SNP24, SNP33, SNP43, SNP44 and SNP45 failed the frequency test (MAF < 0.05)

** Significant SNPs after Bonferroni correction (BONF).

Supplementary Table 9. Control population of cats for the analysis of Abyssinian amyloidosis, excluding the reference genome of Cinnamon.

N.	Breed	99 Lives	MALDi-TOF	N.	Breed	99 Lives	MALDi-TOF
1	American Shorthair	1	8	19	Norwegian		9
2	American Wirehair	1		20	Ocicat	1	7
3	Bengal	8	6	21	Oriental	8	1
4	Birman	8	7	22	Persian	3	7
5	British Shorthair	2	6	23	Peterbald	1	
6	Burmese	5	7	24	Ragdoll	2	7
7	Chartreux		7	25	Russian blue		5
8	Cornish Rex		8	26	Savannah	1	
9	Devon rex	2	7	27	Selkirk rex	1	
10	Egyptian Mau	3	6	28	Siamese	6	
11	Hymalaian	1		29	Siberian	1	6
12	Japanese bobtail		7	30	Sphynx	2	7
13	Korat		3	31	Tennessee rex	1	
14	LaPerm	1		32	Tonkinese	1	7
15	Likoi	1		33	Toyger	1	
16	Maine Coon	8	9	34	Turkish Angora		5
17	Manx		6	35	Turkish Van		6
18	Napoleon	3		36	Random bred	39	15
					Total	112	169

Supplementary Table 10. Genotypes at the six disease-associated SNPs in the Control population and in the Unknown Abyssinians.

Breed	Cat ID	SNP						Source
		9	21	22	26	32	45	
Abyssinian	Fcat 20681	GA	C	C	G	T	A	99Lives
American Shorthair	S20431	A	C	C	G	T	A	99Lives
American Shorthair	Fcat 14899	A	C	TC	G	T	A	MALDI-TOF
American Shorthair	Fcat 16672	A	C		G	T		MALDI-TOF
American Shorthair	Fcat 3123	GA	C		G	T	A	MALDI-TOF
American Shorthair	Fcat 4371	GA	C	TC	G	T	A	MALDI-TOF
American Shorthair	Fcat 4399	A	C	TC	G	T	A	MALDI-TOF
American Shorthair	Fcat 4437	GA	C		G	T	A	MALDI-TOF
American Shorthair	Fcat 4948	GA	C		G	T	A	MALDI-TOF
American Shorthair	Fcat 6416	GA	T	T	G			MALDI-TOF
American Wirehair	Fcat 20784	A	C	C	G	T	A	99Lives
Bengal	K452	A	C	C	G	T	A	99Lives
Bengal	MVX1	A	C	C	G	T	A	99Lives
Bengal	S20423	A	C	C	G	T	A	99Lives
Bengal	Fcat 15571	GA	C	C	G	T	A	MALDI-TOF
Bengal	Fcat 15587							MALDI-TOF
Bengal	Fcat 15588	A	C		G	T		MALDI-TOF
Bengal	S10946	A	C	C	G	T	A	99Lives
Bengal	Fcat 9257	A	C	TC	G	T	A	MALDI-TOF
Bengal	Fcat 9588	GA	C	C	G	T	A	MALDI-TOF
Bengal	Fcat 15576		C	C	G	T	A	MALDI-TOF
BengalxJapaneseBobtail	S17799	A	C	C	G	T	A	99Lives
BengalxJapaneseBobtail	S13227	A	C	C	G	T	A	99Lives
BengalxJapaneseBobtail	S13230	A	C	C	G	T	A	99Lives
BengalxSomali	S15240	A	C	C	G	T	A	99Lives
Birman	Fcat 20796	A	C	C	G	./	A	99Lives
Birman	Fcat 20795	A	C	C	G	T	A	99Lives
Birman	Fcat 20794	A	C	C	G	T	A	99Lives
Birman	MVX2	A	C	C	G	T	A	99Lives
Birman	MVX3	A	C	C	G	T	A	99Lives
Birman	MVX4	A	C	C	G	T	A	99Lives
Birman	CR1397	A	C	C	G	T	A	99Lives
Birman	CR1219	A	C	C	G	T	A	99Lives
Birman	Fcat 16526	A	C		G	T		MALDI-TOF
Birman	Fcat 14349	GA	C	C	G	T	A	MALDI-TOF
Birman	Fcat 11049	A	C	TC	G	T		MALDI-TOF
Birman	Fcat 11351	A	C	TC	G	T	A	MALDI-TOF

Birman	Fcat 13055		CT	TC	G	T		MALDI-TOF
Birman	Fcat 7526	A	C	TC	G	T	A	MALDI-TOF
Birman	Fcat 7578	A	C	TC	G	T		MALDI-TOF
British Shorthair	Fcat 16308	A	C		G	T		MALDI-TOF
British Shorthair	S19747	A	C	C	G	T	A	99Lives
British Shorthair	S19748	A	C	C	G	T	A	99Lives
British Shorthair	Fcat 15928	A	C		G	T	A	MALDI-TOF
British Shorthair	Fcat 12784				G			MALDI-TOF
British Shorthair	Fcat 12786	A	C	TC	G	T	A	MALDI-TOF
British Shorthair	Fcat 7787	GA	C	TC	G	T	A	MALDI-TOF
British Shorthair	Fcat 7957	GA	C	TC	G	T	A	MALDI-TOF
Burmese	Fcat 20682	A	C	C	G	T	A	99Lives
Burmese	Fcat 20580	A	C	C	G	T	A	99Lives
Burmese	S19197	A	C	C	G	T	A	99Lives
Burmese	S19194	A	C	C	G	T	A	99Lives
Burmese	S19190	A	C	C	G	T	A	99Lives
Burmese	Fcat 15345		C	T	G	T		MALDI-TOF
Burmese	Fcat 16344	A	C		G	T		MALDI-TOF
Burmese	Fcat 10532	A	C	TC	G	T	A	MALDI-TOF
Burmese	Fcat 11667	GA	C	C	G	T	A	MALDI-TOF
Burmese	Fcat 13090	GA	C	TC	G	T		MALDI-TOF
Burmese	Fcat 16411	A	C	C	G	T	A	MALDI-TOF
Burmese	Fcat 16762	A	C		G	T		MALDI-TOF
Chartreux	Fcat 14581	GA				AT	G	MALDI-TOF
Chartreux	Fcat 12494	A	C		G	T	A	MALDI-TOF
Chartreux	Fcat 12557	A	C	TC	G	T	A	MALDI-TOF
Chartreux	Fcat 12756	A	C	TC	G	T	A	MALDI-TOF
Chartreux	Fcat 2526							MALDI-TOF
Chartreux	Fcat 7872	GA	C	TC	G	T	A	MALDI-TOF
Chartreux	Fcat 14740	A	C		G	T		MALDI-TOF
Cornish Rex	Fcat 7739	A	C	TC	G	T	A	MALDI-TOF
Cornish Rex	Fcat 7722	A	C	TC	G	T	A	MALDI-TOF
Cornish Rex	Fcat 15927		C		G	T		MALDI-TOF
Cornish Rex	Fcat 4935	GA	C	C	G	T	A	MALDI-TOF
Cornish Rex	Fcat 8550	A	C	TC	G	T	A	MALDI-TOF
Cornish Rex	Fcat 8551	GA		T	G			MALDI-TOF
Cornish Rex	Fcat 14652	GA	C		G	T		MALDI-TOF
Cornish Rex	Fcat 15832	GA			G			MALDI-TOF
Devon Rex	Fcat 15485							MALDI-TOF
Devon Rex	Fcat 18014	GA	C		G	T	A	MALDI-TOF
Devon Rex	S15020	A	C	C	G	T	A	99Lives
Devon Rex	S10361	A	C	C	G	T	A	99Lives

Devon Rex	Fcat 15558	GA	C	C	G	T	A	MALDI-TOF
Devon Rex	Fcat 11273	GA	CT	T	G	AT	GA	MALDI-TOF
Devon Rex	Fcat 12151	GA	C	TC	G	T	A	MALDI-TOF
Devon Rex	Fcat 14105	A	C	TC	G	T		MALDI-TOF
Devon Rex	Fcat 14114	A	C	T	G	T		MALDI-TOF
Egyptian Mau	K469	A	C	C	G	T	A	99Lives
Egyptian Mau	Fcat 20680	A	C	C	G	T	A	99Lives
Egyptian Mau	S20406	A	C	C	G	T	A	99Lives
Egyptian Mau	Fcat 15357	GA	C	TC	G	T		MALDI-TOF
Egyptian Mau	Fcat 15792	A	C	TC	G	T		MALDI-TOF
Egyptian Mau	Fcat 16646	A	C	TC	G	T		MALDI-TOF
Egyptian Mau	Fcat 17542	A	C		G	T		MALDI-TOF
Egyptian Mau	Fcat 14737		C	C	G	T	A	MALDI-TOF
Egyptian Mau	Fcat 14738	A	C	C	G	T	A	MALDI-TOF
Himalayan	S6259	A	C	C	G	T	A	99Lives
Japanese Bobtail	Fcat 11662	GA	C	C	G	T	A	MALDI-TOF
Japanese Bobtail	Fcat 16616	GA			G	AT		MALDI-TOF
Japanese Bobtail	Fcat 16687	A	C		G	AT		MALDI-TOF
Japanese Bobtail	Fcat 11417	GA	C	C	G	T	A	MALDI-TOF
Japanese Bobtail	Fcat 3685	GA	C	TC	G	T	A	MALDI-TOF
Japanese Bobtail	Fcat 9274	A	C	TC	G	T		MALDI-TOF
Japanese Bobtail	Fcat 9276	A	C	TC	G	T	A	MALDI-TOF
Korat	Fcat 7890	GA	C	TC	G	T		MALDI-TOF
Korat	Fcat 5467							MALDI-TOF
Korat	Fcat 14602	A	C	TC	G	T		MALDI-TOF
LaPerm	S19650	A	C	C	G	T	A	99Lives
Lykoi	S19725	A	C	C	G	T	A	99Lives
Maine Coon	K443	A	C	C	G	T	A	99Lives
Maine Coon	MVX5	A	C	C	G	T	A	99Lives
Maine Coon	S20425	A	C	C	G	T	A	99Lives
Maine Coon	Fcat 15347	A	C		G	T	A	MALDI-TOF
Maine Coon	Fcat 15577		C		G	T		MALDI-TOF
Maine Coon	Fcat 15851		C	TC	G	T		MALDI-TOF
Maine Coon	Fcat 14310	GA	C	C	G	T		MALDI-TOF
Maine Coon	Fcat 11544	A	C	TC	G	T	A	MALDI-TOF
Maine Coon	Fcat 12718	A	C	TC	G	T	A	MALDI-TOF
Maine Coon	Fcat 13901	GA	C	TC	G	T	A	MALDI-TOF
Maine Coon	Fcat 4934	GA	C	TC	G	T	A	MALDI-TOF
Maine Coon	Fcat 5983	GA	C	TC	G	T	A	MALDI-TOF
Maine Coon Cross	S_801	A	C	C	G	T	A	99Lives
Maine Coon Cross	S_802	A	C	C	G	T	A	99Lives
Maine Coon Cross	S_799	A	C	C	G	T	A	99Lives

Maine Coon Cross	S_798	A	C	C	G	T	A	99Lives
Maine Coon Cross	S_488	GA	C	C	G	T	A	99Lives
Manx	Fcat 13427	A	C	TC	G	T	A	MALDI-TOF
Manx	Fcat 13428			T	G	T	A	MALDI-TOF
Manx	Fcat 14233		C	T	G			MALDI-TOF
Manx	Fcat 15572		C	C	G	T	A	MALDI-TOF
Manx	Fcat 10676	GA	T		G	T		MALDI-TOF
Manx	Fcat 9084		C	TC	G	T		MALDI-TOF
Mixed breed	Fcat 16734	A	C		G	T		MALDI-TOF
Mixed breed	Fcat 16761		C	TC	G	T		MALDI-TOF
Mixed breed	Fcat 16763	A	C	TC	G	T		MALDI-TOF
Napoleon	S19050	A	C	C	G	T	A	99Lives
Napoleon	S19067	A	C	C	G	T	A	99Lives
Napoleon	S19060	A	C	C	G	T	A	99Lives
Norwegian Forest Cat	Fcat 15632	A	C	TC	G	T		MALDI-TOF
Norwegian Forest Cat	Fcat 16148		C		G	T	A	MALDI-TOF
Norwegian Forest Cat	Fcat 16277		C	T	G	T		MALDI-TOF
Norwegian Forest Cat	Fcat 16632	GA	C	C	G	T	A	MALDI-TOF
Norwegian Forest Cat	Fcat 14240							MALDI-TOF
Norwegian Forest Cat	Fcat 11550	A	C	TC	G	T	A	MALDI-TOF
Norwegian Forest Cat	Fcat 12388	GA	C	TC	G	T	A	MALDI-TOF
Norwegian Forest Cat	Fcat 4572		T	T	G	T		MALDI-TOF
Norwegian Forest Cat	Fcat 6004	GA	C	TC	G	T	A	MALDI-TOF
Ocicat	Fcat 20822	A	C	C	G	T	A	99Lives
Ocicat	Fcat 15063		C		G	T		MALDI-TOF
Ocicat	Fcat 12803	GA	C	TC	G	T	A	MALDI-TOF
Ocicat	Fcat 14111	A	C	TC	G	T	GA	MALDI-TOF
Ocicat	Fcat 3515		T	T	G	T		MALDI-TOF
Ocicat	Fcat 5548	GA	CT		G	AT	GA	MALDI-TOF
Ocicat	Fcat 5744	A	C	TC	G	T	A	MALDI-TOF
Ocicat	Fcat 6976	A	C	TC	G	T		MALDI-TOF
Oriental	Fcat 20684	A	C	TC	G	T	A	99Lives
Oriental	Fcat 16513	GA	C	C	G	T	A	MALDI-TOF
Oriental	Fcat 20793	A	C	C	G	T	A	99Lives
Oriental Shorthair	S5015	A	C	C	G	T	A	99Lives
Oriental Shorthair	S5012	A	C	C	G	T	A	99Lives
Oriental Shorthair	S4780	A	C	C	G	T	A	99Lives
Oriental Shorthair cross	S17994	A	C	C	G	T	A	99Lives
Oriental Shorthair cross	S18307	A	C	C	G	T	A	99Lives
Oriental Shorthair cross	S17993	A	C	C	G	T	A	99Lives
Persian	S14056	A	C	C	G	T	A	99Lives
Persian	CAT807	A	C	C	G	T	A	99Lives

Persian	Fcat 14887	A	C	TC	G	T		MALDI-TOF
Persian	Fcat 16273	A	C	TC	G	T		MALDI-TOF
Persian	Fcat 18440		C	TC	G	T		MALDI-TOF
Persian	Fcat 15925							MALDI-TOF
Persian	Fcat 10846	A	C	TC	G	T	A	MALDI-TOF
Persian	Fcat 13532	GA	C	TC	G	T	A	MALDI-TOF
Persian	Fcat 6167		C	TC	G	T	A	MALDI-TOF
PersianxJapaneseBobtail	S16628	A	C	C	G	T	A	99Lives
Peterbald	S20382	A	C	C	G	T	A	99Lives
Ragdoll	CAT1943	A	C	C	G	T	A	99Lives
Ragdoll	CAT160	A	C	C	G	T	A	99Lives
Ragdoll	Fcat 15956		C	C	G	T	A	MALDI-TOF
Ragdoll	Fcat 16657	A	CT	TC	G	AT		MALDI-TOF
Ragdoll	Fcat 16659		C	C	G	T	A	MALDI-TOF
Ragdoll	Fcat 17252				G			MALDI-TOF
Ragdoll	Fcat 17280	GA	C	TC	G	T		MALDI-TOF
Ragdoll	Fcat 16274	GA	C		G	T	A	MALDI-TOF
Ragdoll	Fcat 16456	A	C	TC	G	T		MALDI-TOF
Random bred	DS05501	A	C	C	G	T	A	99Lives
Random bred	DS05183	A	C	C	CG	T	A	99Lives
Random bred	DS04985	A	C	C	G	T	A	99Lives
Random bred	DS04852	A	C	C	G	T	A	99Lives
Random bred	K473	(T)GA	C	C	G	T	A	99Lives
Random bred	K472	(T)GA	C	C	G	T	A	99Lives
Random bred	K471	A	C	C	G	T	A	99Lives
Random bred	Fcat 20831	A	C	C	G	T	A	99Lives
Random bred	Fcat 20830	A	C	C	G	T	A	99Lives
Random bred	Fcat 20829	A	C	C	G	T	A	99Lives
Random bred	Fcat 19532	A	C	C	G	T	A	99Lives
Random bred	Fcat 19531	A	C	C	G	T	A	99Lives
Random bred	Fcat 20683	A	C	C	G	T	A	99Lives
Random bred	Katze169	A	C	C	G	T	A	99Lives
Random bred	S19983	A	C	C	G	T	A	99Lives
Random bred	MVX6	A	C	C	G	T	A	99Lives
Random bred	S18857	A	C	C	G	T	A	99Lives
Random bred	MVX7	A	C	C	G	T	A	99Lives
Random bred	MVX8	A	C	C	G	T	A	99Lives
Random bred	S19984	A	C	C	G	T	A	99Lives
Random bred	S19982	A	C	C	G	T	A	99Lives
Random bred	K450	A	C	C	G	T	A	99Lives
Random bred	S18421	A	C	C	G	T	A	99Lives
Random bred	S11615	A	C	C	G	T	A	99Lives

Random bred	Fcat 18132	A	C	TC	G	T	A	MALDI-TOF
Random bred	Fcat 18134		C		G	T		MALDI-TOF
Random bred	Fcat 18403		C	C	G	T	A	MALDI-TOF
Random bred	Fcat 18404	GA	C		G	T		MALDI-TOF
Random bred	Fcat 18405	GA	C	C	G	AT	A	MALDI-TOF
Random bred	Fcat 18406	GA	C	C	G	T		MALDI-TOF
Random bred	S18579	A	C	C	G	T	A	99Lives
Random bred	S18801	A	C	C	G	T	A	99Lives
Random bred	S18980	A	C	C	G	T	A	99Lives
Random bred	S18476	A	C	C	G	T	A	99Lives
Random bred	S11849	A	C	C	G	T	A	99Lives
Random bred	CAT3462	A	C	C	G	T	A	99Lives
Random bred	S5384	A	C	C	G	T	A	99Lives
Random bred	S10250	A	C	C	G	T	A	99Lives
Random bred	S18849	A	C	C	G	T	A	99Lives
Random bred	S18666	A	C	C	G	T	A	99Lives
Random bred	S18528	A	C	C	G	T	A	99Lives
Random bred	Fcat 6780		T	T	G	T	A	MALDI-TOF
Random bred	Fcat 6781			T	G			MALDI-TOF
Random bred	Fcat 16902	A	C		G	T	A	MALDI-TOF
Random bred	Fcat 4649	GA	C	C	G	T	A	MALDI-TOF
Random bred	SJA149	A	C	C	G	T	A	99Lives
Random bred	Fcat 2578		C	C	G	T	A	MALDI-TOF
Random bred	Fcat 2579	GA	C	C	G	T	A	MALDI-TOF
Russian Blue	Fcat 3720	GA	C	TC	G	T	A	MALDI-TOF
Russian Bue	Fcat 12329		C	TC	G	T	A	MALDI-TOF
Russian Blue	Fcat 12453	A	C	TC	G	T		MALDI-TOF
Russian Blue	Fcat 12521	A	C	TC	G	T	A	MALDI-TOF
Russian Blue	Fcat 7824	GA	C	TC	G	T		MALDI-TOF
Savannah	Fcat 20679	A	C	C	G	T	A	99Lives
Selkirk Rex	Fcat 20907	A	C	C	G	T	A	99Lives
Siamese	S_813	A	C	C	G	T	A	99Lives
Siamese	S_792	A	C	C	G	T	A	99Lives
Siamese	S_690	A	C	C	G	T	A	99Lives
Siamese	S_695	A	C	C	G	T	A	99Lives
Siamese	S624	A	C	C	G	T	A	99Lives
Siamese cross	S19349	A	C	C	G	T	A	99Lives
Siberian	S20424	A	C	C	G	T	A	99Lives
Siberian	Fcat 15351		C	TC	G	T	A	MALDI-TOF
Siberian	Fcat 16650	A	C	TC	G	T		MALDI-TOF
Siberian	Fcat 11777	A	C	TC	G	T	A	MALDI-TOF
Siberian	Fcat 12973	A	C	TC	G	T	A	MALDI-TOF

Siberian	Fcat 13503	A	C	TC	G	T	A	MALDI-TOF
Siberian	Fcat 15597							MALDI-TOF
Sphynx	Fcat 20034	A	C	C	G	T	A	99Lives
Sphynx	Katze159	A	C	C	G	T	A	99Lives
Sphynx	Fcat 16416	A	C	TC	G	T		MALDI-TOF
Sphynx	Fcat 15931		C	C	G	T	A	MALDI-TOF
Sphynx	Fcat 14925	A	C	TC	G	T		MALDI-TOF
Sphynx	Fcat 15582	A	C		G	T		MALDI-TOF
Sphynx	Fcat 16896	A	C	C	G	T		MALDI-TOF
Sphynx	Fcat 2048	GA	CT	TC	G		A	MALDI-TOF
Sphynx	Fcat 14586	GA	C	C	G	T	A	MALDI-TOF
Tennessee Rex	S20448	A	C	C	G	T	A	99Lives
Tonkinese	S10001	A	C	C	G	T	A	99Lives
Tonkinese	Fcat 16676	A	C	C	G	T	A	MALDI-TOF
Tonkinese	Fcat 18036		C	C	G	T	A	MALDI-TOF
Tonkinese	Fcat 18444	GA	C		G	T	A	MALDI-TOF
Tonkinese	Fcat 10439		CT	T	G	T	A	MALDI-TOF
Tonkinese	Fcat 14087	A	C	TC	G	T	A	MALDI-TOF
Tonkinese	Fcat 5549		T	T	G			MALDI-TOF
Tonkinese	Fcat 7261	GA	CT	T	G	AT	GA	MALDI-TOF
Toyger	Fcat 20781	A	C	C	G	T	A	99Lives
Turkish Angora	Fcat 15592							MALDI-TOF
Turkish Angora	Fcat 18279	A	C		G	T		MALDI-TOF
Turkish Angora	Fcat 14954	A	C		G	T	A	MALDI-TOF
Turkish Angora	Fcat 11537		C	TC	G	AT	A	MALDI-TOF
Turkish Angora	Fcat 12856	A	C	TC	G	T	A	MALDI-TOF
Turkish Van	Fcat 13556	GA	C	TC	G	T		MALDI-TOF
Turkish Van	Fcat 17372		C	TC	G	T		MALDI-TOF
Turkish Van	Fcat 9623	A	C		G	T	A	MALDI-TOF
Turkish Van	Fcat 13613	GA	C	T	G	AT	A	MALDI-TOF
Turkish Van	Fcat 9568	A	C	TC	G	T	A	MALDI-TOF
Turkish Van	Fcat 9905	A	C		G	T		MALDI-TOF
Abyssinian	Fcat 19911	GA	T	C	G	T	GA	UNKNOWN 1
Abyssinian	20038	G	CT	C	G	T	A	UNKNOWN 2
Abyssinian	20046	GA	C	C	G	AT	A	UNKNOWN 3
Abyssinian	20048	G	T	C	G	T	A	UNKNOWN 4
Abyssinian	21975	G	T	C	C	T	A	UNKNOWN 5
Abyssinian	21980	G	C	C	CG	T	A	UNKNOWN 6
Abyssinian	6934	GA	CT	C	C	T	A	UNKNOWN 7
Abyssinian	Mi1198	G	C	C	C	AT	A	UNKNOWN 8
Abyssinian	Mi1606	GA	C	C	G	T	A	UNKNOWN 9
Abyssinian	Mi1654	GA	T	C	CG	T	A	UNKNOWN 10

Abyssinian	Mi2077	GA	T	C	G	AT	A	UNKNOWN 11
Abyssinian	Mi2243	GA	CT	C	CG	T	GA	UNKNOWN 12
Abyssinian	Mi2624	GA	CT	C	CG	T	A	UNKNOWN 13
Abyssinian	Mi2747	GA	CT	C	G	AT	GA	UNKNOWN 14
Abyssinian	Mi2749	G	CT	C	G	AT	GA	UNKNOWN 15
Abyssinian	Fcat 14889	GA	T	TC	G	AT		UNKNOWN 16
Abyssinian	Fcat 15620	G	CT	TC		AT		UNKNOWN 17
Abyssinian	Fcat 12762	GA	T	TC	G	AT	A	UNKNOWN 18
Abyssinian	Fcat 13437	GA	C	TC	G	T	A	UNKNOWN 19
Abyssinian	Fcat 9904	A	C	TC	CG	T	GA	UNKNOWN 20
Abyssinian	AB 16	G	T	T	G			UNKNOWN 21
Abyssinian	AB 17	GA	T	T	CG	T		UNKNOWN 22
Abyssinian	AB 18	A	C	T	G	T		UNKNOWN 23
Abyssinian	AB 19			T	G			UNKNOWN 24
Abyssinian	AB 20		C	TC	G	AT		UNKNOWN 25
Abyssinian	AB 21	GA		T	G		GA	UNKNOWN 26
Abyssinian	AB 22							UNKNOWN 27
Abyssinian	AB 23			T	G			UNKNOWN 28
Abyssinian	AB 24	GA	T	T	G	T		UNKNOWN 29
Abyssinian	AB 25			T	G			UNKNOWN 30
Abyssinian	AB 26	G	T	T	G	T		UNKNOWN 31
Abyssinian	AB 27	GA	T	C	CG	T		UNKNOWN 32
Abyssinian	AB 28		T	T	G			UNKNOWN 33
Abyssinian	AB 29	G	T	TC	G	T		UNKNOWN 34
Abyssinian	AB 30	GA	T	TC	G	T		UNKNOWN 35
Abyssinian	AB 8	GA	T	TC	G	T		UNKNOWN 36

Supplementary Table 11. Allele frequencies and HWE tests for the six disease-associated SNPs in the Abyssinian cats.

	SNP9					SNP21				
	Rare	Affected + Healthy	Affected	Healthy	Unknown	Rare	Affected + Healthy	Affected	Healthy	Unknown
A/A		33	9	24	10		33	10	23	15
A/B		28	13	15	18		16	5	11	7
B/B		4	3	1	2		16	10	6	9
n		65	25	40	30		65	25	40	31
f(B)-rare allele	A	0.277	0.380	0.213	0.367	C	0.369	0.500	0.288	0.403
A/A		G/G	G/G	G/G	G/G		T/T	T/T	T/T	T/T
A/B		G/A	G/A	G/A	G/A		T/C	T/C	T/C	T/C
B/B		A/A	A/A	A/A	A/A		C/C	C/C	C/C	C/C
e[A/A]		33.985	9.610	24.806	12.033		25.861	6.250	20.306	11.040
e[A/B]		26.031	11.780	13.387	13.933		30.277	12.500	16.387	14.919
e[B/B]		4.985	3.610	1.806	4.033		8.861	6.250	3.306	5.040
χ^2 (Yates)		0.137	0.049	0.148	1.691		12.946	7.300	3.150	7.148
P		0.711	0.825	0.700	0.193		0.0003	0.007	0.075	0.007
	SNP22					SNP26				
	Rare	Affected + Healthy	Affected	Healthy	Unknown	Rare	Affected + Healthy	Affected	Healthy	Unknown
A/A		55	20	35	17		22	5	17	23
A/B		6	1	5	8		21	7	14	8
B/B		4	4	0	10		22	13	9	3
n		65	25	40	35		65	25	40	34
f(B)-rare allele	T	0.108	0.180	0.063	0.400	C	0.500	0.660	0.400	0.206
A/A		C/C	C/C	C/C	C/C		G/G	G	G/G	G/G
A/B		C/T	C/T	C/T	C/T		G/C	CG	G/C	G/C
B/B		T/T	T/T	T/T	T/T		C/C	C	C/C	C/C
e[A/A]		51.754	16.810	35.156	12.600		16.250	2.890	14.400	21.441
e[A/B]		12.492	7.380	4.687	16.800		32.500	11.220	19.200	11.118
e[B/B]		0.754	0.810	0.156	5.600		16.250	10.890	6.400	1.441
χ^2 (Yates)		13.024	14.049	0.767	8.024		7.115	2.368	2.146	1.446
P		0.0003	0.0001	0.381	0.005		0.008	0.124	0.143	0.229

	SNP32					SNP45				
	Rare	Affected + Healthy	Affected	Healthy	Unknown	Rare	Affected + Healthy	Affected	Healthy	Unknown
A/A		46	14	32	20		35	10	25	13
A/B		13	6	7	9		19	7	12	6
B/B		6	5	1	0		9	7	2	0
n		65	25	40	29		63	24	39	19
f(B)-rare allele	A	0.192	0.320	0.113	0.155	G	0.294	0.438	0.205	0.158
A/A		T/T	T	T/T	T/T		A/A	A	A/A	A/A
A/B		T/A	AT	T/A	T/A		A/G	GA	A/G	A/G
B/B		A/A	A	A/A	A/A		G/G	G	G/G	G/G
e[A/A]		42.404	11.56	31.506	20.698		31.432	759.375	24.641	13.474
e[A/B]		20.192	10.88	7.987	7.603		26.135	118.125	12.718	5.053
e[B/B]		2.404	2.560	0.506	0.698		5.432	459.375	1.641	0.474
χ^2 (Yates)		6.432	3.559	0.030	0.164		3.716	2.844	0.017	0.041
P		0.011	0.059	0.863	0.685		0.054	0.092	0.897	0.839

Supplementary Table 12. Single SNP logistic regression.

Single SNP	Additive	Recessive	Incomplete	D
Logistic Regressions			Dominance	
SNP9				
χ^2	4.690	2.355	4.771	
P	0.030	0.125	0.09	
Intercept	-1.012	-0.572	-0.981	
SE(Int)	0.377	0.267	0.391	
P(Int)	0.007	0.032	0.012	
Slope	0.927	0.836	1.040	0.838
SE(Slope)	0.443	0.592	0.609	0.544
P(Slope)	0.036	0.158	0.088	0.124
R ₂	0.072	0.037	0.073	
-2LogLikelihood	81.926	84.262	81.845	0.775
SNP21				
χ^2	4.047	5.082	5.086	
P	0.044	0.024	0.079	
Intercept	-0.957	-0.818	-0.833	
SE(Int)	0.368	0.31	0.379	
P(Int)	0.009	0.008	0.028	
Slope	0.624	0.665	0.672	0.044
SE(Slope)	0.316	0.301	0.320	0.659
P(Slope)	0.048	0.027	0.036	0.946
R ₂	0.064	0.080	0.080	
-2LogLikelihood	82.569	81.534	81.530	0.308
SNP22				
χ^2	2.843	8.700	9.106	
P	0.092	0.004	0.010	
Intercept	-0.648	-0.644	-0.560	
SE(Int)	0.281	0.269	0.280	
P(Int)	0.021	0.017	0.046	
Slope	0.800	10.405	10.363	-11.412
SE(Slope)	0.497	3.628.436	3.628.436	3.628.436
P(Slope)	0.107	0.998	0.998	0.997
R ₂	0.043	0.105	0.119	
-2LogLikelihood	83.773	78.546	77.510	0.012

SNP26				
χ^2	6.316	5.931	6.533	
P	0.012	0.015	0.038	
Intercept	-1.334	-0.949	-1.224	
SE(Int)	0.465	0.34	0.509	
P(Int)	0.004	0.005	0.016	
Slope	0.812	0.658	0.796	0.531
SE(Slope)	0.337	0.275	0.334	0.688
P(Slope)	0.016	0.017	0.017	0.440
R ₂	0.097	0.092	0.100	
-2LogLikelihood	80.300	80.685	80.083	0.641
SNP32				
χ^2	6.238	5.647	6.730	
P	0.012	0.017	0.035	
Intercept	-0.874	-0.668	-0.827	
SE(Int)	0.316	0.275	0.320	
P(Int)	0.006	0.015	0.010	
Slope	1.004	1.139	1.218	0.672
SE(Slope)	0.426	0.565	0.571	0.642
P(Slope)	0.018	0.044	0.033	0.295
R ₂	0.096	0.086	0.103	
-2LogLikelihood	80.378	80.969	79.886	0.483
SNP45				
χ^2	6.630	6.752	7.620	
P	0.01	0.009	0.022	
Intercept	-1.054	-0.747	-0.955	
SE(Int)	0.363	0.286	0.372	
P(Int)	0.004	0.009	0.010	
Slope	0.936	1	1.104	0.55
SE(Slope)	0.379	0.426	0.442	0.589
P(Slope)	0.014	0.019	0.012	0.350
R ₂	0.103	0.105	0.117	
-2LogLikelihood	79.986	79.864	78.996	0.320

Supplementary Table 13. Multivariate additive logistic regression.

Multivariate Logistic regressions	Additive	Recessive	Incomplete	D
			Dominance	
χ^2	36.517	24.687	45.205	
P	0	0.0004	0	
Intercept	-5.359	-1.728	-6.595	
SE(Int)	1.302	0.463	1.802	
P(Int)	0	0.0002	0.0003	
Slope(SNP9)	2.110	0.923	2.076	4.058
SE(Slope)	0.746	0.661	1.294	1.370
P(Slope)	0.005	0.162	0.109	0.003
Slope(SNP21)	0.806	0.518	0.946	0.702
SE(Slope)	0.469	0.358	0.530	1.108
P(Slope)	0.086	0.149	0.074	0.526
Slope(SNP22)	1.425	10.328	11.098	-11.456
SE(Slope)	0.759	3.531.972	3.500.245	3.500.246
P(Slope)	0.060	0.998	0.997	0.997
Slope(SNP26)	1.172	0.457	1.147	2.458
SE(Slope)	0.496	0.370	0.556	1.286
P(Slope)	0.018	0.217	0.039	0.056
Slope(SNP32)	1.526	0.999	1.951	0.356
SE(Slope)	0.572	0.642	0.854	1.013
P(Slope)	0.008	0.120	0.022	0.725
Slope(SNP45)	1.197	0.637	1.968	0.546
SE(Slope)	0.570	0.518	0.814	1.039
P(Slope)	0.036	0.218	0.016	0.599
R ₂	0.479	0.338	0.565	
-2LogLikelihood	50.099	61.929	41.411	0.192

Supplementary Table 14. Probability of developing disease P(disease) in the Abyssinian Population for each of the 729 possible genotype combinations at the six disease-associated SNPs.

	Obs.	SNP9	SNP21	SNP22	SNP26	SNP32	SNP45	Disease	Predict
Cat_ID	Phen	A	C	T	C	A	G	Probab.	Phen
6	Aff.	2	1	0	2	0	2	0.988	Aff.
11	Aff.	0	2	2	2	0	2	0.979	Aff.
10	Aff.	1	2	0	2	0	2	0.956	Aff.
19	Aff.	2	2	0	1	0	(0.875)	0.937	Aff.
9	Aff.	0	2	0	2	1	2	0.924	Aff.
4	Aff.	2	0	0	1	0	2	0.919	Aff.
2	Aff.	0	2	1	2	0	2	0.917	Aff.
15	Aff.	0	2	2	0	2	0	0.896	Aff.
16	Aff.	1	0	0	2	1	1	0.861	Aff.
8	Aff.	0	0	2	0	2	1	0.855	Aff.
12	Aff.	0	0	1	1	2	1	0.816	Aff.
23	Aff.	1	1	0	2	1	0	0.807	Aff.
14	Aff.	0	1	2	1	1	0	0.730	Aff.
22	Aff.	1	2	0	1	0	1	0.674	Aff.
20	Aff.	1	2	0	2	0	0	0.669	Aff.
17	Aff.	1	0	0	2	1	0	0.652	Aff.
5	Aff.	1	0	0	2	0	1	0.573	Aff.
21	Aff.	1	1	0	1	1	0	0.564	Aff.
13	Aff.	0	1	0	2	0	2	0.542	Aff.
7	Aff.	0	0	0	2	2	0	0.509	Aff.
58	Heal.	1	0	0	0	1	1	0.372	Heal.
52; 79	Heal.	0	0	0	2	0	2	0.347	Heal.
47	Heal.	0	1	0	1	1	1	0.339	Heal.
55	Heal.	1	0	0	1	0	1	0.294	Heal.
40	Heal.	1	0	0	2	0	0	0.290	Heal.
56	Heal.	1	1	0	0	1	0	0.286	Heal.
62	Heal.	0	2	0	2	0	(0.401)	0.282	Heal.
31	Heal.	1	1	0	0	0	1	0.223	Heal.
68	Heal.	0	2	0	2	0	0	0.196	Heal.
65	Heal.	1	2	0	0	0	0	0.163	Heal.
54	Heal.	1	0	0	0	1	0	0.153	Heal.
35; 37	Heal.	1	0	1	0	0	0	0.140	Heal.
32	Heal.	0	1	1	1	0	0	0.124	Heal.
44	Heal.	0	1	0	1	0	1	0.100	Heal.
53	Heal.	0	1	0	2	0	0	0.099	Heal.

66	Heal.	0	2	0	1	0	0	0.070	Heal.
49	Heal.	0	0	0	1	1	0	0.065	Heal.
45; 46; 48; 61	Heal.	0	0	0	1	0	1	0.048	Heal.
39	Heal.	0	0	0	2	0	0	0.047	Heal.
34; 38;63	Heal.	1	0	0	0	0	0	0.038	Heal.
64	Heal.	0	1	0	1	0	0	0.033	Heal.
60	Heal.	0	1	0	0	0	1	0.033	Heal.
41	Heal.	0	0	1	0	0	0	0.019	Heal.
36; 43	Heal.	0	0	0	1	0	0	0.015	Heal.
78	Heal.	0	1	0	0	0	0	0.011	Heal.
33; 42	Heal.	0	0	0	0	0	0	0.005	Heal.
24*	Aff.	1	0	0	0	2	0	0.453	Heal.
1*	Aff.	1	2	0	0	0	0	0.163	Heal.
30*	Aff.	1	0	0	0	0	1	0.114	Heal.
18*	Aff.	0	2	0	1	0	0	0.070	Heal.
3*	Aff.	0	0	0	1	0	1	0.048	Heal.
51*	Heal.	0	2	0	2	2	0	0.838	Aff.
59*	Heal.	2	1	1	0	0	0	0.751	Aff.
67*	Heal.	1	2	0	2	0	0	0.669	Aff.
57*	Heal.	1	0	0	1	1	1	0.656	Aff.
50*	Heal.	1	1	0	0	1	1	0.569	Aff.

For each Abyssinian cat having an observed phenotype (Obs. Phen) and a genetic profile at the six loci, the probability for disease development (Disease Prob.) and the expected phenotype (Predict. Phen) were estimated.

Cat ID: cat identification number-see Table 1. Aff. = affected, Heal. = not affected. SNPs = variants at each locus identified as associated to the disease in the Abyssinian population. Genotypes are reported as 1 = heterozygous, 2 = homozygous for the variant.

For Cat_ID 19 and 62, genotype values at SNP45 were imputed.

*Discordant cats considering actual health status versus probability of disease.

Supplementary Table 15. Amino acid sequence alignment of the regions containing the six associated variants in different species. The amino acid change site is framed. Sequences from Ensembl *Felis catus* vs 6.2, *Canis familiaris* vs 3.1, *Equus caballus* vs2, *Homo sapiens* vs GRCh38.p10, *Bos taurus* vs UMD3.1.

Species	Ensembl ID	Chr.	SNP 9 - GPRC6A															
<i>Felis catus</i>	ENSFCAT000000315691	B2	T	C	T	E	V	T	V	A	M	A	A	A	L	R	F	L
<i>Canis familiaris</i>	ENSCAFT000000014142	1	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.
<i>Equus caballus</i>	ENSECAT000000203501	10	.	.	.	D	.	.	.	.	.	.	.	.	.	.	.	.
<i>Homo sapiens</i>	ENST000003103577	6	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.
SNP 21 - AEN																		
<i>Felis catus</i>	ENSFCAT000000315161	B3	A	G	P	N	A	K	D	V	L	R	R	K	H	K	T	R
<i>Canis familiaris</i>	ENSCAFT000000182863	3	.	.	.	.	T	.	.	.	.	.	.	.	.	R	K	.
<i>Bos taurus</i>	ENSBTAT000000234545	1	.	S	.	.	.	.	.	.	.	.	.	.	.	R	K	.
<i>Homo sapiens</i>	ENST000003328103	15	.	.	.	.	.	.	.	.	.	.	K	R	.	R	.	.
SNP 22 - EFCAB6																		
<i>Felis catus</i>	ENSFCAT000000073413	B4	Y	L	D	L	A	A	G	P	G	D	A	P	R	N	G	P
<i>Canis familiaris</i>	ENSCAFT000000014004	10	.	.	.	F	.	.	F	E	.	.	.	K	M	S	R	.
<i>Equus caballus</i>	ENSECAT000000080961	28	.	.	.	F	.	.	F	E	V	P	.	M	Q	E	R	.
<i>Bos taurus</i>	ENSBTAT0000000107965	5	.	.	.	F	.	.	F	E	.	P	.	K	M	.	E	.
SNP 26 - STAC3																		
<i>Felis catus</i>	ENSFCAT000000001903	B4	N	G	E	A	V	G	A	G	G	G	P	I	Y	Y	I	Y
<i>Canis familiaris</i>	ENSCAFT000000003233	10	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
<i>Equus caballus</i>	ENSECAT000000256971	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
<i>Bos taurus</i>	ENSBTAT0000000244225	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
<i>Homo sapiens</i>	ENST000003327826	12	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SNP 32 - TIAM1																		
<i>Felis catus</i>	ENSFCAT000000023643	C2	L	S	D	P	S	P	S	P	Q	D	S	T	G	P	Q	L
<i>Bos taurus</i>	ENSBTAT0000000237165	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.
<i>Homo sapiens</i>	ENST000004919271	21	.	.	.	Q	.	.	.	.	.	.	.	.	.	.	.	.
SNP 45 - NCOR1																		
<i>Felis catus</i>	ENSFCAT000000100043	E1	Y	Q	S	T	M	S	R	G	S	P	M	M	N	R	T	S
<i>Canis familiaris</i>	ENSCAFT000000287044	5	.	P	N	.	I	.	.	.	.	.	.	.	.	.	A	.
<i>Equus caballus</i>	ENSECAT000000082721	11	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.
<i>Bos taurus</i>	ENSBTAT0000000613432	19	.	A	.	.	V	.	.	.	.	.	.	L	S	.	A	.
<i>Homo sapiens</i>	ENST000002687127	17	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.

2.3. Genomic Localization and Genetic Characterization of Amyloidosis in the Siamese Cat.

Abstract

The whole genome sequences of two affected Siamese, available through the 99 Lives Project, were used to detect variants related only to affected individuals of this breed. The sequences were compared to other 113 whole genome sequences of other cats in the database that were used as control, and the version 6.2 of the *Felis catus* genome assembly was used as the reference genome. Fifteen new variants have been found, underlying the presence of several and different genes involved in amyloidosis disease, even within the same species. These variants haven't been genotyped in the population yet, due to an essential low number of cases and control, not sufficient at the moment. A worldwide call has been established via Winn Feline Foundation for Siamese/Oriental samples collection to genotype the variants, to perform a statistical analysis and to confirm the presence of amyloidosis associated variants.

Materials and Methods

Samples

The WGS of two affected Siamese with approximately 30x genome coverage were available as part of the 99 Lives Cat Genome Sequencing Initiative (www.felinegenetics.missouri.edu/99Lives), together with the WGS of other 113 domestic cats that were used as controls for variant exclusion (Control Population; **Table1**). The two affected individuals were selected for the study to be ≤ 3 years of age and with a confirmatory diagnosis of hepatic rupture due to primary liver amyloidosis. The Control Population included other 12 cats between Siamese and Oriental Shorthair breeds (OSH). These individuals joined the 99 Lives project for other purpose and they were considered healthy concerning the amyloidosis status, representing therefore an additional control for variants exclusion within the breed (**Table 1**).

Table 1. Control population of cats for the analysis of Abyssinian amyloidosis, excluding the reference genome of Cinnamon.

N.	Breed	99 Lives	MALDi-TOF	N.	Breed	99 Lives	MALDi-TOF
1	Abyssinian	2		20	Norwegian		9
2	American Shorthair	1	8	21	Ocicat	1	7
3	American Wirehair	1		22	Oriental	6	1
4	Bengal	8	6	23	Persian	3	7
5	Birman	8	7	24	Peterbald	1	
6	British Shorthair	2	6	25	Ragdoll	2	7
7	Burmese	5	7	26	Russian blue		5
8	Chartreux		7	27	Savannah	1	
9	Cornish Rex		8	28	Selkirk rex	1	
10	Devon rex	2	7	29	Siamese	6	
11	Egyptian Mau	3	6	30	Siberian	1	6
12	Hymalaian	1		31	Sphynx	2	7
13	Japanese bobtail		7	32	Tennessee rex	1	
14	Korat		3	33	Tonkinese	1	7
15	LaPerm	1		34	Toyger	1	
16	Likoi	1		35	Turkish Angora		5
17	Maine Coon	8	9	36	Turkish Van		6
18	Manx		6	37	Random bred	39	15
19	Napoleon	3			Total	112	169

Variants Detection

All the analyses were performed and referred to the version 6.2 of the *Felis catus* genome assembly (<http://www.ncbi.nlm.nih.gov/assembly/320798>). The genome-wide variant analysis was carried out using different active filters through the Maverix Biomics, Inc. platform (Santa Cruz, CA). To screen for potential causal variants, the filtering considered at least one of the two affected Siamese to be homozygous and heterozygous for the variants and the other 113 cats in the 99 Lives set with an allele frequency of < 1%, thus, considering genotyping errors or the odd chance of other cats carrying the variants but at a low frequency.

Results

The variants detection performed through Maverix Biomics platform led to the identification of 15 potential amyloidosis related variants reported in **Table 2**. Only one variant was homozygous in both the affected cats, with high severity effect and a non-synonymous coding in MYO1B gene. Fcat_20684 resulted to be homozygous for other three SNPs in ENSFCAG00000027369, HPX and ENSFCAG00000026113 genes respectively. All of these variants were not present in Fcat_20973 but those in HPX and ENSFCAG00000026113 were shared by other cats. The only variant with a high effect on the gene coding is the TG deletion on chromosome A2, causing a frameshift in the C7orf63 gene (**Table 2**).

Table 2. List of the 15 variants identified in the two affected Siamese and of their predicted effect.

CHR.	POS.	Ref.	Alt.	Effect	Severity	Aminoacid Change	Gene Name	Fcat_20684	Fcat_20793
A1	112122236	A	G	Non-synonymous coding	Moderate	N954S	SEC24A	X	X
A2	94324276	C	G	Non-synonymous coding	Moderate	L317V	STEAP1	X	X
A2	94448119	CTG	C	Frame shift	High	TV930	C7orf63	X	X
A2	94495873	C	T	Non-synonymous coding	Moderate	P157S	ENSFCAG00000019336	X	X
C1	178130023	C	T	Non-synonymous coding	Moderate	R703C	MYO1B	X*	X*
C1	18031427	A	G	Non-synonymous coding	Moderate	E169G	ENSFCAG00000010324	X	X
D1	116490750**	G	A	Non-synonymous coding	Moderate	S71F	PHRF1	X	X
D1	64054514	C	T	Non-synonymous coding	Moderate	H270Y	ENSFCAG00000027369	X*	
D1	65425127**	C	A	Non-synonymous coding	Moderate	A380S	HPX	X*	
E1	43098853	C	T	Non-synonymous coding	Moderate	R272C	NAGLU	X	X
E3	19602169**	T	C	Non-synonymous coding	Moderate	L21P	PRSS8	X	X
E3	19656216**	C	T	Non-synonymous coding	Moderate	R51C	ZNF668	X	X
E3	19907310**	C	T	Non-synonymous coding	Moderate	G380E	PHKG2	X	X
F1	15169283**	C	T	Non-synonymous coding	Moderate	G119S	ENSFCAG00000026113	X*	
F1	36084396**	C	T	Non-synonymous coding	Moderate	S473L	CRB1	X	X

*Cat homozygous for the variant

**Variant shared by other cats

Conclusions

This study has identified variants that could be related to amyloidosis disease within the Siamese breed. These 15 variants are not shared by the Abyssinian breed, which means that feline amyloidosis could be addressed to different factors, even when two breeds of the same species are considered. The genes found in this study need to be investigated but first of all, variants have to be confirmed through their genotyping in a Siamese population with both affected and healthy individuals. To proceed in this way, additionally samples have to be included in the study so that a correct statistical analysis can be performed.

2.4. Multi-omics Characterization of Primary Amyloidosis in the Abyssinian Cat

Abstract

Three different omics approaches were applied in this study to try to identify the mechanisms underlying Abyssinian amyloidosis. The whole genome sequences of two affected cats were compared to those of other 195 cats (control population) for the detection of disease associated variants. Proteome and miRNAome analyses were performed on the same FFPE kidney specimens of additional 3 affected and 3 healthy Abyssinians to characterize the amyloid deposits and to investigate potentially dysfunctional aspects of the cellular gene expression regulation. The work allowed to detect i) 139 potential disease associated genomic variants; ii) 196 proteins expressed only in the pathological kidneys, 54 proteins related only to the healthy tissue and 21 proteins differentially expressed in the two tissues; iii) 60 miRNAs differentially expressed in affected and healthy kidneys. All the omics data were matched to reconstruct interactions and bring new insights in the mechanism of this feline amyloidosis.

Materials and Methods

Samples

Eight Abyssinian cats were included in the study. Six individuals (PN#) were archival FFPE renal specimens of the Department of Veterinary Medicine – University of Milan. Two cats (Fcat#) were available as whole genome sequences, part of the 99Lives cat WGS initiative (www.felinegenetics.missouri.edu/99Lives) [Mauler *et al.*, 2017].

All the Abyssinians were accurately selected basing on their pathology reports (necropsy and histology). The cats were classified as “affected” when they were 5 years old or younger, with pathology report of amyloidosis confirmed by renal tissue positivity to Congo red staining and excluding other intercurrent pathologies. One affected cat (Fact#16515 – S2) was included in the study despite being 6 years old, as it had a report of primary amyloidosis and it belongs to a North American line of

familial amyloidosis. The cats were classified as “healthy” when older than 5yo and with a report attesting the absence of amyloidosis (**Table 1**).

Table 1. Abyssinian samples used in the study.

Sample_ID	Cat_ID	Status	Age at death (years)	Type	Analysis
PN# 1/08 A	A1	Affected	2	FFPE	Proteomics and miRNA
PN# 160/09	A2	Affected	5	FFPE	Proteomics and miRNA
PN# 233/02	A3	Affected	4	FFPE	Proteomics and miRNA
PN# 119/11	A4	Healthy	5,8	FFPE	Proteomics and miRNA
PN# 131/12	A5	Healthy	6	FFPE	Proteomics and miRNA
PN# 44/14	A6	Healthy	12	FFPE	Proteomics and miRNA
Fact#18778	S1	Affected	4	WGS	Genomics
Fact#16515	S2	Affected	6	WGS	Genomics

WGS of additional 195 cats of different breeds and random bred cats from the 99 Lives project (Control Population – **Supplementary Table 1**) were used for genomic variant detection. The Control Population included two other Abyssinian cats. One of them was the Abyssinian used as the reference genome (Cinnamon) and the other one was included in the 99Lives Project for a different disease.

Genomic Variation

Detection of disease related variants

All the WGS data were available as Variant Call Format (VCF) file via the 99 Lives project. The file included 84005772 annotated variants. VCFtools [Danecek *et al.*, 2011] was used to perform the filtering and to obtain a new VCF file with high quality variants only. The function –missing-indv was used to remove the individuals with >20% of missing genotypes, while the function –max-missing was used to exclude all the variants that were missing in >20% of the individuals. Finally, -minGQ and -minDP were used to set the Genotype Quality (GQ) ≥ 20 and the Read Depth (RD) ≥ 15 respectively.

The VCF file with the high-quality variants was used to detect variants related to the amyloidosis affected status, using the Golden Helix SVS software [Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com]. Different options were set at the beginning of the analysis: samples were considered as individual unrelated samples and the two affected Abyssinians were set on “TRUE”, while all the other cats were

set on “FALSE” to be considered as controls. The analysis was done using the function “Activate Variants by Genotype Count Threshold” available through the SNP & Variation Suite v8.4 [Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com], which allowed to activate those variants with genotypes that meet a count or percentage threshold based on user’s specific filters. As the Control Population could theoretically include disease carriers, thresholds were set to consider variants mostly unique to the two cases and rare in the controls. Therefore, variants were selected when present in both the two affected cats, either with a heterozygous or homozygous genotype, and when present in the remaining 195 cats with genotype frequency < 2% (counting both homozygous and heterozygous).

Resulting variants were prioritized based on their integration with the data from proteomics and miRNAs and they were all analyzed with the Ensembl Variant Effect Predictor (VEP) [McLaren *et al.*, 2016] to be annotated and to investigate their potential function. The nucleotide sequences of the regions harboring prioritized variants were aligned to the genome of other species with the BLAST function of Ensembl genome browser 97 (<https://www.ensembl.org/index.html>) to evaluate the conservation grade of each region.

Kidney Proteome Analysis

Protein extraction from Formalin-Fixed Paraffin-Embedded (FFPE) Blocks

Eight sections of 10 µm of thickness were sliced from each FFPE block of three affected and three healthy cats (A1, A2, A3 and A4, A5, A6 respectively; **Table 1**) and temporary stored at 4 °C. Affected kidneys presented massive amyloid deposits. For protein extraction, FFPE tissue sections were deparaffinized and rehydrated by successive washes in xylene (2×), 100% (2×), 95% (1×), 70% (1×) ethanol, water (1×) and 50 mM Tris HCl buffer. After air-drying, the tissue sections were harvested with a scalpel blade, collected in a clean tube and homogenized using a Potter homogenizer in 500 µL of protein extraction buffer (8 M urea, 20 mM Hepes pH 8, with mini protease inhibitors cocktail) at full speed for 1-3 minutes. The homogenate was centrifuged at 15000 rpm for 15 minutes at 4 °C to sediment unhomogenized tissue and large cellular debris. The pellet was discarded, and the recovered proteins were quantified with Bradford’s protein assay.

Shotgun Mass Spectrometry and Label Free Quantification

Extracted proteins were subjected to reduction with 13 mM dithioerythritol (DTE) (15 min at 50°C) and alkylation with 26 mM iodoacetamide (IAA; 30 min at RT). Peptide digestion was conducted using sequence-grade trypsin (Roche) over night at 37°C using a protein:trypsin ratio of 20:1 [Vernocchi *et al.*, 2014; Maffioli *et al.*, 2018]. Digestion was quenched by the addition of 2 µL of 98% formic acid, into the sample solution. The proteolytic digest was desalted using Zip-Tip C18 (Millipore) and evaporated in a SpeedVac. Prior to LC-MS, samples were reconstituted in 0.1% formic acid [Tedeschi *et al.*, 2011]. LC-ESI-MS/MS analysis was performed on a DionexUltiMate 3000 HPLC System with a PicoFritProteoPrep C18 column (200 mm, internal diameter of 75 µm) (New Objective, USA). Gradient: 2% ACN in 0.1 % formic acid for 10 min, 2-4 % ACN in 0.1% formic acid for 6 min, 4-30% ACN in 0.1% formic acid for 147 min and 30-50 % ACN in 0.1% formic for 3 min at a flow rate of 0.3 µL/min. The eluate was electrosprayed into an LTQ-OrbitrapVelos (Thermo Fisher Scientific, Bremen, Germany) through a Proxeon nanoelectrospray ion source (Thermo Fisher Scientific) [Galli *et al.* 2018]. The LTQ-Orbitrap was operated in positive mode in data-dependent acquisition mode to automatically alternate between a full scan (m/z 350-2000) in the Orbitrap (at resolution 60000, AGC target 1000000) and subsequent CID MS/MS in the linear ion trap of the 20 most intense peaks from full scan (normalized collision energy of 35%, 10 ms activation). Isolation window: 3 Da, unassigned charge states: rejected, charge state 1: rejected, charge states 2+, 3+, 4+: not rejected; dynamic exclusion enabled (60 s, exclusion list size: 200) [Maffioli *et al.* 2017]. Data acquisition was controlled by Xcalibur 2.0 and Tune 2.4 software (Thermo Fisher Scientific).

Statistical and bioinformatic analysis of the proteomic data

MS spectra were searched against the *Felis catus* NCBI sequence database (release 31.01.2017) by MaxQuant (version 1.3.0.5) [Maffioli *et al.*, 2017]. The following parameters were used: initial maximum allowed mass deviation of 15 ppm for monoisotopic precursor ions and 0.5 Da for MS/MS peaks, trypsin enzyme specificity, a maximum of two missed cleavages, carbamidomethyl cysteine as fixed modification, N-terminal acetylation and methionine oxidation, as variable modifications. False protein identification rate (1%) was estimated by searching

MS/MS spectra against the corresponding reversed-sequence (decoy) database. Minimum required peptide length was set to 6 amino acids and minimum number of unique peptides supporting protein identification was set to 1. Quantification in MaxQuant was performed using the built-in label-free quantification algorithms (LFQ) based on extracted ion intensity of precursor ions [Cox & Mann 2008; Zhang *et al.* 2012].

Three replicates for both healthy control and pathological groups were carried out respectively. Only proteins present and quantified in at least 2 out of 3 were considered as positively identified in a sample and used for statistical analyses. Statistical analyses of Max Quant results were performed using the Perseus software module (version 1.4.0.8, www.biochem.mpg.de/mann/tools/). A t-test ($p\text{-value} \leq 0.05$) was carried out to identify proteins differentially expressed among the different conditions. Proteins were considered to be differentially expressed if they were present only in one condition or showed significant t-test difference ($p\text{-value} \leq 0.05$).

Bioinformatic analyses were carried out by Panther software (Version 12.1) [Mi *et al.*, 2013] and DAVID software (release 6.7) [Huang *et al.*, 2009], to cluster enriched annotation groups of Cellular Component (CC), Biological Process (GOBP), Molecular Function (GOMF) and Pathways within the set of up-regulated or present only in affected cats' proteins. Functional grouping was based on modified Fisher exact EASE score $p\text{ value} \leq 0.05$ and at least three counts.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [Vizcaíno *et al.*, 2016] partner repository with the dataset identifier PXD015703.

Kidney miRNAome Analysis

miRNA isolation

Five sections of 10 μm of thickness were sliced from each FFPE block of three affected and three healthy cats (A1, A2, A3 and A4, A5, A6 respectively; **Table 1**), the same blocks used for the proteomic analysis. The total miRNAome was isolated using the miRNeasy FFPE kit by Qiagen with minor modification to the manufacturer instructions (Handbook 06/2015). Single or double steps of deparaffinization were applied according to the quantity of paraffin that covered the specimens. Moreover,

for samples A1, A2 and A6, the second step of deparaffinization was characterized by an incubation time increased to 8 minutes, while samples A4 and A5 were incubated twice at 56° C for 30 seconds before the second round of deparaffinization. Samples were then eluted in 25µl of RNase-free water. Extracted miRNAs were quantified with the RNA HS Assay of the Qbit 2.0 by Invitrogen (https://assets.thermofisher.com/TFSA-Assets/LGS/manuals/Qubit_RNA_HS_Assay_UG.pdf). Length and quality were then evaluated with the Bioanalyzer Agilent RNA 6000 Nano Assay (https://www.agilent.com/cs/library/usermanuals/public/RNA-6000-Nano_QSG.pdf).

Sequencing

A total amount of 200ng of extracted miRNA/sample was sequenced using the smallRNA-seq kit by Illumina and the Illumina NextSeq500 platform, according to the Company's protocol. The read length obtained was 1x75bp for a total of 30 million reads. The quality control of the reads was assessed with FastQC [<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>] and the adapter sequences were removed using Cutadapt [Martin M. 2011].

miRNA Identification and Expression

MiRNA mapping and profiling were obtained with MiRDeep2 [Friedländer *et al.* 2012]. The function -mapper.pl was used to collapse the reads and to map the fastq files against the vs 9.0 of the cat genome assembly, that was used as reference sequence and retrieved from NCBI online database [<https://www.ncbi.nlm.nih.gov/genome?term=felis%20catus>]. The -mirDeep2.pl function was used to identify miRNAs and to match them to the human ones using miRbase database vs22.1 [Griffiths-Jones *et al.* 2006], as the cat miRNAome is still not available on online databases. Finally, the -quantifier.pl function was used to determine miRNAs expression.

MiRNAs having less than 10 reads per sample were filtered out. Raw counts were normalized with TMM method (Trimmed mean of M-value) and expressed as log2 CPM (counts per million) to reduce errors. Finally, the differential expression

between affected and control cats was assessed with a moderated t test by limma and nominal P-values were adjusted by Benjamin-Hochberg method.

Data analyses

The pre-miRNA sequences obtained with miRDeep2 that did not match any human miRNAs, were used as input data for the alignment with the BLASTN function on miRbase database (vs22.1). The pre-sequences were aligned against human and against the genome of *Canis familiaris*, *Bos Taurus*, *Suis scrofa* and *Mus musculus* that were available in the database. This analysis was carried out to double check miRDeep2 output and to identify possible matches in other species.

All the filtered miRNAs were used to evaluate the samples distribution through a Multidimensional Scaling (MDS) plot performed with R Studio routine [<https://www.rstudio.com/products/rpackages/>].

Following the differential expression analysis, the significant miRNAs (P-val<0.05) were represented with a Heatmap [<https://www.rstudio.com/products/rpackages/>] and their pre and mature sequences were matched with miRNAs already reported in the domestic cat [Laganà *et al.* 2017; Sun *et al.* 2014] to identify potential novel feline miRNAs.

For the identification of miRNA genes, pre-miRNA sequences were aligned to the cat reference genome vs9.0 using the BLAST tool of Ensembl genome browser 97 [<https://www.ensembl.org/index.html>].

MiRNet database [Fan *et al.* 2016] was used to verify if the identified Abyssinian miRNAs could target and regulate the proteins detected with the proteomic approach. The prediction was done using human as a model, since the cat miRNAome is not available online yet, and the kidney was specified as the tissue in which the interactions take place.

The matching between miRNAs and the mutated genes identified with the genomic analysis was performed through a customized algorithm retrieved from TargetScan vs7.2 [Riffo-Campos *et al.* 2016]. A specific Perl script was used to identify the binding sites of the feline miRNAs on the 3'Untranslated Region (UTR) of the mutated genes.

Finally, GO and Pathway analysis of the significant differentially expressed miRNAs were evaluated with MiRNA Enrichment Analysis and Association database [Backes *et al.* 2016].

Results and Discussion

Genomic detection of disease related variants

With the filtering of the 99Lives VCF file, three individuals were removed as they had >20% of missing genotypes, leaving 194 cats for the further analyses. Additional filtering for excluding variants missing in >20% of the individuals, led to consider 81123337 variants, which all passed the filters for GQ and RD.

Starting from the high-quality variants and comparing the data of the two affected Abyssinians to those of the remaining 192 cats of the Control Population, 139 potential disease related variants were detected with Golden Helix Software (**Supplementary Table 2**). The majority of the variants mapped in intergenic regions (n.80) or in intronic regions (n.55), but none of them was found in regulatory regions, according to the Ensembl VEP [McLaren *et al.*, 2016]. One single variant was located into an exon but was synonymous, and three variants mapped within 3'UTRs (**Table 2**). When focusing on the genotypes, heterozygous variants were the most common (n.86) while n.53 homozygous variants were identified, five of them being homozygous in both the affected cats. Regarding variant types, 121 Single Nucleotide Polymorphism (SNP), 14 deletions and 4 insertions were detected (**Table 2, Supplementary Table 2**).

Table 2. Overview of the variants identified in the two affected Abyssinians compared to the Control Population with the Golden helix SVS Software.

	TOT	HOMOZYGOUS	HETEROZYGOUS		SNPs	INSERTIONS	DELETIONS
INTERGENIC	80	30	50		69		11
INTRONIC	55	21	34		49	4	2
EXONIC	1	0	1		1		
3'UTR	3	2	1		2		1
TOT	139	53	86		121	4	14

The 139 genomic variants were prioritized to be integrated and match with the Proteome and miRNAome results. It is well known that in animals, miRNAs regulate the gene expression through their binding to the 3'UTR transcripts [Dweep *et al.* 2013] and that SNPs within 3'UTR regions can disrupt regulatory sites compromising the gene expression [Maurano *et al.* 2012] or they can alter the

binding of regulatory factors such as miRNAs [Moszyńska *et al.* 2017]. Therefore, the priority was given to variants mapping into 3'UTR that were annotated in Septin 8 (SEPT8; A1:111852820), Erythrocyte Membrane Protein Band 4.1 like 5 (EPB41L5; C1:120549939) and ATP Binding Cassette Subfamily A Member 8 (ABCA8; E1:51790720) genes. The variants occurring in SEPT8 and ABCA8 were C>G and a C>T SNPs, respectively. The variant in the EPB41L5 gene was an adenine deletion (**Supplementary Table 2**). According to the Human Protein Atlas database (<http://www.proteinatlas.org>) [Pontén *et al.*, 2008], EPB41L5 is an intracellular protein, while ABCA8 is a membrane protein transporter. Both of them are expressed in all tissues including kidney. SEPT8 was a gene already reported in literature to be involved in the human amyloid pathway. It is a member of the conserved septin family of nucleotide binding proteins and is expressed in kidney [Wasik *et al.*, 2019] as well as in all the other human tissues. In human neurons, it was shown that SEPT8 modulate the beta-amyloidogenic processing of APP through the regulation of the Beta Amyloid cleaving enzyme 1 (BACE1). Particularly, decreased levels of *SEPT8* results in a reduction of BACE1 levels and subsequently decreased β APP [Kurkinen *et al.* 2016].

The sequences harboring the three prioritized variants were aligned to those of other species to evaluate the region conservation. Only 3'UTR in SEPT8 was conserved in different species, while the 3'UTR regions of ABCA8 and EPB41L5 were conserved only within the Felidae family. (**Supplementary Figure 1**).

Kidney Proteomic Analysis

Proteome and miRNAome analyses were conducted on the same kidney FFPE blocks from three affected and three healthy cats (**Table 1**). Pathological specimens presented massive amyloid deposits. To characterize the protein component of the deposits, a quantitative label free shotgun proteomic approach was adopted. This non-gel-based proteomic technique has emerged as powerful tool for studying large scale differential protein expression [Zhu *et al.* 2010] allowing to examine the impact of different conditions by achieving the simultaneous identification of thousands of proteins and their quantification in each sample. The MS/MS data identify 740 proteins in the kidney with amyloid deposits and 598 proteins from the control kidneys. 544 proteins were expressed in both pathological and healthy tissues,

while 196 proteins were exclusively present in the deposits and 54 proteins were expressed only in the controls. Out of the 544 proteins common to healthy and affected samples, 21 were differentially expressed ($p\text{-value} \leq 0.05$, Student's t-test): 19 proteins were up-regulated, and 2 proteins were down regulated in kidneys with amyloid deposits (**Figure 1**).

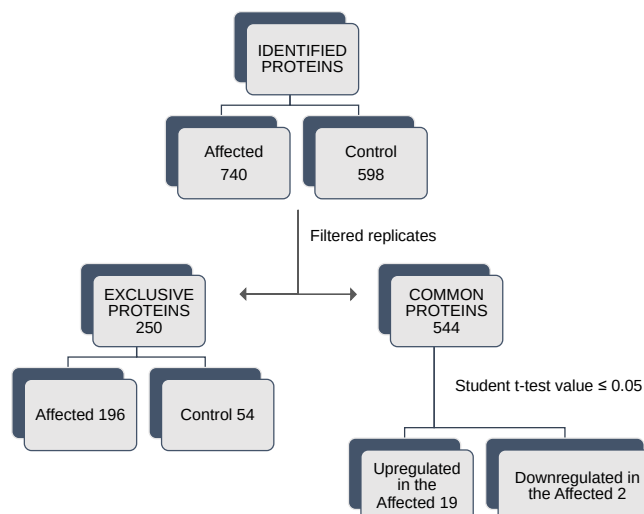


Figure 1. Workflow of the proteomic approach. A shotgun proteomic analysis was performed on FFPE kidney tissue from cats affected by renal amyloidosis and healthy cats (control). Statistical analyses were performed using the Perseus software (version 1.4.0.8, www.biochem.mpg.de/mann/tools/).

The exclusive and the differentially expressed proteins are listed in **Supplementary Tables 3 and 4**. SAA, APOA and TTR are within the most known proteins in the classic literature and in the cat too, for being related to kidney amyloid deposits. They were among the proteins exclusively detected in the pathological kidneys, confirming this form of amyloidosis as an SAA-like amyloidosis [Di Bartola *et al.* 1985]. Both exclusively and differentially expressed proteins (n.271) found in the deposits were analyzed for Gene Ontology (GO) annotation enrichment and functional grouping by Panther and DAVID software to cluster enriched annotation groups in terms of the highest enrichment score. The results showed that renal amyloidosis is related to main cellular component such as ribosome and extracellular matrix, as well as to cellular pathways involving oxidoreductase activity,

TCA cycle and blood coagulation (**Supplementary Table 5; Supplementary Table 6**). The functional groupings identified through the analysis suggested that the proteins detected in the pathological kidney could be representative of different phases and aspects of the progressive tissue damage occurring in the late disease.

- Extracellular space region and matrix (ECM)

The results obtain from Panther and DAVID data analyses (**Supplementary Table 5; Supplementary Table 6**) clearly revealed an important enrichment in extracellular matrix proteins (AEBP1, SERPINF1, MGP, ECM1, AHSG) and in extracellular space proteins (AEBP1, DDB1, GUSB, CES5A, ECM1, PRDX1, TTR, LAMP2, CKM, SERPINF1, SERPINF2, LOC101097240, LOC101099498, PON1, LTF, SERPIND1, CFI, CTSH, MUC5B), in agreement with previous publications reporting that the composition of protein insoluble deposits in the extracellular spaces can include ECM components [Brandan & Inestrosa 1993], and suggesting that the ECM modulates the level of amyloidogenic peptides [Bronfman et al. 1996]. These effects could be due to a direct regulation of the amyloidogenic protein expression by ECM signaling molecules and/or to a modulation of protein processing by the ECM related cytoskeletal network in the cytoplasmic compartment [Clark & Brugge 1995].

- Metabolism

David analysis (**Supplementary Table 5**), showed that renal amyloidosis affects catabolic processes such as fatty acid degradation (GCDH, ACADS, ALDH2, CPT1A, ALDH3A2, ACSBG1), citric acid cycle (ACO2, OGDHL, CS, PCK2, MDH2, FH), pyruvate metabolism (ME1, LDHA, ALDH2, ACACB, PCK2, MDH2, ALDH3A2, FH) and more generally the carbon metabolism (ME1, ACO2, PFKL, ACADS, OGDHL, CS, PGD, PGLS, G6PD, H6PD, HAO2, RGN, MDH2, FH), in accordance with previous findings suggesting that ATP at concentrations of physiological significance to both intra- and extracellular environments actively participated in the fibrillogenesis and is involved in a number of amyloidosis [Exley C. 1997].

- Inhibition of serine protease activity and coagulation

Bleeding manifestation in human patients with amyloidosis is a multifactorial event which can occur in advanced stage of the disease. In general, the vascular infiltration of amyloid fibrils can disrupt vascular integrity, and bleeding manifestations in patients with renal amyloidosis may be augmented by acquired haemostatic abnormalities associated with renal dysfunction [Gastineau *et al.* 1991]. Moreover, acquired deficiency of serine endo peptidase factor X (Stuart factor) is the most common coagulation factor deficiency that has been identified in patients with amyloidosis [Choufani *et al.* 2001], probably due to the apparent ability of amyloid fibrils to bind factors X [Gastineau *et al.* 1991; Schroeder *et al.* 1978; Furie *et al.* 1977; Greipp *et al.* 1981]. The data reported in **Supplementary Table 5**, clearly show an increase in categories related to serine-type endopeptidase inhibitor activity. This evidence is strongly supported by Panther data analysis. **Supplementary Table 6** shows a significant enrichment in proteins involved in this process such as those encoded by SERPINF1, SERPINF2, SERPIND1, ITIH3 and SERPINH1 genes.

- Antioxidant activity

Amyloidosis is a complex disease process with numerous pathological sequelae. Without doubt, the progressive accumulation of amyloid in organs contributes to architectural damage and organ dysfunction. Increasing levels of oxidative stress around the amyloid deposits have been detected in a variety of amyloid diseases [Liu *et al.* 2008; Saito *et al.* 2005; Witkowski *et al.* 2018], but it is not clear whether oxidative stress is involved in the progression of amyloidosis. The results obtained by Panther analysis reveal an increased oxidoreductase and antioxidant activity (**Supplementary Table 5; Supplementary Table 6**).

- Impact on structural constituent of ribosome and translation

From Panther data analysis reported in **Supplementary Table 6**, renal amyloidosis in cats has a deep influence on both ribosomes (RPS26, RPS19, SLC25A13, LOC101092708, RPS3A, RPL9, SLC25A3, RPLP2, RPL37A, RPS20, RPS7), and translation processes (RPS26, RPS19, SLC25A13,

LOC101092708, RPS3A, SLC25A3, RPL37A, RPS20, RPS7). These data are in agreement with David software data analysis (**Supplementary Table 5**) and to previous studies in which the in vivo accumulation of protein aggregates mainly composed of ribosomal components has been reported [Pathak et al. 2017].

Kidney miRNAome Analysis

Total miRNAs isolation was more successful when two steps of deparaffinization were carried out on the specimens as shown in the length and quality plots of the **Supplementary Figure 2**. Weak peaks, corresponding to small amount of other longer types of RNAs, were noticed for samples A4 and A5 (**Supplementary Figure 2**).

The sequencing data analysis showed that the number of reads per sample ranged from 27,607,431 to 52,126,090 (mean value 40,267,032). For each sample, the highest number of reads fell in the range of the 18-23 bp length, which is what is expected considering that the standard length of a mature miRNA sequence is around 18-22bp. Samples A1 and A4 showed a higher amount of reads with a 10-17 bp length (data not shown) which could be addressed to the presence of other longer RNAs that were not perfectly removed during the isolation step. Overall, a good sequencing quality was obtained (**Supplementary Figure 3**).

The library size, the miRNA density and the reads count were evaluated before and after the normalization with the TMM method and are reported in **Supplementary Figure 4**. A total of 854 different miRNA precursor sequences were identified with miRDeep2, corresponding to 618 unique mature sequences of feline miRNAs, reported with the same miRNA ID. Among these miRNAs with the same ID, those having the genomic region supported by the highest number of reads were used in the differential expression analysis (**Supplementary Table 7**). After filtering and normalization with miRDeep2, only 260/618 miRNA mature sequences were considered (**Supplementary Table 7**). Out of these 260 miRNAs, 71 had no match with human miRNAs, therefore they kept the name assigned by the software with “NC” as a prefix, meanwhile the other 189 were named with the human

nomenclature (prefix “hsa-mir”) (**Supplementary Table 7**). For a further check, the miRNA pre-sequences of these 71 NC_miRNAs were compared to those ones present in dog, bovine, mouse, swine, and human as well, using the BLASTN function on miRbase database. Eight miRNA pre-sequences turned out to perfectly match the human ones and were added to the 189 hsa-miRNA for a total of 197 human miRNAs (**Supplementary Table 8**). Two miRNAs were found only in bovine and swine, while 61 remained “NC” and were therefore considered as specific feline miRNAs.

All the filtered miRNAs were then used to evaluate the samples distribution with an MDS plot. The affected cats grouped in a separate cluster (**Figure 2**).

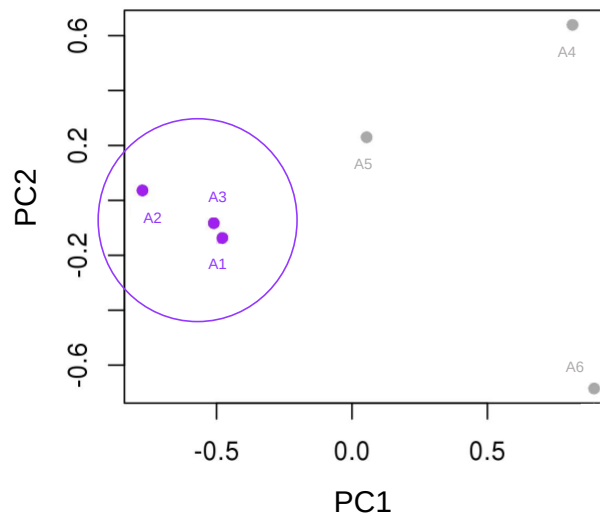


Figure 2. MDS plot realized with R Studio [<https://www.rstudio.com/products/rpackages/>], showing the two clusters that represent the affected and the healthy cats. Affected individuals (purple) are grouped on the left, while the healthy individuals (grey) are grouped on the right.

When the differential expression between miRNAs identified in affected and healthy cats was evaluated, 60 miRNAs resulted significant based on the nominal P-value (<0.05) and seven of them had significant adjusted P-value too (<0.05) after the Benjamin-Hochberg correction (**Table 3; Supplementary Table 9**).

Table 3. List of the 7 miRNAs that were significant also after the Benjamin-Hochberg correction (Adjusted P-value<0.05).

miRNA_ID (miRDeep2)	P-Value	Adj. P-Val
NC_018730.3_21281	0.0001	0.0189
NC_018727.3_13107	0.0002	0.0189
hsa-miR-4451	0.0002	0.0189
NC_018731.3_22438	0.0005	0.0318
NC_018726.3_8619	0.0007	0.0318
hsa-miR-4654	0.0007	0.0318
hsa-miR-186-5p	0.0011	0.0426

The result shown by the MDS (**Figure 2**) was confirmed with the heatmap based on the P-values of the 60 significant miRNAs. The plot shows four perfectly distinct blocks, with 27 miRNAs downregulated in the affected cats and upregulated in the healthy ones, and the remaining 33 miRNAs with the opposite trend (**Figure 3**).

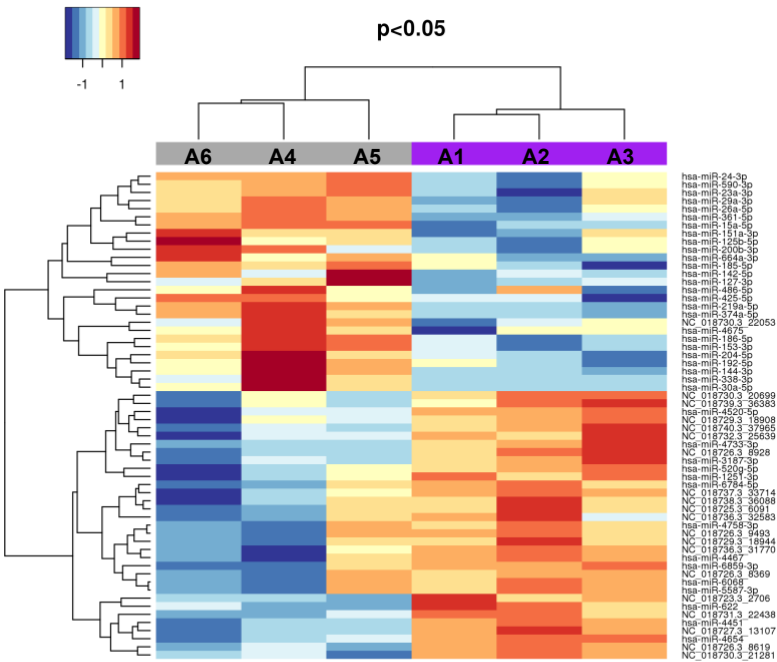


Figure 3. Heatmap realized with R Studio [<https://www.rstudio.com/products/rpackages/>], with the 60 significant miRNAs (Pval<0.05). The grey bar and the purple bar represent the healthy (A4, A5 and A6) and the affected (A1, A2, A3) Abyssinians respectively. The shades of blue refer to downregulated miRNAs while the shades of red refer to upregulated miRNAs.

The 60 significant miRNAs were compared to those reported in previous studies of miRNAs in the domestic cat, and 37 resulted to be already known in feline literature [Laganà *et al.* 2017; Sun *et al.* 2014], while the remaining ones have never been described in cats, with 3 of them matching the human miRNAs and 20 exclusive feline miRNAs (**Supplementary Table 10**).

Alignment was performed to the cat genome reference sequence vs 9 to identify the miRNA genes coding for the 60 significant miRNAs. MiRNA genes were already annotated on Ensembl database for 21 miRNAs, 37 miRNAs aligned in regions which still not have the corresponding gene annotated and only one did not have any alignment (**Supplementary Table 10**).

Gene Ontology (GO) analysis of the 60 significant miRNAs was performed and showed the “endoplasmic reticulum unfolded protein response” pathway as the most associated pathway (P-value=0.0018) with 18 miRNAs involved. All these 18 miRNAs belonged to the group of downregulated miRNAs in the affected cats after the differential expression analysis, nine of them with an already confirmed role in AD in literature (hsa-miR-590-3p [Villa *et al.* 2011]; hsa-miR-26a-5p [Chang *et al.* 2017]; hsa-miR-15a-5p [Kumar & Reddy 2016]; hsa-miR-125b-5p [Jin *et al.* 2018]; hsa-miR-200b-3p [Higaki *et al.* 2018]; hsa-miR-153-3p [Zhou *et al.* 2019]; hsa-miR-144-3p [Li Kuo *et al.* 2016]; hsa-miR-338-3p [Qian *et al.* 2019]). In particular, hsa-miR-186-5p was also among the seven miRNAs with significant Adjusted P-value (P-val<0.05) and it is known to play an important role in the regulation of BACE1 protein, the enzyme responsible of A β production, that is considered one of the main targets in AD [Ben Halima *et al.* 2016]. Finally, hsa-miR-142-5p and hsa-miR-425-5p are reported in literature, but interestingly they result to be upregulated in AD patients [Sierksma *et al.* 2018; Martinez & Peplow 2019]. This discrepancy could be addressed to the differences between the species considered and by the different roles that miRNAs could play in cats compared to humans.

MiRNA targets were predicted for the 60 final miRNAs, considering proteomic and genomic results separately. The match between detected proteins versus miRNAs was evaluated with miRNet, specifying in the analysis that the kidney was the tissue in which the interactions take place. The analysis showed 89 proteins regulated by 22 of the identified miRNAs (**Figure 4**).

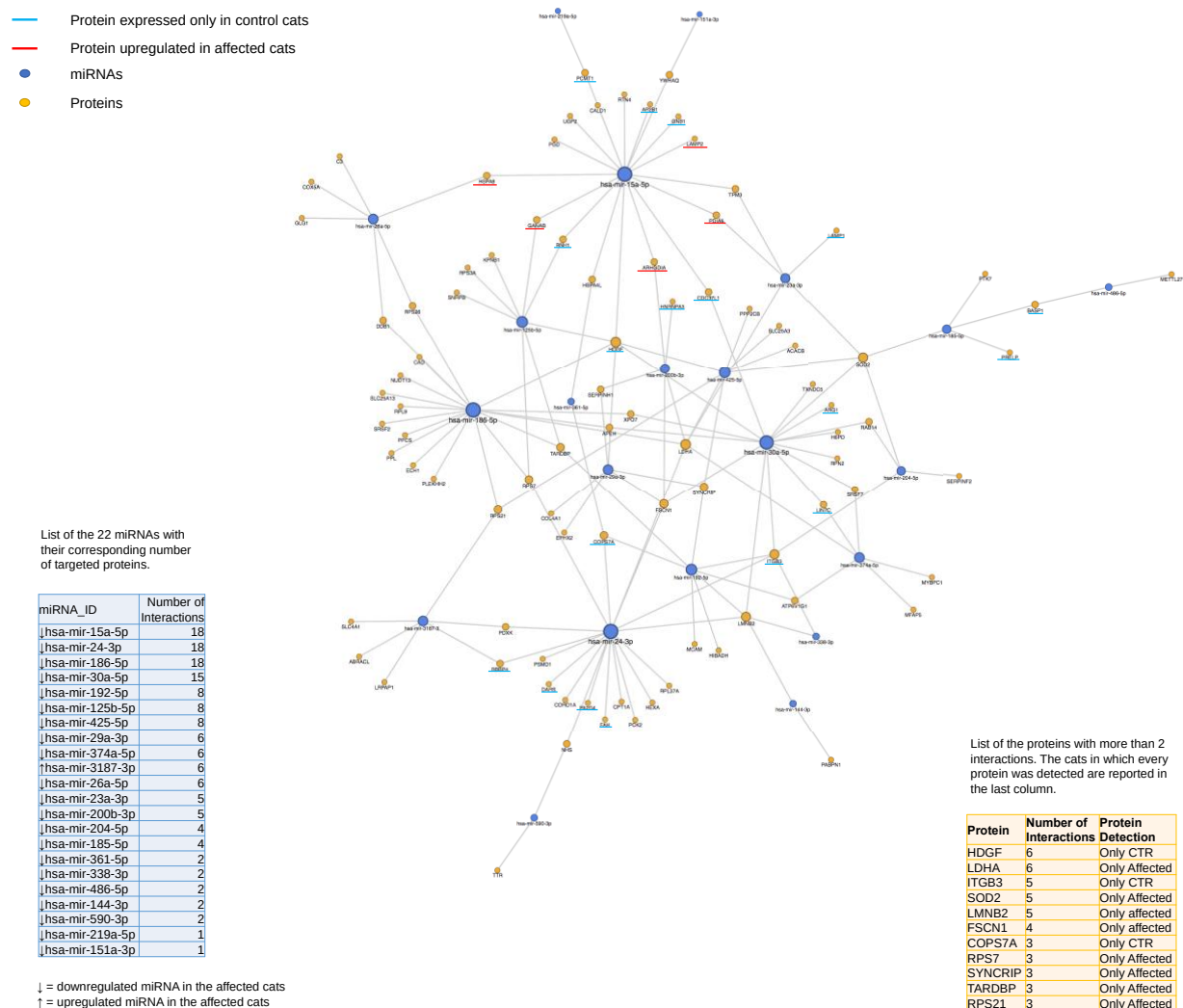


Figure 4. Network plot showing the interaction in kidney tissue among proteins and miRNAs identified in the study, according to miRNet database. The dot dimension is directly proportional to the number of the miRNA interactions.

Hsa-miR-186-5p was again among the most relevant miRNAs since, together with hsa-miR-15a-5p and hsa-miR-24-3p, it represented the node of the network with more interactions (n.18). Almost all the proteins in the network were proteins expressed exclusively in the affected cats (**Figure 4**), with only 18 out of 89 proteins present only in the control cats. All the miRNAs in the network were among the downregulated miRNAs in amyloidosis cats, except for hsa-miR-3187-3p, which was upregulated in the affected cats and it was also one of the miRNAs with a high

number of interactions (n.6). Among the proteins targeted by hsa-miR-3187-3p, there was COPS7A protein, which resulted to be expressed only in control cats. This is interesting as COPS7A is a component of the COP9 signalosome complex (CSN), which regulates the activity of the ubiquitin conjugation pathway in the process of protein degradation. It could be hypothesized that upregulated levels of miR-3187-3p in the affected cats can affect COPS7A transcription, explaining the missed detection of the proteins in the deposits.

The proteins whose genes represented the main targets of the identified miRNAs, with the higher number of interactions in the network, were in order HDGF, LDHA, ITGB3, SOD2, LMNB2 and FSCN1 (**Figure 4**). Briefly, HDGF and ITGB3 were proteins expressed only in healthy tissues with no reports in amyloidosis literature, to our knowledge. Instead, LDHA is among the proteins expressed only in the affected tissues and was already reported in literature as an important prognostic marker in AL amyloidosis. High levels of LDH were detected in the 40% of older age patients with severe AL amyloidosis forms [Mughtar et al. 2017]. The upregulation of SOD2 was recently found to be associated to the β amyloid-induced neurotoxicity in neuronal cells [Du et al. 2019]. This could resemble the Abyssinian kidney condition as SOD2 is among the proteins detected only in the affected cats and its high expression could be interpreted as one of the responses to the toxicity induced by the presence of amyloid deposits. LMNB2 is a laminin type B, a fibrillary protein that has already been shown to be involved in the onset of tauopathies, including AD [Frost et al. 2016].

Even if with minor interactions, the presence of LAMP1 and LAMP2 in the network was interesting. These are two different lysosomal trans-membrane proteins representing the 50% of all the lysosomal proteins. They are responsible for providing selectins with carbohydrate ligands in the lysosomal uptake followed by protein degradation. It is suggestive that, in the present study, despite no mutation at DNA level, LAMP1 resulted to be expressed only in control cats, while LAMP2 resulted to be exclusively present in the affected cats whereas, in knockout mice, it was shown that the deficiency of one is compensated by the upregulation of the other and vice versa [Eskelinen et al. 2002].

Finally, the match between the mutated 3'UTR regions of *SEPT8*, *EPB41L5*, *ABCA8* genes and miRNAs was investigated. In this case, it was evaluated which miRNAs had these 3'UTR regions as a target and if the binding occurred in the mutation site. Only *SEPT8* and *EPB41L5* were regulated, matching 17 and 27 different miRNAs respectively (**Supplementary Table 11**). The deletion in *EPB41L5* gene was the only variant to map in the binding site of miR-186-5p, one of the most significant downregulated miRNAs (Adj. P-value= 0.0426). The 3'UTR region of the *EPB41L5* gene with the corresponding binding site for miR-186-5p are reported in **Figure 5**. However, no correlation between this gene and amyloidosis are reported in literature to our knowledge and no hypothesis could be advanced based on the results obtained through this approach.

EPB41L5-201 3'UTR SEQUENCE (4040-4617)

TGAAGAATGGACATTTCTGTATTATTGTCCCTTTTCTCTCAAAAATTCAAACACTCTGGTTTTGAATGAT
CAGACAAAAGTTGGAAGCTTTTCTATCTATTAGCCTTCAGGATGAGGAGACTTGTGGATTATAGAAATT
TGCTCTGAGCTTAAGTCTGCTGTGTGAACATGTGCCAGAGCCGGAATTATTTCCACTTTTAAGATAGGATAC
GTGAGTCTCTTTATCACATTGCTTGAGATAATAATTCACTCACTCCCAGGATTTCTTTTATCAGGGAGGTG
TTTTAGTACAGATCCCTTGCTCTTTTATTTTCATAAAATAAATTTCTTAACCTGGGTATGATAGACAGTTTAA
GGACCTCACTAACTATTTGTGTTTACTTTTATATATGTAAAGCTTTCAGTAGCTTTAAAGTGTAATTTAT
TTTTATATCAAGTGCACTATTCATTGTGTCCACAGCTTTATATGACTTGTATTAGAAGCACTGCACTGAACT
GAGTTGTTTGTATACTGACATGACAACCGTAGATGTGACCACAAATAAAGCTTTTGTGAAACATT

Figure 5. 3'UTR region of the *EPB41L5* gene from nucleotide 4040 to nucleotide 4617. Nucleotides in red show the binding site of the miR-186-5p mature sequence and the deleted nucleotide in the mutated individuals is highlighted in blue.

Conclusions

Amyloidosis is a disease with difficult diagnosis. The samples enrolled in the present study have to be considered rare specimens, therefore the large amount of data obtained represents an important source for further meta-analyses and annotations.

This study represents the first investigation with a multi-omics approach of feline renal amyloidosis in the Abyssinian breed. The genomic analysis revealed no obvious amyloidogenic candidate genes and no strong evidences for mono or major genes. Almost all the variants mapped into intronic or intergenic regions and their functional investigation required the identification of their target gene. However, the new discoveries on the 3D structure of the genome have shown that regulatory elements can be found in distant regions from the genes that they regulate [Lieberman-Aiden *et al.*, 2009], making the association between the variant and the target gene difficult. This aspect, together with the poor annotation of the cat assembly, suggested to avoid any interpretation on the function of these polymorphisms. The only genomic variants showing a potential functional relation to the other data of the study were 3'UTR regions of three different coding genes (**Supplementary Table 2**).

The proteomic analysis on the pathological kidneys compared to the normal ones, showed several proteins exclusive of the amyloid tissue. The deposits were characterized by well-known potentially amyloidogenic proteins, mainly SAA, TTR and APOA4 (**Supplementary Table 3**). SAA was assumed to be genetically wild type, not being mutant in the WGS of the affected cats, in contrast with the suggestion of an “amyloidogenic” polymorphism reported in feline literature [Di bartola *et al.* 1985; Niewold *et al.* 1999]. Previous Sanger sequencing of the SAA N-terminal “amyloidogenic” region of three affected Abyssinians and a larger cohort of controls revealed very little variability, not associated to the affected status (**Supplementary Figure 5**). The massive presence of SAA in the deposits could be not in contrast with the low serum concentration detected in a longitudinal study of affected Abyssinians [Paltrinieri *et al.* 2015]. In fact, in experimental amyloidosis, it has been shown that rapid incorporation of SAA into amyloid reduces circulating SAA concentration by up to 90% [Simons *et al.*, 2013].

The exclusive identification of TTR in the deposits is in accordance with the human forms where both wild-type and mutated TTR can be detected [Merlini *et al.*, 2003; Adams *et al.*, 2019].

The data that gave a deeper insight into the cellular mechanisms of the disorder resulted from miRNA analysis. Hsa-mir-186-5p was one of the most significant downregulated miRNAs in the affected kidney (Adj. P-value 0.0426). Moreover, it was among the three miRNAs with the highest number of interactions in the

proteins-miRNAs network. This miRNA is considered one of the main targets in AD for its role in the regulation of the enzyme BACE1, which is directly responsible for the A β production [Ben Halima *et al.* 2016]. Even if BACE1 protein was not detected through the proteomic analysis, the present study shows that hsa-miR-186-5p target also SEPT8 (**Supplementary Table 12**), whose downregulation involves a reduction of the BACE1 protein with a consequent decrease in the levels of APP β [Kurkinen *et al.* 2016]. So, it is not surprising that BACE1 was not identified in the deposits.

Finally, the study characterized the feline miRNAome of both normal and amyloid affected kidneys and led to the identification of 260 miRNAs, with 20 of them representing novel feline miRNAs.

The limitations of this study could be represented by the low number of samples and by the type of tools used for the data analysis, which are mainly based on humans and do not always have the cat among the species included in the database.

Abyssinian amyloidosis is a juvenile form with extracellular deposits, in contrast with the intracellular ones that characterize AD, a condition with a late onset. However, the study performed through the integration of omics results allowed to collect a considerable amount of data which emphasize the similarity of the mechanisms underlying amyloidosis and AD, both complex diseases with different factors involved.

The biochemical characterization of the kidney deposits by SAA and some known amyloid related proteins, such as TTR, confirmed what reported in the classical feline literature and classifies the Abyssinian form within the “SAA amyloidosis” or “SAA-TTR amyloidosis”. However, no evidences on gene mutations that might determine misfolding or insolubility of those proteins in the extracellular matrix have been found. The proteomic data together with the miRNAome analysis allowed to zoom into the biological mechanisms of the renal cells and of their external environment in affected and normal tissue. Quite as expected, the results showed a mixed panorama of pathways, with very few of them that might be more directly related to the initial trigger of the disorder (perhaps ECM), and others more evidently related to secondary effects (such as altered coagulation). However, the trigger of the disease remained unsolved.

Supplementary Tables

Supplementary Table 1. Number of cats per breed in the 197 WGS available via the 99Lives Initiative.

Breed	Cat/breed
Abyssinian	4
American Shorhair	1
American Wirehair	1
Australian Mist	1
Bengal	5
Bengal x Japanese Bobtail	3
Bengal x Somali	1
Birman	10
British Shorthair	2
Burmese	13
Cross-breed colony	4
Devon Rex	5
Domestic Longhair	1
Don Sphynx	1
Egyptian Mau	3
Himalayan	1
Korat	2
LaPerm	1
Lykoi	1
Maine Coon	11
Maine Coon Cross	20
Napoleon	3
Ocicat	1

Breed	Cat/breed
Oriental	4
Oriental Shorthair	3
Oriental Shorthair cross	3
Persian	3
Persian cross	1
Persian x Bengal	2
Persian x Japanese Bobtail	1
Peterbald	1
Ragdoll	4
Random bred	55
Savannah	1
Selkirk Rex	1
Siamese	7
Siamese cross	1
Siberian	2
Sphynx	3
Tennessee Rex	1
Tonkinese	1
Toybob	2
Toyger	1
Turkish Angora	1
Unknown	4
Total	197

Supplementary Table 2. List of the variants detected in the two affected Abyssinians compared to the Control Population with the Golden Helix SVS Software. Variants are referred to the *Felis catus* genome vs9.0.

	LOCATION	REF	ALT	VARIANT ANNOTATION	GENE	S1	S2	Other cats with the variant
HOMOZYGOUS VARIANTS	A1:111852820-SNV	C	G	3'UTR	SEPT8	X*	X	
	A1:119678149-SNV	C	G	INTRONIC	DIAPH1	X*	X	
	A1:16984630-SNV	A	T	INTRONIC	UFM1	X*	X	
	A1:177001232-SNV	G	A	INTERGENIC		X	X*	
	A1:177022342-SNV	G	A	INTERGENIC		X	X*	
	A2:64580052-SNV	G	A	INTRONIC	ABCA13	X	X*	
	A2:64590355-SNV	C	A	INTRONIC	ABCA13	X	X*	
	A2:130695581-Del	/	DEL T	INTERGENIC		X	X*	
	A2:130925698-SNV	C	T	INTERGENIC		X	X*	
	A2:130952930-SNV	G	A	INTERGENIC		X	X*	
	A2:155040067-SNV	A	G	INTERGENIC		X*	X	
	A2:155042376-SNV	A	G	INTERGENIC		X*	X	
	A2:155056520-SNV	G	A	INTERGENIC		X*	X	
	A2:155126405-Del	/	DEL T	INTRONIC	TRIM24	X*	X	
	A2:155674197-SNV	C	T	INTRONIC	ZC3HAV1	X*	X*	
	A3:10826792-Ins	/	INS T	INTRONIC	NFATC2	X*	X	
	A3:73255707-SNV	A	T	INTERGENIC		X	X*	
	B2:16064796-Del	/	DEL G	INTRONIC	PHACTR1	X*	X	
	B2:19703579-SNV	G	A	INTERGENIC		X*	X	
	B2:19736178-SNV	A	G	INTERGENIC		X*	X	
	B2:42707393-SNV	C	T	INTERGENIC		X*	X*	
	B2:8777038-SNV	G	A	INTRONIC	CDKAL1	X*	X*	Napoleon1 (Heterozygous)
	B2:89190917-SNV	G	A	INTRONIC	FUT9	X	X*	
	B2:90834045-SNV	G	A	INTERGENIC		X	X*	
	B3:88127652-SNV	C	T	INTRONIC	SLC25A21	X	X*	
	C1:91241195-SNV	G	T	INTRONIC	ENSFCAG00000045482	X	X*	
	C1:92259380-SNV	C	T	INTRONIC	VAV3	X	X*	
	C1:117438182-SNV	G	A	INTERGENIC		X	X*	
	C1:120549939-Del	/	DEL A	3'UTR	EPB41L5	X	X*	BengalXSomali (Heterozygous)
	C1:135680694-SNV	A	G	INTRONIC	LRP1B	X*	X	
	C1:135938363-SNV	G	A	INTERGENIC		X*	X	
	C1:135939497-SNV	C	T	INTERGENIC		X*	X	
	C1:136047575-SNV	A	G	INTERGENIC		X*	X	
	C1:136048770-SNV	A	G	INTERGENIC		X*	X	
	D1:71376338-SNV	G	A	INTRONIC	ENSFCAG00000045288	X	X*	Random Bred1 (Heterozygous)
	D1:78680550-SNV	C	T	INTRONIC	NELL1	X	X*	

HETEROZYGOUS VARIANTS	D1:110195591-SNV	C	A	INTERGENIC		X*	X	Abyssinian1 (Heterozygous)
	D3:3130253-SNV	A	G	INTRONIC	GLT1D1	X	X*	Maine Coon cross1 (Heterozygous)
	D4:23796830-SNV	A	G	INTERGENIC		X*	X	
	D4:23927917-SNV	T	C	INTERGENIC		X*	X	
	E1:16576203-SNV	G	A	INTERGENIC		X	X*	
	E1:51539347-SNV	C	T	INTERGENIC		X*	X	Napoleon1 (Heterozygous)
	E2:17348953-SNV	A	G	INTRONIC	ZNF461	X	X*	Abyssinian2 (Heterozygous)
	E3:2906924-SNV	G	A	INTERGENIC		X	X*	
	E3:1146848-SNV	C	T	INTERGENIC		X	X*	Domestic Longhair (Heterozygous)
	E3:15242919-SNV	A	G	INTRONIC	ENSFCAG00000008708	X*	X	
	E3:27789066-SNV	A	C	INTRONIC	TMEM159	X*	X	
	F1:36054657-SNV	C	T	INTERGENIC		X*	X	
	F1:37513834-SNV	C	T	INTERGENIC		X*	X	
	F1:37695845-SNV	G	A	INTERGENIC		X*	X	
	F2:5978362-SNV	A	G	INTERGENIC		X*	X	
	F2:69511589-SNV	T	G	INTRONIC	TMEM65	X*	X*	Abyssinian1 (Heterozygous)
	F2:75400451-SNV	A	G	INTERGENIC		X*	X*	Abyssinian1 (Heterozygous)
	A1:238324409-SNV	C	G	INTRONIC	ADAMTS16	X	X	
	A1:241637222-SNV	G	A	INTRONIC	TPPP	X	X	
	A2:146503307-SNV	G	A	INTRONIC	ENSFCAG00000042722	X	X	
	A2:159295682-SNV	G	A	INTRONIC	TCAF1	X	X	
	A3:54253063-SNV	G	C	INTRONIC	CCDC138	X	X	
	A3:54313901-Ins	/	INS AAA	INTRONIC	CCDC138	X	X	
	B1:72770115-SNV	A	G	INTRONIC	GLRB	X	X	
	B1:72839639-SNV	C	G	INTRONIC	GLRB	X	X	
	B1:72847525-SNV	C	A	INTRONIC	GLRB	X	X	
	B1:72948314-SNV	T	C	INTRONIC	PDGFC	X	X	
	B1:176551596-Ins	/	INS TTG	INTRONIC	RELL1	X	X	
	B1:176551597-Ins	/	INS ACTCTGTGCTTTGAACAC	INTRONIC	RELL1	X	X	
	B2:138716854-SNV	T	C	INTRONIC	PLEKHG1	X	X	
	B3:18756555-SNV	G	A	INTRONIC	ADAMTS17	X	X	
	B3:138448681-SNV	C	T	INTRONIC	ITPK1	X	X	
	B3:138578082-SNV	A	T	INTRONIC	ENSFCAG00000022177	X	X	
	C1:216086348-SNV	T	A	INTRONIC	DGKD	X	X	
	C1:220145561-SNV	G	A	INTRONIC	ASB1	X	X	
	D2:31721229-SNV	G	T	INTRONIC	TET1	X	X	
	D2:41954837-SNV	C	T	INTRONIC	TSPAN14	X	X	
	D2:42026387-SNV	G	C	INTRONIC	SH2D4B	X	X	
	D2:84594662-SNV	G	A	INTRONIC	DOCK1	X	X	
	D2:84641559-SNV	T	C	INTRONIC	DOCK1	X	X	
	D3:71323080-SNV	C	T	INTRONIC	DYM	X	X	
	D3:71337519-SNV	C	T	INTRONIC	DYM	X	X	

	D4:24894003-SNV	A	G	INTRONIC	GKAP1	X	X	
	D4:24963097-SNV	A	T	INTRONIC	KIF27	X	X	
	D4:25018898-SNV	G	A	INTRONIC	KIF27	X	X	
	D4:25031984-SNV	C	A	INTRONIC	KIF27	X	X	
	D4:26970845-SNV	A	G	INTRONIC	ISCA1	X	X	
	D4:26991696-SNV	G	A	INTRONIC	TUT7	X	X	
	D4:96337354-SNV	C	T	EXONIC	CACNA1B (Synonymous mutation)	X	X	
	E1:51790720-SNV	C	T	3'UTR	ABCA8	X	X	
	E3:34556986-SNV	A	C	INTRONIC	SNX29	X	X	
	F1:49259649-SNV	C	A	INTRONIC	ENSCAG00000041780	X	X	
	F1:65085198-SNV	T	C	INTRONIC	NOS1AP	X	X	
HETEROZYGOUS INTERGENIC VARIANTS	A1:3721622-Del	/	DEL GAG	INTERGENIC		X	X	
	A1:3723336-SNV	A	G	INTERGENIC		X	X	
	A1:12473207-SNV	A	T	INTERGENIC		X	X	
	A1:12489481-Del	/	DEL ATC	INTERGENIC		X	X	
	A1:13704950-SNV	C	T	INTERGENIC		X	X	
	A1:103510608-SNV	A	C	INTERGENIC		X	X	
	A1:103549403-SNV	C	G	INTERGENIC		X	X	
	A1:149297420-SNV	C	T	INTERGENIC		X	X	
	A1:149309231-Del	/	DEL G	INTERGENIC		X	X	
	A2:36574081-SNV	C	T	INTERGENIC		X	X	
	A2:130475206-SNV	A	G	INTERGENIC		X	X	
	A2:163583037-Del	/	DEL CT	INTERGENIC		X	X	
	A2:163599262-Del	/	DEL T	INTERGENIC		X	X	
	A2:163603433-SNV	G	T	INTERGENIC		X	X	
	A3:2762833-SNV	G	C	INTERGENIC		X	X	
	A3:46725435-SNV	A	T	INTERGENIC		X	X	
	A3:53199582-SNV	A	T	INTERGENIC		X	X	
	A3:54647107-SNV	G	A	INTERGENIC		X	X	
	A3:54649029-SNV	G	A	INTERGENIC		X	X	
	A3:55724505-SNV	G	T	INTERGENIC		X	X	
	A3:91142621-Del	/	DEL TTAT	INTERGENIC		X	X	
	B1:73715398-SNV	G	A	INTERGENIC		X	X	
	B1:73716499-SNV	C	T	INTERGENIC		X	X	
	B1:73732838-SNV	C	G	INTERGENIC		X	X	
	B1:175653864-SNV	A	G	INTERGENIC		X	X	
	B1:175662987-SNV	C	T	INTERGENIC		X	X	
	B1:175965365-SNV	G	A	INTERGENIC		X	X	
	B1:176097479-Del	/	DEL CAAAA	INTERGENIC		X	X	
	B2:91675239-SNV	G	A	INTERGENIC		X	X	
	B3:112248722-SNV	T	C	INTERGENIC		X	X	
	C1:71596955-SNV	A	G	INTERGENIC		X	X	

C1:127649275-SNV	C	T	INTERGENIC		X	X	
C1:218597807-SNV	C	T	INTERGENIC		X	X	
D1:34379693-SNV	G	A	INTERGENIC		X	X	
D1:116903406-Del	/	DEL AAAAAAC	INTERGENIC		X	X	
D3:49498304-SNV	G	A	INTERGENIC		X	X	
D3:49638483-SNV	C	T	INTERGENIC		X	X	
D4:23196160-SNV	G	A	INTERGENIC		X	X	
D4:24821535-SNV	G	C	INTERGENIC		X	X	
D4:25071395-SNV	C	G	INTERGENIC		X	X	
D4:25217008-Del	/	DEL TTGC	INTERGENIC		X	X	
D4:25239416-SNV	G	A	INTERGENIC		X	X	
D4:25242963-SNV	A	G	INTERGENIC		X	X	
D4:25288919-Del	/	DEL G	INTERGENIC		X	X	
D4:25293630-SNV	C	T	INTERGENIC		X	X	
E1:52514780-SNV	C	T	INTERGENIC		X	X	
E1:52628135-SNV	A	G	INTERGENIC		X	X	
E3:23032187-SNV	C	T	INTERGENIC		X	X	
E3:23103793-SNV	C	G	INTERGENIC		X	X	
F2:50114460-SNV	A	G	INTERGENIC		X	X	

*Cat homozygous for the variant

INS = Insertion

DEL = Deletion

Supplementary Table 3 (data not shown). Proteins up regulated or exclusively expressed in cats affected by renal amyloidosis (AMY). Proteins were considered differentially expressed if they were present only in one condition or showed statistically significant difference (t-test p value ≤ 0.05).

Supplementary Table 4 (data not shown). Proteins down regulated in cats affected by renal amyloidosis (AMY) or exclusively expressed in healthy cats (CTR). Proteins were considered differentially expressed if they were present only in one condition or showed statistically significant difference (t-test p value ≤ 0.05).

Supplementary Table 5. Functional grouping analysis performed by David software on proteins up regulated and expressed in cats affected by renal amyloidosis.

Clustering	Term	Count	PValue	Genes
Extracellular space region and matrix	Extracellular space	19	9,8E-03	AEBP1, DDB1, GUSB, CES5A, ECM1, PRDX1, TTR, LAMP2, CKM, SERPINF1, SERPINF2, LOC101097240, LOC101099498, PON1, LTF, SERPIND1, CFI, CTSB, MUC5B
	Extracellular matrix	5	1,9E-02	AEBP1, SERPINF1, MGP, ECM1, AHSG
	Collagen fibril organization	3	5,0E-02	SERPINF2, PLOD3, SERPINH1
Metabolism	Metabolic pathways	55	3,0E-12	SGSH, LDHA, HEXA, PGD, OGDHL, PPCS, CAD, HIBADH, RGN, RPN2, NDUFS3, HPD, PDXK, PFKL, ACO2, ACADS, ATP6V1H, NAPRT, PGM2, PGLS, G6PD, CKM, H6PD, LOC101083433, HAO2, MGAM, OAT, UGP2, MDH2, MPST, XYLB, ME1, LOC101081808, NDUFB4, GANAB, LOC101098145, LOC101094311, ATP6V1G1, ALDH3A2, ACSBG1, GMPPB, LOC101083474, CSAD, ARG2, TSTA3, FH, BCKDHA, GCDH, GUSB, MAOA, CS, EPHX2, ACACB, PCK2, ALDH2
	Fatty acid degradation	6	6,8E-04	GCDH, ACADS, ALDH2, CPT1A, ALDH3A2, ACSBG1
	Fatty acid beta-oxidation using acyl-CoA dehydrogenase	4	1,1E-03	GCDH, ACADS, ETFB, ETFA
	Fatty-acyl-CoA binding	3	3,9E-02	GCDH, ACADS, ETFA
	Citrate cycle (TCA cycle)	6	1,6E-04	ACO2, OGDHL, CS, PCK2, MDH2, FH
	Tricarboxylic acid cycle	5	1,8E-04	ACO2, OGDHL, CS, MDH2, FH
	Pyruvate metabolism	8	4,7E-06	ME1, LDHA, ALDH2, ACACB, PCK2, MDH2, ALDH3A2, FH
	Carbon metabolism	14	6,9E-08	ME1, ACO2, PFKL, ACADS, OGDHL, CS, PGD, PGLS, G6PD, H6PD, HAO2, RGN, MDH2, FH
	Pentose phosphate pathway	7	7,4E-06	PGM2, PGLS, G6PD, PFKL, H6PD, PGD, RGN
	Carbohydrate metabolic process	8	6,6E-05	LDHA, GANAB, GUSB, HEXA, MGAM, CS, SHPK, MDH2
	Mitochondrial inner membrane	10	1,3E-04	SLC25A13, GSTK1, LOC101094311, SLC25A3, VDAC2, ABCB6, CPT1A, MDH2, ALDH3A2, SOD2
	Mitochondrion	22	6,0E-04	ME1, GCDH, GPD1, ECH1, ACO2, NUDT13, ACADS, OGDHL, PRDX4, LOC101098145, ACACB, PCK2, PRDX1, BBOX1, GMPPB, TPP1, ARG2, HAO2, CTU2, ALDH2, OAT, FH
	Amino sugar and nucleotide sugar metabolism	6	1,8E-03	GMPPB, PGM2, HEXA, TSTA3, NAGK, UGP2
	Lysine degradation	6	2,5E-03	GCDH, PLOD3, OGDHL, ALDH2, BBOX1, ALDH3A2
	Arginine and proline metabolism	6	1,9E-03	CKM, MAOA, ARG2, ALDH2, OAT, ALDH3A2
	Acetylation	5	2,1E-03	RPS3A, MAOA, AARS, EIF3M, RPS7
	Malate metabolic process	3	3,5E-03	ME1, MDH2, FH

	Glycolysis / Gluconeogenesis	6	4,4E-03	PGM2, LDHA, PFKL, ALDH2, PCK2, ALDH3A2
	Acyl-CoA dehydrogenase activity	3	8,5E-03	GCDH, ACADS, ETFA
	Tryptophan metabolism	5	5,5E-03	GCDH, MAOA, OGDHL, ALDH2, ALDH3A2
	Valine, leucine and isoleucine degradation	5	1,1E-02	BCKDHA, ACADS, ALDH2, HIBADH, ALDH3A2
	Glycoside hydrolase, superfamily	4	1,3E-02	GANAB, GUSB, HEXA, MGAM
	Glyoxylate and dicarboxylate metabolism	4	1,4E-02	ACO2, HAO2, CS, MDH2
	Galactose metabolism	4	1,6E-02	PGM2, PFKL, MGAM, UGP2
	Ascorbate and aldarate metabolism	3	2,3E-02	ALDH2, RGN, ALDH3A2
Inhibition of serine protease activity and coagulation	Blood microparticle	10	1,8E-07	APOA4, LOC101083474, SERPINF2, PON1, JCHAIN, VTN, SLC4A1, CPN2, AHSB, C8G
	Serine-type endopeptidase inhibitor activity	5	8,9E-03	SERPINF1, SERPINF2, SERPIND1, ITIH3, SERPINH1
	Protease inhibitor I4, serpin, conserved site	4	4,8E-03	SERPINF1, SERPINF2, SERPIND1, SERPINH1
	Complement and coagulation cascades	7	2,1E-03	F13B, C8B, CFB, SERPINF2, CFI, SERPIND1, C8G
	SERPIN	4	2,8E-03	SERPINF1, SERPINF2, SERPIND1, SERPINH1
	Serpin family	4	5,3E-03	SERPINF1, SERPINF2, SERPIND1, SERPINH1
	Serpin domain	4	5,8E-03	SERPINF1, SERPINF2, SERPIND1, SERPINH1
Redox balance	Oxidoreductase	14	2,1E-07	ME1, GCDH, GPD1, LDHA, ACADS, MAOA, PGD, LOC101094311, HIBADH, ALDH3A2, SOD2, ALDH2, NDUFS3, MDH2
	NAD	5	4,8E-03	GPD1, LDHA, NDUFS3, MDH2, HIBADH
	Dehydrogenase, multihelical	3	5,1E-03	GPD1, PGD, HIBADH
	NAD(P)-binding domain	10	1,8E-04	ME1, GPD1, LDHA, H6PD, PGD, LOC101098145, TSTA3, DDO, MDH2, HIBADH
	Peroxisome	10	2,5E-05	PAOX, ECH1, NUDT13, GSTK1, HAO2, LOC101098145, EPHX2, DDO, PRDX1, SOD2
	Cell redox homeostasis	4	4,2E-02	TXNDC5, PRDX4, PDIA6, PRDX1
	Electron carrier activity	4	2,2E-02	GCDH, ACADS, ETFB, ETFA
	Peroxisome	4	2,6E-02	GSTK1, HAO2, LOC101098145, EPHX2
RNA translation	Structural constituent of ribosome	11	2,5E-04	RPS26, RPS19, SLC25A13, LOC101092708, RPS3A, RPL9, SLC25A3, RPLP2, RPL37A, RPS20, RPS7
	Ribosome	8	1,1E-02	RPS26, RPS19, RPS3A, RPL9, RPLP2, RPL37A, RPS20, RPS7
	Translation	9	1,2E-03	RPS26, RPS19, SLC25A13, LOC101092708, RPS3A, SLC25A3, RPL37A, RPS20, RPS7

	poly(A) RNA binding	21	1,8E-02	RTN4, LOC101082749, GANAB, FSCN1, CS, SYNCRIP, SERPINH1, PRDX1, LOC101089263, APEH, CORO1A, LRP1, RPS19, SRSF7, RPS3A, SNRNPB, MAP4, RPL37A, RPS20, KPNB1, MDH2
	Cytosolic small ribosomal subunit	5	6,9E-04	RPS26, RPS19, RPS3A, RPS20, RPS7
	Cytosolic large ribosomal subunit	4	2,5E-02	LOC101092708, RPL9, RPLP2, RPL37A
Other	Signal	47	1,2E-06	RAB7A, GANAB, HEXA, SUSP2, PRDX4, JCHAIN, PTK7, PDIA6, VTN, SERPINH1, CPN2, AHSG, F13B, APOA4, TTR, TPP1, PLOD3, LTF, ITIH3, RPN2, NDUFS3, GPNMB, MATN2, EFEMP2, GUSB, CDHR2, ERP29, EPHX2, CES5A, MGP, NID2, ACACB, MCAM, ECM1, C8G, C8B, LAMP2, LRP1, SERPINF1, H6PD, SERPINF2, GLMP, LOC101099498, ARSA, SERPIND1, ENG, MFAP5
	Myelin sheath	10	3,9E-05	LOC101081808, ACO2, FSCN1, LOC101094311, SLC25A3, NDUFS3, VDACC2, PRDX1, MDH2, SOD2
	Epithelial cell differentiation	6	3,9E-04	PTER, TPP1, GSTK1, CDHR2, ANXA4, CPT1A
	Six-bladed beta-propeller, TolB-like	5	9,5E-04	APEH, LRP1, PON1, RGN, NID2
	Muscle contraction	4	4,3E-03	MYBPC1, CALD1, MYL1, ACTN2
	Focal adhesion	11	3,3E-03	RPS19, LRP1, RPS3A, RPL9, YWHAQ, PTK7, RPLP2, RPL37A, ACTN2, MCAM, ENG
	Lysosome	8	5,9E-03	SGSH, LAMP2, TPP1, GUSB, HEXA, ARSA, ATP6V1H, CTSH
	Membrane	20	6,1E-03	ECH1, GANAB, PFKL, HEXA, GUSB, AARS, ERP29, SYNCRIP, RPLP2, CAD, RPS26, LAMP2, RPS19, RPL9, PSMD1, PSMB2, YWHAQ, RPS20, CFI, MVP
	Thioredoxin domain	4	8,9E-03	TXNDC5, PRDX4, PDIA6, PRDX1
	Thioredoxin-like fold	6	9,1E-03	GSTK1, TXNDC5, ERP29, PRDX4, PDIA6, PRDX1
	Cytosol	20	9,6E-03	RAB7A, PDXK, AARS, EPHX2, PRDX4, CAD, ATP6V1G1, NAPRT, PGM2, CORO1A, LOC101083433, RPS3A, CTU2, GLMP, RAB14, SNRNPB, HSPA4L, CA3, CTSH, FH
	6-phosphogluconate dehydrogenase, C-terminal-like	3	1,1E-02	GPD1, PGD, HIBADH
	Lysosome	6	1,6E-02	RAB7A, TPP1, GUSB, HEXA, GLMP, CTSH
	Cytoplasm	8	2,2E-02	PFKL, RPS3A, FSCN1, AARS, CTU2, MAP4, EIF3M, RPS7
	Retina homeostasis	3	2,3E-02	JCHAIN, LTF, PRDX1
	Complement C1r-like EGF domain	3	2,5E-02	LRP1, EFEMP2, NID2
	Protein processing in endoplasmic reticulum	8	2,9E-02	GANAB, TXNDC5, ERP29, HSPA4L, PDIA6, SEC13, RPN2, HSPA8
	Lysosomal membrane	6	3,3E-02	LAMP2, LRP1, RAB14, ATP6V1H, MYLPP, ATP6V1G1
	Calcium ion binding	14	3,4E-02	MATN2, EFEMP2, MYL1, CDHR2, S100A11, MGP, MYLPP, ACTN2, NID2, ANXA4, LRP1, SLC25A13, PON1, RGN
	Pentose and glucuronate interconversions	3	3,4E-02	GUSB, UGP2, XYLB

	Lipid homeostasis	3	4,3E-02	GCDH, ACADS, ETFA
	Amoebiasis	6	4,5E-02	C8B, RAB7A, COL4A1, ARG2, ACTN2, C8G
	Glycosaminoglycan degradation	3	4,7E-02	SGSH, GUSB, HEXA

Supplementary Table 6. Gene Ontology Biological process, Molecular Function, Cellular Component and Pathway analyses by Panther software on proteins up regulated and expressed in cats affected by renal amyloidosis. Enrichment scores, nominal P-values and False Discovery Rate (FDR) are reported for each GO term.

GO biological process complete			
	Fold Enrichment	Raw P-value	FDR
negative regulation of peptidase activity (GO:0010466)	6.15	7.06E-05	2.04E-02
translation (GO:0006412)	3.54	6.02E-05	1.82E-02
peptide biosynthetic process (GO:0043043)	3.46	7.72E-05	2.18E-02
fatty acid beta-oxidation using acyl-CoA dehydrogenase (GO:0033539)	69.19	1.62E-06	8.26E-04
NADPH regeneration (GO:0006740)	64.87	9.10E-08	6.72E-05
pentose-phosphate shunt (GO:0006098)	64.87	9.10E-08	6.37E-05
glucose 6-phosphate metabolic process (GO:0051156)	39.92	5.83E-07	3.69E-04
NADP metabolic process (GO:0006739)	32.43	1.35E-06	7.49E-04
tricarboxylic acid cycle (GO:0006099)	20.76	8.83E-06	3.67E-03
aerobic respiration (GO:0009060)	13.31	6.04E-05	1.78E-02
drug catabolic process (GO:0042737)	11.53	1.12E-04	3.05E-02
glucose metabolic process (GO:0006006)	10.81	1.49E-04	3.88E-02
monosaccharide metabolic process (GO:0005996)	8.86	2.30E-05	8.06E-03
cellular respiration (GO:0045333)	8.35	3.29E-05	1.07E-02
organic acid catabolic process (GO:0016054)	8.12	3.01E-06	1.48E-03
carboxylic acid catabolic process (GO:0046395)	8.12	3.01E-06	1.43E-03
drug metabolic process (GO:0017144)	7.53	1.27E-08	1.20E-05
carbohydrate metabolic process (GO:0005975)	7.39	1.16E-11	1.71E-08
generation of precursor metabolites and energy (GO:0006091)	6.89	3.69E-09	3.78E-06
carboxylic acid metabolic process (GO:0019752)	6.35	2.33E-14	7.74E-11
oxoacid metabolic process (GO:0043436)	6.02	8.06E-14	2.14E-10
organic acid metabolic process (GO:0006082)	5.85	1.57E-13	3.47E-10
small molecule catabolic process (GO:0044282)	5.77	4.02E-05	1.27E-02
cellular amino acid metabolic process (GO:0006520)	5.52	2.14E-05	7.68E-03
small molecule metabolic process (GO:0044281)	5.42	1.34E-21	1.78E-17
oxidation-reduction process (GO:0055114)	5.33	2.72E-17	1.81E-13
cofactor metabolic process (GO:0051186)	4.89	1.34E-04	3.57E-02

carbohydrate derivative metabolic process (GO:1901135)	4.24	2.18E-08	1.81E-05
nucleobase-containing small molecule metabolic process (GO:0055086)	4.11	5.30E-05	1.64E-02
peptide metabolic process (GO:0006518)	4.06	8.23E-07	4.97E-04
organophosphate metabolic process (GO:0019637)	3.91	1.39E-06	7.42E-04
cellular amide metabolic process (GO:0043603)	3.90	9.23E-08	6.14E-05
amide biosynthetic process (GO:0043604)	3.45	2.42E-05	8.24E-03
organonitrogen compound biosynthetic process (GO:1901566)	3.29	1.45E-08	1.28E-05
cellular catabolic process (GO:0044248)	2.95	8.14E-06	3.49E-03
cellular nitrogen compound biosynthetic process (GO:0044271)	2.85	1.26E-06	7.28E-04
catabolic process (GO:0009056)	2.84	3.48E-06	1.59E-03
organic substance catabolic process (GO:1901575)	2.82	1.61E-05	5.96E-03
organic substance biosynthetic process (GO:1901576)	2.27	6.36E-06	2.82E-03
biosynthetic process (GO:0009058)	2.20	1.04E-05	4.08E-03
cellular biosynthetic process (GO:0044249)	2.09	1.10E-04	3.04E-02
cellular nitrogen compound metabolic process (GO:0034641)	2.04	1.34E-05	5.07E-03
metabolic process (GO:0008152)	2.00	2.26E-16	1.00E-12
cellular metabolic process (GO:0044237)	1.97	2.85E-12	4.74E-09
organic substance metabolic process (GO:0071704)	1.95	4.19E-13	7.96E-10
organonitrogen compound metabolic process (GO:1901564)	1.95	3.83E-08	3.00E-05
primary metabolic process (GO:0044238)	1.89	5.24E-11	6.96E-08
nitrogen compound metabolic process (GO:0006807)	1.63	9.71E-06	3.91E-03
cellular process (GO:0009987)	1.32	2.80E-05	9.29E-03
biological process (GO:0008150)	1.31	1.31E-09	1.45E-06
Unclassified (UNCLASSIFIED)	.45	1.31E-09	1.58E-06
GO molecular function complete			
oxidoreductase activity, acting on the aldehyde or oxo group of donors (GO:0016903)	13.31	6.04E-05	1.59E-02
oxidoreductase activity, acting on CH-OH group of donors (GO:0016614)	11.41	4.51E-08	2.38E-05
oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor (GO:0016616)	10.99	2.89E-07	1.33E-04
oxidoreductase activity (GO:0016491)	5.37	8.44E-15	7.79E-12
antioxidant activity (GO:0016209)	9.44	2.68E-04	4.50E-02
coenzyme binding (GO:0050662)	7.74	2.22E-10	1.64E-07
exopeptidase activity (GO:0008238)	7.33	2.38E-04	4.39E-02
peptidase regulator activity (GO:0061134)	5.91	2.54E-04	4.47E-02
carbohydrate binding (GO:0030246)	5.57	1.36E-04	2.94E-02
cofactor binding (GO:0048037)	5.19	4.08E-09	2.51E-06
structural constituent of ribosome (GO:0003735)	4.58	1.93E-05	5.94E-03
structural molecule activity (GO:0005198)	3.18	6.36E-05	1.57E-02
identical protein binding (GO:0042802)	2.67	6.32E-06	2.34E-03
catalytic activity (GO:0003824)	2.22	6.31E-17	7.76E-14
hydrolase activity (GO:0016787)	1.99	5.07E-05	1.44E-02
cation binding (GO:0043169)	1.83	1.23E-04	2.83E-02
metal ion binding (GO:0046872)	1.78	2.90E-04	4.65E-02

ion binding (GO:0043167)	1.67	1.38E-05	4.63E-03
molecular_function (GO:0003674)	1.45	1.24E-17	2.29E-14
binding (GO:0005488)	1.42	6.70E-07	2.75E-04
Unclassified (UNCLASSIFIED)	.30	1.24E-17	4.57E-14
transmembrane signaling receptor activity (GO:0004888)	.09	1.80E-04	3.70E-02
G protein-coupled receptor activity (GO:0004930)	< 0.01	1.98E-04	3.84E-02
GO cellular component complete			
ribosome (GO:0005840)	4.21	4.27E-05	1.03E-02
ribonucleoprotein complex (GO:1990904)	2.67	2.85E-04	4.81E-02
extracellular space (GO:0005615)	4.00	5.97E-08	2.53E-05
extracellular region part (GO:0044421)	3.63	2.25E-08	1.27E-05
extracellular region (GO:0005576)	2.77	8.98E-07	2.53E-04
mitochondrion (GO:0005739)	3.10	2.65E-07	8.96E-05
cytoplasmic part (GO:0044444)	1.93	1.88E-11	3.18E-08
cytoplasm (GO:0005737)	1.78	4.26E-11	3.60E-08
intracellular (GO:0005622)	1.35	7.75E-05	1.64E-02
intracellular part (GO:0044424)	1.35	7.75E-05	1.46E-02
GO Pathways			
TCA cycle (P00051)	34.77	1.74E-04	1.43E-02
Blood coagulation (P00011)	9.87	1.20E-05	1.97E-03

Supplementary Table 7 (data not shown). Output of miRDeep2 showing the 854 identified miRNAs divided into miRNAs kept after filtering (<10reads/sample; n258) and filtered out miRNAs (>10reads/sample; n596). For each feline miRNA, the predicted start and precursor sequences are reported, as well as the read count. The human miRNA ID is reported only for those feline miRNAs with the same mature sequence of the human ones.

Supplementary Table 8. Output of the “NC” miRNA pre-sequences alignment using the BLASTN function of miRBase. Only the 10 miRNAs that matched those in humans or in other species are reported. For each alignment, the reliability of the match with the other species is shown through the E-value and the bit-score. Bit-score $\geq$ or = 100; E-value < 0.01.

miRDeep2 provisional ID	consensus precursor sequence	precursor sequence length (bp)	Homo sapiens				Canis familiaris			
			bp alignment	bit-score	E-value	miRNA_ID	bp alignment	bit-score	E-value	miRNA_ID
NC_018730.3_21320	caguuaucacagug cugaugcuguccau ucuaaagguacagu acugugauaacuga	56	1-22;	110	9,00E-04	hsa-miR-101-5p	37-56	100	0.006	cfa-miR-101
			37-56	100	0.006	hsa-miR-101-3p				
NC_018737.3_34091	cagugguuuuaccc uaugguagguuacg ucaugcuguucua cacaggguaagaacc acggac	62	1-22;	110	0.001	hsa-miR-140-5p	41-61	105	0.003	cfa-miR-140
			40-60	105	0.003	hsa-miR-140-3p				
NC_018725.3_6039	aggacuguccaacc ugagaauaggugucu ccggggucuaucuc agguucgucagccc gug	59	–	–	–	–	–	–	–	–
NC_018724.3_5609	auggcacugguaga auucacugugaaca gucucggucaguga auuaccgaagggcc auaaa	61	39-60;	110	0.001	hsa-miR-183-3p	1-21	105	0.003	cfa-miR-183
			1-21	105	0.003	hsa-miR-183-5p				
NC_018741.3_39910	uuauaaauacaaccu gauaaguguuacag cacuaccagguug uauuguaauu	52	1-22	110	8,00E-04	hsa-miR-374a-5p	1-21	105	0.002	cfa-miR-374a
NC_018728.3_14704	auaguagaccguau agcguacgcuuuau cugugacguaugua acacgguccacuaa c	57	2-22;	105	0.002	hsa-miR-411-5p	1-22	110	0.001	cfa-miR-411
			37-57	105	0.002	hsa-miR-411-3p				
NC_018725.3_7102	agcugguaaaaugg aaccaaaucaacug uucaauggauuugg ucccuucaaccagc ugu	60	38-59	110	0.001	hsa-miR-133a-3p	39-60	110	0.001	cfa-miR-133a
NC_018731.3_23601	gagagaucagaggc gcagagugcgucag ugucaaugaagccu gugccuuuaccuc uuuaa	61	1-22;	110	0.001	hsa-miR-6529-5p	40-61	110	0.001	cfa-miR-6529
			40-61	101	0.006	hsa-miR-6529-3p				
NC_018725.3_7074	cgggcugacgaacc ugagauugaccccg gagacaccauucuc agguuggacagucc ugga	60	–	–	–	–	–	–	–	–
NC_018735.3_30916	ucgguuaucauggu accgaugcuguau ucugaagguuacag uacugugauaacug a	57	1-22	110	0.001	hsa-miR-101-2-5p	38-57	100	0.006	cfa-miR-101
			38-57	100	0.006	hsa-miR-101-3p				

			Bos taurus				Mus musculus			
			bp alignment	bit-score	E-value	miRNA_ID	bp alignment	bit-score	E-value	miRNA_ID
NC_018730.3_21320	caguuaucacagug cugaugcguccau ucuaaagguacagu acugugauaacuga	56	37-56	100	0.006	bta-miR-101	1-21;	105	0.002	mmu-miR-101a-5p
							37-56	100	0.006	mmu-miR-101a-3p
NC_018737.3_34091	cagugguuuuaccc uauuguagguuacg ucaugcguuucac cacagguagaacc acggac	62	40-61	110	0.001	bta-miR-140	1-22;	110	0.001	mmu-miR-140-5p
							40-60	105	0.003	mmu-miR-140-3p
NC_018725.3_6039	aggacuguccaacc ugagaauugugucu ccggggucaaucuc agguucgucagccc gug	59	1-21	105	0.003	bta-miR-1388-5p	–	–	–	–
NC_018724.3_5609	auggcacugguaga auucacugugaaca gucucggucaguga auuaccgaagggcc auaaa	61	1-22	110	0.001	bta-miR-183	39-60;	110	0.001	mmu-miR-183-3p
							1-21	105	0.003	mmu-miR-183-5p
NC_018741.3_39910	uuauaaauacaaccu gauaaguguuacag cacuaccagguug uauuguaauu	52	1-22	110	8,00E-04	bta-miR-374a	–	–	–	–
NC_018728.3_14704	auaguagaccguau agcguacgcuuuau cugugacguauugua acacgguccacuaa c	57	1-22	110	0.001	bta-miR-411a	37-57	105	0.002	mmu-miR-411-3p
							2-22	105	0.002	mmu-miR-411-5p
NC_018725.3_7102	agcugguaaaaugg aaccaaaucacug uucaauggauuugg ucccuucaaccagc ugu	60	38-59	110	0.001	bta-miR-133a	–	–	–	–
NC_018731.3_23601	gagagaucagaggc gcagagugcgucag ugucaaugaagccu gugccuuuaccuc uuuaa	61	1-21;	105	0.003	bta-miR-6529a	–	–	–	–
			2-21	100	0.007	bta-miR-6529b				
NC_018725.3_7074	cgggcugacgaacc ugagauugaccccg gagacaccauucuc agguuggacagucc ugga	60	37-57	105	0.003	bta-miR-1388-5p	–	–	–	–
NC_018735.3_30916	ucgguuaucauggu accgaugcuguau ucugaaagguacag uacugugauaacug a	57	38-57	100	0.006	bta-miR-101	1-23	115	4,00E-04	mmu-miR-101b-5p
							38-57	100	0.006	mmu-miR-101a-3p

			Suis scrofa			
			bp alignment	bit-score	E-value	miRNA_ID
NC_018730.3_21320	caguuauacacagug cugaugcuguccau ucuaaagguacagu acugugauaacuga	56	37-56	100	0.006	ssc-miR-101
NC_018737.3_34091	cagugguuuuaccc uaugguagguuacg ucaugcuguucac cacagguagaacc acggac	62	40-62;	115	4,00E-04	ssc-miR-140-3p
			2-22	105	0.003	ssc-miR-140-5p
NC_018725.3_6039	aggacuguccaacc ugagaauaggugucu ccggggucuaucuc agguucgucagccc gug	59	1-22	101	0.005	ssc-miR-1388
NC_018724.3_5609	auggcacugguaga auucacugugaaca gucucggucaguga auuaccgaagggcc auaaa	61	1-22	110	0.001	ssc-miR-183
NC_018741.3_39910	uuauaauacaaccu gauaaguguuacag cacuaccagguug uauuguaauu	52	1-22	110	8,00E-04	ssc-miR-374a-5p
			31-52	101	0.004	ssc-miR-374a-3p
NC_018728.3_14704	auaguagaccguau agcguacgcuuuau cugugacguaugua acacgguccacuaa c	57	38-57	100	0.006	ssc-miR-411
NC_018725.3_7102	agcugguaaaaugg aaccaaaaucaacug uucaauggauuugg uccccuuaaccagc ugu	60	1-22	110	0.001	ssc-miR-133a-5p
NC_018731.3_23601	gagagaucaaggcc gcagagugcgucag ugucaaugaagccu gugccuuuaccuc uuuaa	61	1-21	105	0.003	ssc-miR-6529
NC_018725.3_7074	cgggcugacgaacc ugagauugaccccg gagacaccauucuc agguuggacagucc ugga	60	36-57	101	0.006	ssc-miR-1388
NC_018735.3_30916	ucgguuaucauggu accgaugcuguau ucugaaagguacag uacugugauaacug a	57	38-57	100	0.006	ssc-miR-101

Supplementary Table 9. List of the 60 miRNAs statistically differentially expressed, ordered by significance (P-value<0.05). The first 7 miRNAs were significant also after the Benjamin-Hochberg adjustment (Adjusted P-value<0.05).

miRNA_ID (miRDeep2)	P-Value	sign	Adj. P-Val	sign
NC_018730.3_21281	0.0001	***	0.0189	*
NC_018727.3_13107	0.0002	***	0.0189	*
hsa-miR-4451	0.0002	***	0.0189	*
NC_018731.3_22438	0.0005	***	0.0318	*
NC_018726.3_8619	0.0007	***	0.0318	*
hsa-miR-4654	0.0007	***	0.0318	*
hsa-miR-186-5p	0.0012	**	0.0426	*
NC_018736.3_31770	0.0018	**	0.0589	
hsa-miR-4733-3p	0.0022	**	0.0624	
hsa-miR-26a-5p	0.0028	**	0.0714	
hsa-miR-204-5p	0.0034	**	0.0757	
NC_018740.3_37965	0.0035	**	0.0757	
hsa-miR-153-3p	0.0045	**	0.0900	
hsa-miR-144-3p	0.0055	**	0.0916	
hsa-miR-4467	0.0057	**	0.0916	
hsa-miR-192-5p	0.0062	**	0.0916	
hsa-miR-1251-3p	0.0062	**	0.0916	
hsa-miR-15a-5p	0.0064	**	0.0916	
NC_018737.3_33714	0.0077	**	0.1013	
hsa-miR-3187-3p	0.0079	**	0.1013	
NC_018726.3_8928	0.0092	**	0.1133	
NC_018730.3_20699	0.0101	*	0.1185	
hsa-miR-4520-5p	0.0106	*	0.1185	
hsa-miR-6784-5p	0.0115	*	0.1194	
hsa-miR-361-5p	0.0116	*	0.1194	
hsa-miR-185-5p	0.0142	*	0.1359	
hsa-miR-30a-5p	0.0142	*	0.1359	
hsa-miR-142-5p	0.0159	*	0.1468	
hsa-miR-622	0.0176	*	0.1563	
NC_018723.3_2706	0.0184	*	0.1563	
hsa-miR-127-3p	0.0189	*	0.1563	
hsa-miR-24-3p	0.0199	*	0.1601	
NC_018736.3_32583	0.0231	*	0.1643	

hsa-miR-425-5p	0.0234	*	0.1643	
hsa-miR-4675	0.0237	*	0.1643	
hsa-miR-125b-5p	0.0237	*	0.1643	
hsa-miR-590-3p	0.0242	*	0.1643	
NC_018729.3_18944	0.0248	*	0.1643	
hsa-miR-219a-5p	0.0248	*	0.1643	
hsa-miR-520g-5p	0.0267	*	0.1678	
hsa-miR-5587-3p	0.0267	*	0.1678	
hsa-miR-6068	0.0274	*	0.1678	
hsa-miR-4758-3p	0.0280	*	0.1678	
NC_018726.3_9493	0.0292	*	0.1710	
hsa-miR-29a-3p	0.0300	*	0.1714	
hsa-miR-6859-3p	0.0306	*	0.1714	
hsa-miR-338-3p	0.0314	*	0.1723	
NC_018725.3_6091	0.0329	*	0.1744	
hsa-miR-374a-5p	0.0331	*	0.1744	
NC_018738.3_36088	0.0372	*	0.1884	
NC_018730.3_22053	0.0385	*	0.1884	
NC_018739.3_36383	0.0387	*	0.1884	
NC_018729.3_18908	0.0408	*	0.1884	
hsa-miR-23a-3p	0.0408	*	0.1884	
hsa-miR-486-5p	0.0415	*	0.1884	
NC_018726.3_8369	0.0416	*	0.1884	
hsa-miR-151a-3p	0.0417	*	0.1884	
hsa-miR-664a-3p	0.0430	*	0.1884	
NC_018732.3_25639	0.0431	*	0.1884	
hsa-miR-200b-3p	0.0499	*	0.2147	

Sign = significance

* $0.01 < P\text{-value} < 0.05$

** $P\text{-value} < 0.01$

*** $P\text{-value} < 0.00$

Supplementary Table 10. The table shows the genomic regions in which the miRNA mature sequences aligned on the cat reference genome (vs 9.0) and reports the miRNA genes already annotated in those regions. The last two columns show the comparison between miRNAs identified in the study and those already reported in literature.

miRNA ID (miRDeep2)	Mature sequence alignment on cat genome vs 9.0	miRNA gene	Already present in literature	
			Sun et al. 2014	Laganà et al. 2017
NC_018730.3_21281	chrC1:47329629-47329646			
NC_018727.3_13107	chrB2:152639385-152639403			
hsa-miR-4451	chrC2:71949543-71949560			
NC_018731.3_22438	chrC2:11221012-11221029			
NC_018726.3_8619	chrB1:93903253-93903270			
hsa-miR-4654	chrC1:42173644-42173662			
hsa-miR-186-5p	chrC1:59209948-59209969		mir-186	fca-mir-186
NC_018736.3_31770	chrE1:6112141-6112158			
hsa-miR-4733-3p	chrD4:29141906-29141924			
hsa-miR-26a-5p	chrC2:155146726-155146747	MIR26A1	mir-26a	fca-mir-26a-1
hsa-miR-204-5p	chrD4:13188378-13188399	MIR204		fca-mir-204
NC_018740.3_37965	chrF2:11905892-11905909			
hsa-miR-153-3p	chrC1:204283442-204283463	MIR153-1	mir-153	fca-mir-153-1/2
hsa-miR-144-3p	chrE1:17972486-17972505	MIR144	mir-144	fca-mir-144
hsa-miR-4467	chrB4:134142466-134142483			
hsa-miR-192-5p	chrD1:110099470-110099490	MIR192	mir-192	fca-mir-192
hsa-miR-1251-3p	chrB4:120706655-120706675			fca-mir-1251
hsa-miR-15a-5p	chrA1:21521395-21521415	MIR15A		fca-mir-15a
NC_018737.3_33714	chrE2:9470285-9470303			
hsa-miR-3187-3p	chrC1:23011165-23011182			
NC_018726.3_8928	No Alignment			
NC_018730.3_20699	chrC1:208919059-208919080			
hsa-miR-4520-5p	chrD2:76471955-76471979			
hsa-miR-6784-5p	chrA1:1464447-1464464			
hsa-miR-361-5p	chrX:73838108-73838129	MIR361	mir-361	fca-mir-361
hsa-miR-185-5p	chrD3:28342591-28342612	MIR185	mir-185	fca-mir-185
hsa-miR-30a-5p	chrC1:31853395-31853418	MIR30E	mir-30e	
hsa-miR-142-5p	chrE1:30526043-30526063	MIR142	mir-142	
hsa-miR-622	chrC2:83012233-83012250			
NC_018723.3_2706	chrA1:174205065-174205083			
hsa-miR-127-3p	chrB3:145457051-145457071	MIR127		
hsa-miR-24-3p	chrD4:29138418-29138440	MIR3074	mir-3074	fca-mir-24-1
NC_018736.3_32583	chrE1:E1:57707913-57707931			

hsa-miR-425-5p	chrA2:18165553-18165574	MIR425	mir-425	
hsa-miR-4675	chrX:70753372-70753390			
hsa-miR-125b-5p	chrD1:19302682-19302703	MIR125B1		fca-mir-125b-1
hsa-miR-590-3p	chrE3:10665608-10665629	MIR590	mir-590	fca-mir-590
NC_018729.3_18944	chrB4:139786000-139786017			
hsa-miR-219a-5p	chrB2:34223801-34223823/chrB2:34841076-34841098		mir-506	fca-mir-506
hsa-miR-520g-5p	chrA1:29270363-29270380			
hsa-miR-5587-3p	chrB4:3165764-3165783			
hsa-miR-6068	chrB2:143978648-143978665			
hsa-miR-4758-3p	chrD3:70986240-70986257			
NC_018726.3_9493	chrB1:17059541-17059558			
hsa-miR-29a-3p	chrA2:148601342-148601364	MIR29A	mir-29a	fca-mir-29a
hsa-miR-6859-3p	chrC2:2323582-2323599			
hsa-miR-338-3p	chrE1:61695558-61695579			
NC_018725.3_6091	chrA3:4304543-4304560			
hsa-miR-374a-5p	chrX:62661540-62661561	MIR374B	mir-374b	fca-mir-374b
NC_018738.3_36088	chrE3:43332946-43332963			
NC_018730.3_22053	chrC1:184154403-184154421			
NC_018739.3_36383	chrF1:31198573-31198591			
NC_018729.3_18908	chrB4:136777475-136777492			
hsa-miR-23a-3p	chrAANG04001161:44966-44987/chrA2:9825373-9825394	MIR23A	mir-23a	
hsa-miR-486-5p	chrB1:43294753-43294774	MIR486-1		fca-mir-486-1
NC_018726.3_8369	chrB1:46921986-46922004			
hsa-miR-151a-3p	chrF2:82608506-82608526	MIR151A		fca-mir-151
hsa-miR-664a-3p	chrF1:59463563-59463586			fca-mir-664
NC_018732.3_25639	chrD1:18534677-18534696			
hsa-miR-200b-3p	chrC1:279272-279294	MIR200B	mir-200b	fca-mir-200b

Supplementary Table 11. TargetScan output showing miRNAs regulating SEPT8 and EPB41L5 genes with their corresponding binding site on the 3'UTR region. The last column shows the types of seed-matched sites with the following hierarchy of site efficacy: 8mer > 7mer-m8 > 7mer- A1 > 6mer. In bold the binding site of hsa-mir-186-5p that resulted mutated in the 3'UTR region of the EPB41L5 gene.

Gene_ID	miRNA_family_ID	species_ID	UTR_start	UTR_end	Site_type
SEPT8-201	NC_018723.3_1520	9685	324	330	7mer-m8
SEPT8-201	NC_018724.3_4857	9685	1035	1041	7mer-1a
SEPT8-201	NC_018724.3_4857	9685	930	936	7mer-m8
SEPT8-201	NC_018725.3_7102	9685	37	43	7mer-m8
SEPT8-201	NC_018726.3_8147	9685	452	458	7mer-1a
SEPT8-201	NC_018729.3_18936	9685	881	886	6mer
SEPT8-201	NC_018729.3_18944	9685	937	944	8mer-1a
SEPT8-201	NC_018730.3_20992	9685	623	628	6mer
SEPT8-201	NC_018730.3_20992	9685	661	666	6mer
SEPT8-201	NC_018730.3_21281	9685	870	875	6mer
SEPT8-201	NC_018730.3_21708	9685	1067	1073	7mer-m8
SEPT8-201	NC_018730.3_22053	9685	994	1000	7mer-m8
SEPT8-201	NC_018735.3_30238	9685	19	26	8mer-1a
SEPT8-201	NC_018735.3_30916	9685	842	847	6mer
SEPT8-201	NC_018739.3_36196	9685	995	1000	6mer
SEPT8-201	NC_018741.3_39910	9685	272	278	7mer-1a
SEPT8-201	NC_018741.3_39910	9685	582	587	6mer
SEPT8-201	NC_018741.3_39910	9685	984	990	7mer-1a
SEPT8-201	NW_019368574.1_41036	9685	224	230	7mer-m8
SEPT8-201	miR-23a-3p	9685	641	647	7mer-1a
SEPT8-201	miR-23a-3p	9685	877	882	6mer
SEPT8-201	miR-24-3p	9685	945	950	6mer
SEPT8-201	miR-29a-3p	9685	59	64	6mer
SEPT8-201	miR-425-5p	9685	802	807	6mer
SEPT8-201	miR-425-5p	9685	867	872	6mer
SEPT8-201	miR-520g-5p	9685	818	825	8mer-1a
SEPT8-201	miR-6859-3p	9685	787	792	6mer
SEPT8-202	NC_018723.3_2706	9685	2014	2019	6mer
SEPT8-202	NC_018723.3_2706	9685	2163	2168	6mer
SEPT8-202	NC_018724.3_4857	9685	2026	2031	6mer
SEPT8-202	NC_018724.3_4857	9685	579	585	7mer-m8
SEPT8-202	NC_018724.3_5609	9685	1283	1288	6mer
SEPT8-202	NC_018725.3_7102	9685	1288	1294	7mer-m8
SEPT8-202	NC_018726.3_8928	9685	1339	1345	7mer-1a
SEPT8-202	NC_018726.3_8928	9685	565	571	7mer-m8

SEPT8-202	NC_018727.3_12558	9685	2249	2255	7mer-1a
SEPT8-202	NC_018728.3_14704	9685	1789	1795	7mer-m8
SEPT8-202	NC_018728.3_14704	9685	1920	1926	7mer-m8
SEPT8-202	NC_018729.3_16994	9685	2098	2103	6mer
SEPT8-202	NC_018731.3_23601	9685	19	24	6mer
SEPT8-202	NC_018732.3_26240	9685	1978	1985	8mer-1a
SEPT8-202	NC_018733.3_27088	9685	1163	1168	6mer
SEPT8-202	NC_018734.3_28736	9685	2311	2317	7mer-1a
SEPT8-202	NC_018734.3_28736	9685	799	805	7mer-m8
SEPT8-202	NC_018735.3_30916	9685	1276	1281	6mer
SEPT8-202	NC_018736.3_32583	9685	37	43	7mer-1a
SEPT8-202	NC_018737.3_34091	9685	1181	1186	6mer
SEPT8-202	NC_018738.3_36088	9685	1165	1171	7mer-1a
SEPT8-202	NC_018739.3_36609	9685	2248	2254	7mer-1a
SEPT8-202	NC_018739.3_36609	9685	27	33	7mer-m8
SEPT8-202	NW_019365585.1_40446	9685	731	736	6mer
SEPT8-202	NW_019368574.1_41036	9685	1857	1863	7mer-1a
SEPT8-202	miR-125b-5p	9685	1547	1553	7mer-m8
SEPT8-202	miR-127-3p	9685	90	95	6mer
SEPT8-202	miR-142-5p	9685	175	181	7mer-1a
SEPT8-202	miR-142-5p	9685	882	887	6mer
SEPT8-202	miR-186-5p	9685	312	318	7mer-m8
SEPT8-202	miR-192-5p	9685	2238	2245	8mer-1a
SEPT8-202	miR-204-5p	9685	604	610	7mer-1a
SEPT8-202	miR-23a-3p	9685	2087	2092	6mer
SEPT8-202	miR-29a-3p	9685	104	109	6mer
SEPT8-202	miR-30a-5p	9685	1771	1776	6mer
SEPT8-202	miR-425-5p	9685	1626	1631	6mer
SEPT8-202	miR-425-5p	9685	905	910	6mer
SEPT8-202	miR-4451	9685	1667	1672	6mer
SEPT8-202	miR-4758-3p	9685	1292	1297	6mer
SEPT8-202	miR-590-3p	9685	1836	1842	7mer-m8
SEPT8-202	miR-664a-3p	9685	261	267	7mer-m8
SEPT8-203	NC_018724.3_4857	9685	579	585	7mer-m8
SEPT8-203	NC_018724.3_5609	9685	1283	1288	6mer
SEPT8-203	NC_018725.3_7102	9685	1288	1294	7mer-m8
SEPT8-203	NC_018726.3_8928	9685	1339	1345	7mer-1a
SEPT8-203	NC_018726.3_8928	9685	565	571	7mer-m8
SEPT8-203	NC_018728.3_14704	9685	1789	1795	7mer-m8
SEPT8-203	NC_018728.3_14704	9685	1920	1926	7mer-m8
SEPT8-203	NC_018731.3_23601	9685	19	24	6mer
SEPT8-203	NC_018732.3_26240	9685	1978	1985	8mer-1a

SEPT8-203	NC_018733.3_27088	9685	1163	1168	6mer
SEPT8-203	NC_018734.3_28736	9685	799	805	7mer-m8
SEPT8-203	NC_018735.3_30916	9685	1276	1281	6mer
SEPT8-203	NC_018736.3_32583	9685	37	43	7mer-1a
SEPT8-203	NC_018737.3_34091	9685	1181	1186	6mer
SEPT8-203	NC_018738.3_36088	9685	1165	1171	7mer-1a
SEPT8-203	NC_018739.3_36609	9685	27	33	7mer-m8
SEPT8-203	NW_019365585.1_40446	9685	731	736	6mer
SEPT8-203	NW_019368574.1_41036	9685	1857	1863	7mer-1a
SEPT8-203	miR-125b-5p	9685	1547	1553	7mer-m8
SEPT8-203	miR-127-3p	9685	90	95	6mer
SEPT8-203	miR-142-5p	9685	175	181	7mer-1a
SEPT8-203	miR-142-5p	9685	882	887	6mer
SEPT8-203	miR-186-5p	9685	312	318	7mer-m8
SEPT8-203	miR-204-5p	9685	604	610	7mer-1a
SEPT8-203	miR-29a-3p	9685	104	109	6mer
SEPT8-203	miR-30a-5p	9685	1771	1776	6mer
SEPT8-203	miR-425-5p	9685	1626	1631	6mer
SEPT8-203	miR-425-5p	9685	905	910	6mer
SEPT8-203	miR-4451	9685	1667	1672	6mer
SEPT8-203	miR-4758-3p	9685	1292	1297	6mer
SEPT8-203	miR-590-3p	9685	1836	1842	7mer-m8
SEPT8-203	miR-664a-3p	9685	261	267	7mer-m8
EPB41L5-201	NC_018723.3_1520	9685	2345	2351	7mer-1a
EPB41L5-201	NC_018723.3_1520	9685	549	555	7mer-m8
EPB41L5-201	NC_018723.3_2706	9685	305	311	7mer-m8
EPB41L5-201	NC_018723.3_2706	9685	4468	4474	7mer-m8
EPB41L5-201	NC_018724.3_4857	9685	3792	3798	7mer-m8
EPB41L5-201	NC_018725.3_6039	9685	1698	1703	6mer
EPB41L5-201	NC_018725.3_6039	9685	2019	2025	7mer-m8
EPB41L5-201	NC_018725.3_6039	9685	4566	4571	6mer
EPB41L5-201	NC_018725.3_6091	9685	3181	3187	7mer-m8
EPB41L5-201	NC_018725.3_6968	9685	3394	3399	6mer
EPB41L5-201	NC_018725.3_7102	9685	1459	1465	7mer-1a
EPB41L5-201	NC_018725.3_7102	9685	1646	1652	7mer-m8
EPB41L5-201	NC_018725.3_7102	9685	1659	1665	7mer-1a
EPB41L5-201	NC_018727.3_12196	9685	2175	2181	7mer-1a
EPB41L5-201	NC_018727.3_13107	9685	1511	1516	6mer
EPB41L5-201	NC_018727.3_13107	9685	3308	3314	7mer-1a
EPB41L5-201	NC_018727.3_13107	9685	466	471	6mer
EPB41L5-201	NC_018727.3_13107	9685	693	698	6mer
EPB41L5-201	NC_018728.3_14215	9685	1551	1556	6mer

EPB41L5-201	NC_018728.3_14215	9685	2777	2783	7mer-m8
EPB41L5-201	NC_018728.3_14215	9685	2953	2958	6mer
EPB41L5-201	NC_018729.3_18936	9685	2995	3001	7mer-1a
EPB41L5-201	NC_018729.3_18936	9685	4403	4408	6mer
EPB41L5-201	NC_018730.3_20992	9685	4393	4398	6mer
EPB41L5-201	NC_018730.3_21281	9685	1981	1987	7mer-1a
EPB41L5-201	NC_018732.3_26240	9685	3100	3105	6mer
EPB41L5-201	NC_018733.3_27088	9685	2129	2135	7mer-1a
EPB41L5-201	NC_018733.3_27088	9685	2807	2812	6mer
EPB41L5-201	NC_018733.3_27088	9685	2870	2875	6mer
EPB41L5-201	NC_018733.3_28048	9685	250	255	6mer
EPB41L5-201	NC_018734.3_28736	9685	3135	3140	6mer
EPB41L5-201	NC_018734.3_28736	9685	415	421	7mer-1a
EPB41L5-201	NC_018734.3_28894	9685	2685	2691	7mer-m8
EPB41L5-201	NC_018734.3_28894	9685	3882	3887	6mer
EPB41L5-201	NC_018735.3_30238	9685	1650	1655	6mer
EPB41L5-201	NC_018735.3_30238	9685	2809	2814	6mer
EPB41L5-201	NC_018736.3_32583	9685	2813	2818	6mer
EPB41L5-201	NC_018737.3_34091	9685	2167	2172	6mer
EPB41L5-201	NC_018737.3_34091	9685	397	402	6mer
EPB41L5-201	NC_018738.3_35045	9685	1142	1149	8mer-1a
EPB41L5-201	NC_018739.3_36383	9685	4578	4584	7mer-m8
EPB41L5-201	NC_018739.3_36609	9685	3884	3889	6mer
EPB41L5-201	NC_018739.3_36967	9685	2858	2863	6mer
EPB41L5-201	NC_018740.3_37965	9685	2495	2500	6mer
EPB41L5-201	NC_018740.3_37965	9685	3477	3483	7mer-m8
EPB41L5-201	NC_018740.3_38215	9685	2824	2829	6mer
EPB41L5-201	NC_018740.3_38215	9685	3929	3934	6mer
EPB41L5-201	NW_019365585.1_40446	9685	2489	2494	6mer
EPB41L5-201	NW_019367678.1_40856	9685	2771	2776	6mer
EPB41L5-201	NW_019368574.1_41036	9685	1948	1954	7mer-1a
EPB41L5-201	NW_019368574.1_41036	9685	3802	3807	6mer
EPB41L5-201	miR-1251-3p	9685	249	255	7mer-1a
EPB41L5-201	miR-125b-5p	9685	2754	2759	6mer
EPB41L5-201	miR-186-5p	9685	1598	1604	7mer-1a
EPB41L5-201	miR-186-5p	9685	3283	3290	8mer-1a
EPB41L5-201	miR-186-5p	9685	3331	3337	7mer-m8
EPB41L5-201**	miR-186-5p	9685	4374	4380	7mer-m8
EPB41L5-201	miR-192-5p	9685	1040	1046	7mer-1a
EPB41L5-201	miR-192-5p	9685	2269	2275	7mer-1a
EPB41L5-201	miR-200b-3p	9685	4520	4525	6mer
EPB41L5-201	miR-204-5p	9685	2183	2188	6mer

EPB41L5-201	miR-219a-5p	9685	2657	2663	7mer-1a
EPB41L5-201	miR-219a-5p	9685	2681	2686	6mer
EPB41L5-201	miR-23a-3p	9685	305	310	6mer
EPB41L5-201	miR-23a-3p	9685	4467	4473	7mer-m8
EPB41L5-201	miR-26a-5p	9685	313	319	7mer-1a
EPB41L5-201	miR-26a-5p	9685	3766	3772	7mer-m8
EPB41L5-201	miR-30a-5p	9685	4063	4068	6mer
EPB41L5-201	miR-30a-5p	9685	4183	4188	6mer
EPB41L5-201	miR-361-5p	9685	1861	1868	8mer-1a
EPB41L5-201	miR-374a-5p	9685	1044	1051	8mer-1a
EPB41L5-201	miR-374a-5p	9685	1277	1282	6mer
EPB41L5-201	miR-374a-5p	9685	4059	4064	6mer
EPB41L5-201	miR-425-5p	9685	639	644	6mer
EPB41L5-201	miR-4467	9685	3873	3879	7mer-m8
EPB41L5-201	miR-4520-5p	9685	2493	2498	6mer
EPB41L5-201	miR-486-5p	9685	1663	1669	7mer-m8
EPB41L5-201	miR-486-5p	9685	4568	4574	7mer-1a
EPB41L5-201	miR-590-3p	9685	1834	1840	7mer-m8
EPB41L5-201	miR-590-3p	9685	3062	3068	7mer-m8
EPB41L5-201	miR-622	9685	3263	3269	7mer-1a
EPB41L5-201	miR-622	9685	4113	4119	7mer-1a
EPB41L5-201	miR-664a-3p	9685	3600	3606	7mer-m8
EPB41L5-201	miR-6859-3p	9685	2983	2989	7mer-1a
EPB41L5-201	miR-6859-3p	9685	3055	3060	6mer
EPB41L5-202	NC_018723.3_1520	9685	2345	2351	7mer-1a
EPB41L5-202	NC_018723.3_1520	9685	549	555	7mer-m8
EPB41L5-202	NC_018723.3_2706	9685	305	311	7mer-m8
EPB41L5-202	NC_018723.3_2706	9685	4468	4474	7mer-m8
EPB41L5-202	NC_018724.3_4857	9685	3792	3798	7mer-m8
EPB41L5-202	NC_018725.3_6039	9685	1698	1703	6mer
EPB41L5-202	NC_018725.3_6039	9685	2019	2025	7mer-m8
EPB41L5-202	NC_018725.3_6039	9685	4566	4571	6mer
EPB41L5-202	NC_018725.3_6091	9685	3181	3187	7mer-m8
EPB41L5-202	NC_018725.3_6968	9685	3394	3399	6mer
EPB41L5-202	NC_018725.3_7102	9685	1459	1465	7mer-1a
EPB41L5-202	NC_018725.3_7102	9685	1646	1652	7mer-m8
EPB41L5-202	NC_018725.3_7102	9685	1659	1665	7mer-1a
EPB41L5-202	NC_018727.3_12196	9685	2175	2181	7mer-1a
EPB41L5-202	NC_018727.3_13107	9685	1511	1516	6mer
EPB41L5-202	NC_018727.3_13107	9685	3308	3314	7mer-1a
EPB41L5-202	NC_018727.3_13107	9685	466	471	6mer
EPB41L5-202	NC_018727.3_13107	9685	693	698	6mer

EPB41L5-202	NC_018728.3_14215	9685	1551	1556	6mer
EPB41L5-202	NC_018728.3_14215	9685	2777	2783	7mer-m8
EPB41L5-202	NC_018728.3_14215	9685	2953	2958	6mer
EPB41L5-202	NC_018729.3_18936	9685	2995	3001	7mer-1a
EPB41L5-202	NC_018729.3_18936	9685	4403	4408	6mer
EPB41L5-202	NC_018730.3_20992	9685	4393	4398	6mer
EPB41L5-202	NC_018730.3_21281	9685	1981	1987	7mer-1a
EPB41L5-202	NC_018732.3_26240	9685	3100	3105	6mer
EPB41L5-202	NC_018733.3_27088	9685	2129	2135	7mer-1a
EPB41L5-202	NC_018733.3_27088	9685	2807	2812	6mer
EPB41L5-202	NC_018733.3_27088	9685	2870	2875	6mer
EPB41L5-202	NC_018733.3_28048	9685	250	255	6mer
EPB41L5-202	NC_018734.3_28736	9685	3135	3140	6mer
EPB41L5-202	NC_018734.3_28736	9685	415	421	7mer-1a
EPB41L5-202	NC_018734.3_28894	9685	2685	2691	7mer-m8
EPB41L5-202	NC_018734.3_28894	9685	3882	3887	6mer
EPB41L5-202	NC_018735.3_30238	9685	1650	1655	6mer
EPB41L5-202	NC_018735.3_30238	9685	2809	2814	6mer
EPB41L5-202	NC_018736.3_32583	9685	2813	2818	6mer
EPB41L5-202	NC_018737.3_34091	9685	2167	2172	6mer
EPB41L5-202	NC_018737.3_34091	9685	397	402	6mer
EPB41L5-202	NC_018738.3_35045	9685	1142	1149	8mer-1a
EPB41L5-202	NC_018739.3_36383	9685	4578	4584	7mer-m8
EPB41L5-202	NC_018739.3_36609	9685	3884	3889	6mer
EPB41L5-202	NC_018739.3_36967	9685	2858	2863	6mer
EPB41L5-202	NC_018740.3_37965	9685	2495	2500	6mer
EPB41L5-202	NC_018740.3_37965	9685	3477	3483	7mer-m8
EPB41L5-202	NC_018740.3_38215	9685	2824	2829	6mer
EPB41L5-202	NC_018740.3_38215	9685	3929	3934	6mer
EPB41L5-202	NW_019365585.1_40446	9685	2489	2494	6mer
EPB41L5-202	NW_019367678.1_40856	9685	2771	2776	6mer
EPB41L5-202	NW_019368574.1_41036	9685	1948	1954	7mer-1a
EPB41L5-202	NW_019368574.1_41036	9685	3802	3807	6mer
EPB41L5-202	miR-1251-3p	9685	249	255	7mer-1a
EPB41L5-202	miR-125b-5p	9685	2754	2759	6mer
EPB41L5-202	miR-186-5p	9685	1598	1604	7mer-1a
EPB41L5-202	miR-186-5p	9685	3283	3290	8mer-1a
EPB41L5-202	miR-186-5p	9685	3331	3337	7mer-m8
EPB41L5-202	miR-186-5p	9685	4374	4380	7mer-m8
EPB41L5-202	miR-192-5p	9685	1040	1046	7mer-1a
EPB41L5-202	miR-192-5p	9685	2269	2275	7mer-1a
EPB41L5-202	miR-200b-3p	9685	4520	4525	6mer

EPB41L5-202	miR-204-5p	9685	2183	2188	6mer
EPB41L5-202	miR-219a-5p	9685	2657	2663	7mer-1a
EPB41L5-202	miR-219a-5p	9685	2681	2686	6mer
EPB41L5-202	miR-23a-3p	9685	305	310	6mer
EPB41L5-202	miR-23a-3p	9685	4467	4473	7mer-m8
EPB41L5-202	miR-26a-5p	9685	313	319	7mer-1a
EPB41L5-202	miR-26a-5p	9685	3766	3772	7mer-m8
EPB41L5-202	miR-30a-5p	9685	4063	4068	6mer
EPB41L5-202	miR-30a-5p	9685	4183	4188	6mer
EPB41L5-202	miR-361-5p	9685	1861	1868	8mer-1a
EPB41L5-202	miR-374a-5p	9685	1044	1051	8mer-1a
EPB41L5-202	miR-374a-5p	9685	1277	1282	6mer
EPB41L5-202	miR-374a-5p	9685	4059	4064	6mer
EPB41L5-202	miR-425-5p	9685	639	644	6mer
EPB41L5-202	miR-4467	9685	3873	3879	7mer-m8
EPB41L5-202	miR-4520-5p	9685	2493	2498	6mer
EPB41L5-202	miR-486-5p	9685	1663	1669	7mer-m8
EPB41L5-202	miR-486-5p	9685	4568	4574	7mer-1a
EPB41L5-202	miR-590-3p	9685	1834	1840	7mer-m8
EPB41L5-202	miR-590-3p	9685	3062	3068	7mer-m8
EPB41L5-202	miR-622	9685	3263	3269	7mer-1a
EPB41L5-202	miR-622	9685	4113	4119	7mer-1a
EPB41L5-202	miR-664a-3p	9685	3600	3606	7mer-m8
EPB41L5-202	miR-6859-3p	9685	2983	2989	7mer-1a
EPB41L5-202	miR-6859-3p	9685	3055	3060	6mer
EPB41L5-204	NC_018723.3_1520	9685	1936	1941	6mer
EPB41L5-204	NC_018723.3_1520	9685	91	97	7mer-m8
EPB41L5-204	NC_018723.3_2706	9685	646	651	6mer
EPB41L5-204	NC_018724.3_5609	9685	992	997	6mer
EPB41L5-204	NC_018725.3_6039	9685	47	52	6mer
EPB41L5-204	NC_018725.3_6039	9685	943	948	6mer
EPB41L5-204	NC_018725.3_6091	9685	1932	1937	6mer
EPB41L5-204	NC_018725.3_6968	9685	977	983	7mer-1a
EPB41L5-204	NC_018727.3_12196	9685	996	1002	7mer-1a
EPB41L5-204	NC_018727.3_13107	9685	1175	1181	7mer-1a
EPB41L5-204	NC_018729.3_18936	9685	1324	1329	6mer
EPB41L5-204	NC_018729.3_18936	9685	681	686	6mer
EPB41L5-204	NC_018730.3_20992	9685	1454	1459	6mer
EPB41L5-204	NC_018730.3_20992	9685	2125	2131	7mer-m8
EPB41L5-204	NC_018730.3_20992	9685	477	482	6mer
EPB41L5-204	NC_018730.3_20992	9685	492	498	7mer-1a
EPB41L5-204	NC_018733.3_27088	9685	1834	1839	6mer

EPB41L5-204	NC_018733.3_27088	9685	2055	2061	7mer-m8
EPB41L5-204	NC_018734.3_28736	9685	2079	2085	7mer-m8
EPB41L5-204	NC_018734.3_28736	9685	53	59	7mer-1a
EPB41L5-204	NC_018738.3_35941	9685	248	253	6mer
EPB41L5-204	NC_018740.3_37965	9685	541	546	6mer
EPB41L5-204	NC_018740.3_37965	9685	785	790	6mer
EPB41L5-204	NW_019368574.1_41036	9685	1565	1570	6mer
EPB41L5-204	miR-125b-5p	9685	100	106	7mer-1a
EPB41L5-204	miR-125b-5p	9685	1326	1331	6mer
EPB41L5-204	miR-142-5p	9685	1056	1061	6mer
EPB41L5-204	miR-142-5p	9685	1754	1759	6mer
EPB41L5-204	miR-142-5p	9685	66	72	7mer-m8
EPB41L5-204	miR-144-3p	9685	747	752	6mer
EPB41L5-204	miR-153-3p	9685	74	79	6mer
EPB41L5-204	miR-185-5p	9685	1311	1317	7mer-m8
EPB41L5-204	miR-186-5p	9685	374	380	7mer-m8
EPB41L5-204	miR-186-5p	9685	676	682	7mer-1a
EPB41L5-204	miR-192-5p	9685	950	956	7mer-m8
EPB41L5-204	miR-204-5p	9685	1803	1810	8mer-1a
EPB41L5-204	miR-204-5p	9685	997	1004	8mer-1a
EPB41L5-204	miR-23a-3p	9685	1681	1687	7mer-1a
EPB41L5-204	miR-23a-3p	9685	257	263	7mer-m8
EPB41L5-204	miR-30a-5p	9685	1274	1280	7mer-1a
EPB41L5-204	miR-30a-5p	9685	1920	1925	6mer
EPB41L5-204	miR-30a-5p	9685	903	909	7mer-m8
EPB41L5-204	miR-425-5p	9685	159	164	6mer
EPB41L5-204	miR-4451	9685	1646	1651	6mer
EPB41L5-204	miR-4758-3p	9685	1837	1842	6mer
EPB41L5-204	miR-4758-3p	9685	916	921	6mer
EPB41L5-204	miR-486-5p	9685	497	502	6mer
EPB41L5-204	miR-520g-5p	9685	380	386	7mer-1a
EPB41L5-204	miR-590-3p	9685	1684	1689	6mer
EPB41L5-204	miR-590-3p	9685	2176	2181	6mer
EPB41L5-204	miR-590-3p	9685	665	671	7mer-1a
EPB41L5-204	miR-622	9685	1024	1029	6mer
EPB41L5-204	miR-664a-3p	9685	1226	1231	6mer
EPB41L5-204	miR-664a-3p	9685	2173	2179	7mer-1a
EPB41L5-205	NC_018725.3_6039	9685	121	127	7mer-m8
EPB41L5-205	NC_018725.3_6091	9685	115	120	6mer
EPB41L5-205	NC_018730.3_22053	9685	76	83	8mer-1a
EPB41L5-205	NC_018735.3_30916	9685	31	36	6mer
EPB41L5-205	NC_018739.3_36196	9685	77	83	7mer-1a

EPB41L5-205	NC_018740.3_37965	9685	117	123	7mer-m8
EPB41L5-205	miR-192-5p	9685	178	184	7mer-m8
EPB41L5-206	NC_018723.3_1520	9685	278	284	7mer-1a
EPB41L5-206	NC_018723.3_2706	9685	132	137	6mer
EPB41L5-206	NC_018723.3_2706	9685	229	234	6mer
EPB41L5-206	NC_018723.3_2706	9685	531	536	6mer
EPB41L5-206	NC_018723.3_2706	9685	537	542	6mer
EPB41L5-206	NC_018723.3_2706	9685	546	551	6mer
EPB41L5-206	NC_018724.3_4857	9685	356	361	6mer
EPB41L5-206	NC_018727.3_13107	9685	407	412	6mer
EPB41L5-206	NC_018728.3_14215	9685	351	356	6mer
EPB41L5-206	NC_018728.3_14704	9685	788	793	6mer
EPB41L5-206	NC_018730.3_20699	9685	486	491	6mer
EPB41L5-206	NC_018732.3_26240	9685	776	782	7mer-m8
EPB41L5-206	NC_018734.3_28736	9685	426	431	6mer
EPB41L5-206	NC_018735.3_30916	9685	31	36	6mer
EPB41L5-206	miR-142-5p	9685	507	512	6mer
EPB41L5-206	miR-144-3p	9685	118	123	6mer
EPB41L5-206	miR-185-5p	9685	775	780	6mer
EPB41L5-206	miR-204-5p	9685	154	160	7mer-m8
EPB41L5-206	miR-23a-3p	9685	303	309	7mer-1a
EPB41L5-206	miR-23a-3p	9685	320	326	7mer-1a
EPB41L5-206	miR-374a-5p	9685	334	339	6mer
EPB41L5-206	miR-590-3p	9685	843	850	8mer-1a
EPB41L5-206	miR-590-3p	9685	849	854	6mer
EPB41L5-206	miR-622	9685	381	386	6mer
EPB41L5-206	miR-664a-3p	9685	388	393	6mer

**match between the site of mutation and the miRNA binding site

6mer = perfect match with the 6-nt miRNA seed

7mer-m8 = comprises the seed match to miRNA nucleotide 8

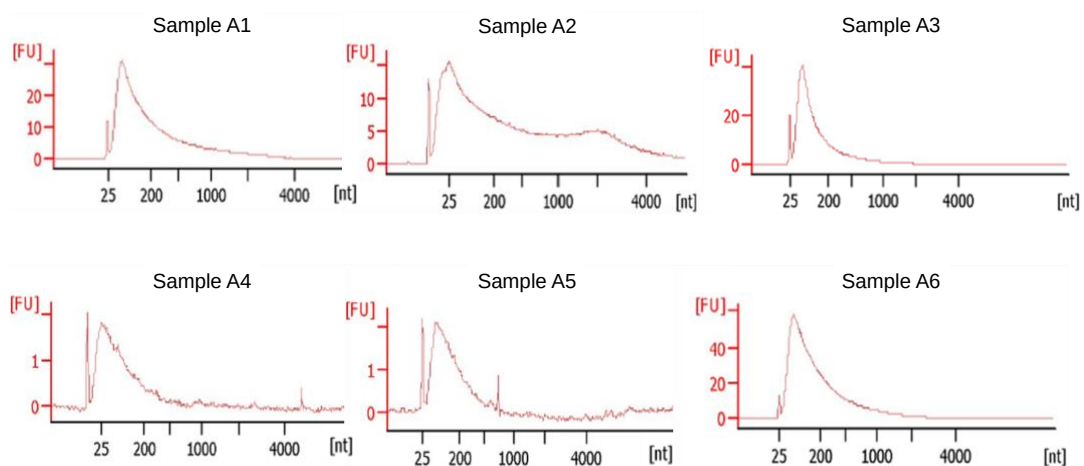
7mer-A1 = comprises the seed match supplemented by an A across from miRNA nucleotide 1

8mer = comprises the seed match supplemented by both the m8 and the A1

Supplementary Figures

[illegible]

Supplementary Figure 1. Alignment of the regions harboring the three prioritized variants in the SEPT8, EPB41L5 and ABCA8 genes. The mutated nucleotides are highlighted in red. Dots indicate the presence of the same nucleotide; a different nucleotide in the sequence is reported with the letter of the corresponding alternate nucleotide.



Supplementary Figure 2. Bioanalyzer Agilent RNA 6000 Nano Assay plots of samples characterized by a double step of deparaffinization. It is possible to notice two well distinguished peaks: the first one on the left represents the marker, while the second peak correspond to the amount of isolated miRNAs. Possible further peaks (sample A4; A5) have to be referred to potential contaminations of other longer RNAs.

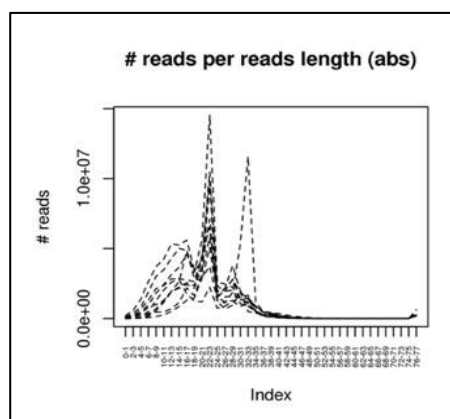


Figure 4A

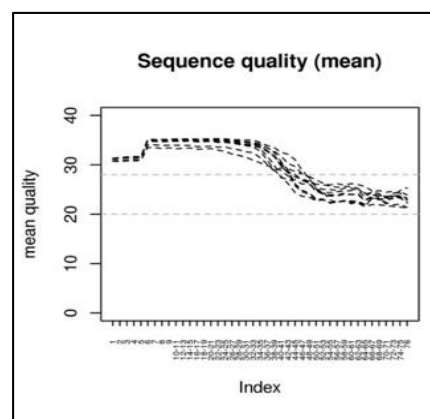
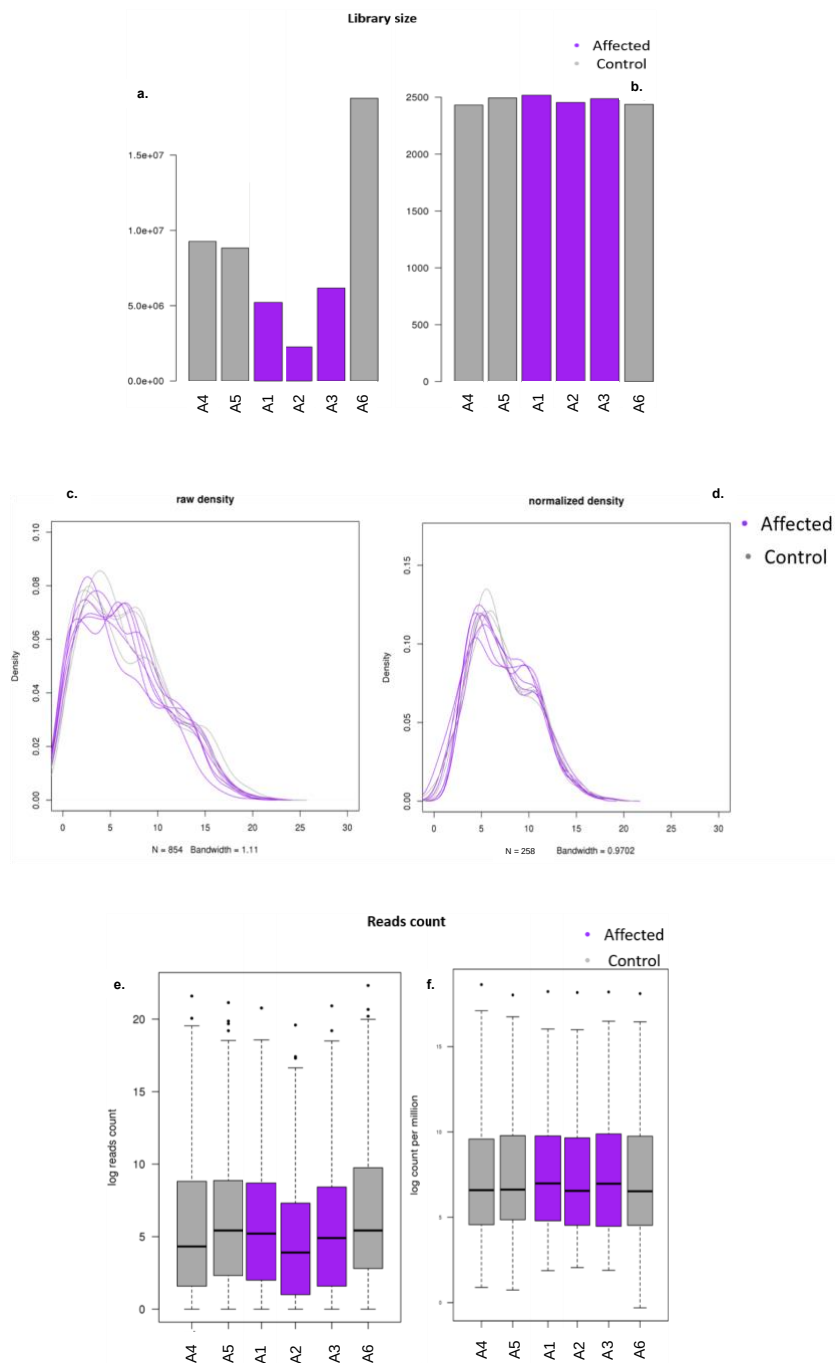
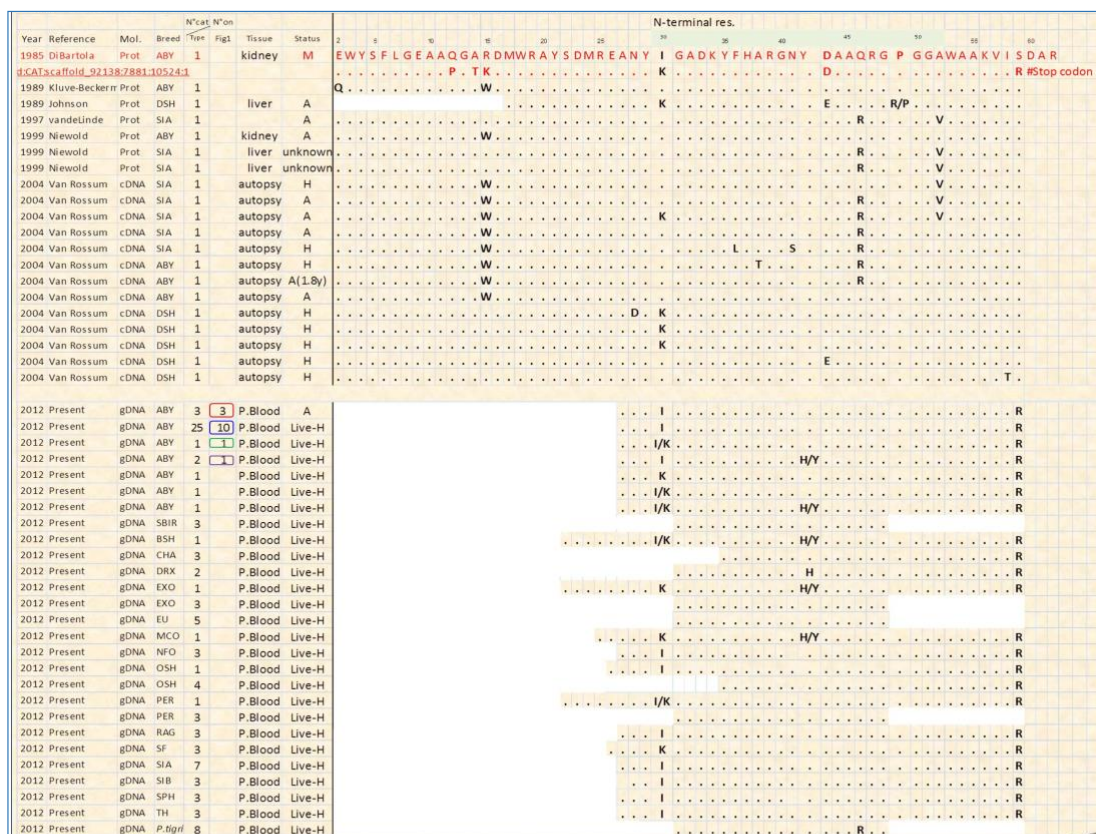


Figure 4B

Supplementary Figure 3. Quality control of the sequencing data with FastQC. **Figure 4A** shows the number of reads for each read length. **Figure 4B** reports the overall sequence quality.



Supplementary Figure 4. The figure shows the sequencing data before and after the normalization with the TMM method. On the left, figures **a**, **c** and **e**, shows the Library Size, the miRNA density and the reads count before the normalization. Figure **b**, **d** and **f**, represent the normalized sequencing data.



Supplementary Figure 5. Amino acid alignment of the putative amyloidogenic SAA coding region. The sequences of the cats of different breeds used in the study (including Abyssinians) are compared with the SAA sequences of previous works [Longeri et al. 2012. Available from https://www.isag.us/Docs/Proceedings/ISAG_Programme_and_Abstract_Book_2012.pdf]

2.5. Localizing the Regions of Causative Mutation in Siamese Amyloidosis through an NGS Approach

Abstract

To genetically identify potential variants associated to amyloidosis in the Siamese/Oriental cat, the whole genome sequences of two amyloidosis affected Siamese were compared to other 195 whole genome cat sequences available through the 99 Lives Initiative. Variants detection was performed with Golden Helix SVS software, using the version 9.0 of the cat genome assembly as the reference genome, and 350 variants were found to be unique in the affected individuals. Variants were annotated, and priority was given to variants mapping in exonic regions of the genome and to homozygous variants. The study is still ongoing, genes hosting the variants still have to be investigated and a population of about 100 Siamese is ready to be genotyped as soon as the main variants will be identified. In the following manuscript the last updated results are reported.

Materials and Methods

Samples

The WGS of two amyloidosis affected OSH cats were available through the 99 Lives project, together with 195 additional whole genome sequences of cats of different breeds and random bred cats (Control Population –**Table 1**). The two affected individuals were selected for the study to be ≤ 3 years of age and with a confirmatory diagnosis of hepatic rupture due to primary liver amyloidosis (**Table 1A**). The Control Population included other 16 cats between Siamese and OSH breeds. These individuals joined the 99 Lives project to study breed diversity and they were therefore considered healthy, representing an additional control for variants exclusion within the breed (**Table 1B**).

Table 1. List of the cats included in the study for the detection of amyloidosis related variants in Siamese/OSH breed. **Table 1A** shows the two affected Siamese, while **Table 1B** shows the number of cats per each breed in the Control Population.

Table 1A. Affected Siamese

Cat_ID	Breed	Age (years old)	Status
Fcat_20684	OSH	3	Affected
Fcat_20793	OSH	3	Affected

Table 1B. Control Population

Breed	Cat/breed
Abyssinian	4
American Shorhair	1
American Wirehair	1
Australian Mist	1
Bengal	5
Bengal x Japanese Bobtail	3
Bengal x Somali	1
Birman	10
British Shorthair	2
Burmese	13
Cross-breed colony	4
Devon Rex	5
Domestic Longhair	1
Don Sphynx	1
Egyptian Mau	3
Himalayan	1
Korat	2
LaPerm	1
Lykoi	1
Maine Coon	11
Maine Coon Cross	20
Napoleon	3
Ocicat	1
Oriental	4
Oriental Shorthair	1
Oriental Shorthair cross	3
Persian	3

Breed	Cat/breed
Persian cross	1
Persian x Bengal	2
Persian x Japanese Bobtail	1
Peterbald	1
Ragdoll	4
Random bred	55
Savannah	1
Selkirk Rex	1
Siamese	7
Siamese cross	1
Siberian	2
Sphynx	3
Tennessee Rex	1
Tonkinese	1
Toybob	2
Toyger	1
Turkish Angora	1
Unknown	4
Total	195

VCFtools

The whole genome sequences of 197 cats, belonging to the 99 Lives Initiative, were available in a VCF file format, with 84005772 annotated variants. The original VCF file was filtered using VCFtools [Danecek *et al.* 2011] to obtain a VCF file containing only the high-quality variants. All the individuals with >20% of missing genotypes were removed (--missing-indv) as well as all the variants that were missing in >20% of the individuals (--max-missing). GQ DP were set at ≥ 20 and ≥ 15 respectively (--minGQ; --minDP).

Golden Helix SVS

Variants detection was performed with the Golden Helix SVS software [Golden Helix, Inc., Bozeman, MT, www.goldenelix.com], using the VCF file with the high-quality variants as the input file. To identify the variants related only to the affected cats, the function “Activate Variants by Genotype Count Threshold” available through the SNP & Variation Suite v8.4 [Golden Helix, Inc., Bozeman, MT, www.goldenelix.com] was used. Samples were considered as not related and specific threshold were considered with custom setting. Since the Control Population included other individuals of the same breed, filters were applied to not completely exclude those variants that could possibly be present in healthy cats (disease carriers). Therefore, variants were selected when present in both the affected cats either with a heterozygous or homozygous genotype, and when present in the Control Population with a low genotype frequency <5% (homozygous and/or heterozygous).

Results*Detection of disease related variants*

VCF file filtering resulted in the removal of three individuals with >20% of missing genotypes, leaving 194 cats for the further analyses. Variants missing in >20% of the individuals were removed and 81123337 variants were finally considered as they all passed the filters for GQ and RD.

Considering the VCF file with the high-quality variants, the data of the two affected OSH were compared to those of the 192 cats of the Control Population and 362 potential disease related variants were detected with Golden Helix Software.

Intergenic variants represented a great part of the overall detected variants (n.213) with seven of them being homozygous in at least one of the affected cats (**Table 2; Supplementary Table 1**). The remaining 206 intergenic variants were heterozygous in the affected cats and therefore not considered in this first part of the study. Excluding the seven homozygous intergenic variants, other nine homozygous variants were detected, all of them mapping into intronic regions. Regarding the 140 heterozygous variants, nine mapped into exonic regions and 131 were found in introns (**Table 2; Supplementary Table 1**). Among the 362 detected variants, n.320 were SNPs, n.27 were deletions and n.15 were insertions (**Table 2**).

Table 2. Overview of the variants identified in the two affected OSH compared to the Control Population with the Golden helix SVS Software.

	TOT	HOMOZYGOUS	HETEROZYGOUS		SNPs	INSERTIONS	DELETIONS
INTERGENIC	213	7	206		193	6	14
INTRONIC	140	9	131		118	9	13
EXONIC	9	0	9		9	0	0
TOT	362	16	346		320	15	27

Supplementary Tables

Supplementary Table 1. List of the 156 variants related to two the affected cats. The 206 intergenic heterozygous variants are not reported.

HOMOZYGOUS VARIANTS						
Location	Ref	Alt	Gene	Intron/Exon	Fcat_20684	Fcat_20793
A2:34384766-SNV**	C	T	INTERGENIC		X*	X*
A2:59536480-SNV**	T	C	UROC1	IN	X*	X
B3:123919337-SNV	T	C	VASH1	IN	X*	X
D4:26308556-SNV	C	T	INTERGENIC		X*	X
D4:26662067-SNV	C	T	INTERGENIC		X*	X
D4:26710653-SNV	A	G	NAA35	IN	X*	X
D4:27265569-SNV	G	A	INTERGENIC		X*	X
E1:35442985-SNV**	G	A	INTERGENIC		X	X*
E1:35669216-SNV**	C	T	INTERGENIC		X	X*
E1:35908439-SNV**	T	A	CA10	IN	X	X*
E1:35908845-Ins**	/	INS A	CA10	IN	X	X*
E1:35908955-Del**	/	DEL TTGCCATTCTT	CA10	IN	X	X*
E1:35910059-SNV**	A	G	CA10	IN	X	X*
E1:35910401-SNV**	A	G	CA10	IN	X	X*
E1:36955655-SNV**	C	T	ENSFCAG0000004413 4	IN	X*	X
E1:37000570-SNV**	C	T	INTERGENIC		X*	X
HETEROZYGOUS VARIANTS WITHIN EXONS						
Location	Ref	Alt	Gene	Intron/Exon	Fcat_20684	Fcat_20793
A2:154094599-SNV	C	T	PTN	EX	X	X
A2:154094967-SNV	C	T	PTN	EX	X	X
B3:122577714-SNV	T	A	FOS	EX	X	X
C1:18932340-SNV	A	G	ENSFCAG0000004309 5	EX	X	X
C1:18952072-SNV	A	G	ENSFCAG0000001032 4	EX	X	X
C1:18958236-SNV	T	C	RSRP1	EX	X	X
C1:18960752-SNV	G	A	RSRP1	EX	X	X
C1:19004480-SNV	T	C	ENSFCAG0000001278 9	EX	X	X
D4:93270935-SNV	G	T	COL5A1	EX	X	X
HETEROZYGOUS VARIANTS WITHIN INTRONS						
Location	Ref	Alt	Gene	Intron/Exon	Fcat_20684	Fcat_20793
A1:208811380-SNV	G	A	MROH2B	IN	X	X
A1:241591317-SNV	C	T	BRD9	IN	X	X
A1:241780536-SNV	C	T	SLC9A3	IN	X	X
A2:15243211-SNV	C	T	LARS2	IN	X	X
A2:15308936-SNV	G	A	LARS2	IN	X	X

A2:154136371-SNV	A	C	PTN	IN	X	X
A2:154140121-SNV	C	G	PTN	IN	X	X
A2:154140133-SNV	G	A	PTN	IN	X	X
A2:154140217-SNV	T	C	PTN	IN	X	X
A2:154333091-SNV	G	A	DGKI	IN	X	X
A2:155218737-SNV	C	T	TRIM24	IN	X	X
A2:15667929-SNV	C	T	ENSFCAG00000027547	IN	X	X
A3:11036209-Del	/	DEL T	ENSFCAG00000041976	IN	X	X
A3:130481846-SNV	T	G	FAM49A	IN	X	X
A3:136269992-SNV	T	C	CPSF3	IN	X	X
A3:14672942-SNV	T	C	SLC13A3	IN	X	X
A3:14708719-Del	/	DEL G	SLC13A3	IN	X	X
B2:101341826-SNV	C	T	ENSFCAG00000006353	IN	X	X
B2:149730754-Del	/	DEL TATGTG	QKI	IN	X	X
B3:103481278-SNV	T	C	ENSFCAG00000041386	IN	X	X
B3:103492889-SNV	A	G	ENSFCAG00000041386	IN	X	X
B3:103675028-SNV	C	A	ENSFCAG00000040790	IN	X	X
B3:105644226-SNV	G	A	PELI2	IN	X	X
B3:112118151-SNV	C	G	PPP2R5E	IN	X	X
B3:122849658-SNV	A	G	FLVCR2	IN	X	X
B3:122877590-SNV	C	T	FLVCR2	IN	X	X
B3:123041725-SNV	C	T	TTLL5	IN	X	X
B3:123243553-SNV	G	A	IFT43	IN	X	X
B3:123392681-SNV	A	G	GPATCH2L	IN	X	X
B3:123397194-SNV	A	G	GPATCH2L	IN	X	X
B3:123398519-SNV	T	C	GPATCH2L	IN	X	X
B3:123398524-SNV	C	T	GPATCH2L	IN	X	X
B3:123399189-SNV	A	G	GPATCH2L	IN	X	X
B3:123399864-SNV	T	C	GPATCH2L	IN	X	X
B3:123400966-SNV	A	G	GPATCH2L	IN	X	X
B3:123401788-SNV	C	G	GPATCH2L	IN	X	X
B3:123412807-SNV	G	A	GPATCH2L	IN	X	X
B3:123427196-SNV	C	T	GPATCH2L	IN	X	X
B3:123429343-SNV	T	C	GPATCH2L	IN	X	X
C1:18933495-SNV	C	T	ENSFCAG00000043095	IN	X	X
C1:18933826-SNV	C	T	ENSFCAG00000043095	IN	X	X
C1:18934150-SNV	C	T	ENSFCAG00000043095	IN	X	X
C1:18935385-SNV	C	A	ENSFCAG00000043095	IN	X	X
C1:18949505-SNV	G	A	ENSFCAG00000010324	IN	X	X
C1:18950576-SNV	C	T	ENSFCAG00000010324	IN	X	X

C1:18951149-SNV	G	T	ENSFCAG00000010324	IN	X	X
C1:18952275-SNV	C	G	ENSFCAG00000010324	IN	X	X
C1:18952322-SNV	C	T	ENSFCAG00000010324	IN	X	X
C1:18952667-SNV	C	T	ENSFCAG00000010324	IN	X	X
C1:18953549-SNV	T	A	ENSFCAG00000010324	IN	X	X
C1:18958853-Ins	/	INS T	RSRP1	IN	X	X
C1:18959089-SNV	G	A	RSRP1	IN	X	X
C1:18959336-SNV	A	G	RSRP1	IN	X	X
C1:18960402-SNV	C	T	RSRP1	IN	X	X
C1:18961926-SNV	C	T	RSRP1	IN	X	X
C1:18985221-SNV	G	C	TMEM50A	IN	X	X
C1:18985940-SNV	T	C	TMEM50A	IN	X	X
C1:18987860-SNV	C	T	TMEM50A	IN	X	X
C1:18987936-SNV	A	G	TMEM50A	IN	X	X
C1:18988796-SNV	G	A	TMEM50A	IN	X	X
C1:18989631-Ins	/	INS TCTTG	TMEM50A	IN	X	X
C1:18991583-SNV	A	G	TMEM50A	IN	X	X
C1:18994364-SNV	C	T	TMEM50A	IN	X	X
C1:19000921-SNV	G	C	TMEM50A	IN	X	X
C1:19001925-SNV	G	T	TMEM50A	IN	X	X
C1:19002062-SNV	G	A	TMEM50A	IN	X	X
C1:19005752-SNV	G	A	ENSFCAG00000012789	IN	X	X
C1:19007555-SNV	T	G	ENSFCAG00000012789	IN	X	X
C1:19012458-Del	/	DEL TTGTT	ENSFCAG00000012789	IN	X	X
C1:19013616-SNV	A	G	ENSFCAG00000012789	IN	X	X
C1:19017395-SNV	G	A	ENSFCAG00000012789	IN	X	X
C1:19018658-SNV	G	A	ENSFCAG00000012789	IN	X	X
C1:19019844-SNV	G	A	ENSFCAG00000012789	IN	X	X
C1:19020415-SNV	A	G	ENSFCAG00000012789	IN	X	X
C1:19020977-SNV	A	G	ENSFCAG00000012789	IN	X	X
C1:19023393-SNV	A	G	ENSFCAG00000012789	IN	X	X
C1:19023909-SNV	T	C	ENSFCAG00000012789	IN	X	X
C1:19023954-SNV	T	C	ENSFCAG00000012789	IN	X	X
C1:19024014-SNV	T	C	ENSFCAG00000012789	IN	X	X
C1:19024691-SNV	A	G	ENSFCAG00000012789	IN	X	X
C1:19024795-SNV	T	C	ENSFCAG00000012789	IN	X	X
C1:19032670-SNV	T	G	ENSFCAG00000012789	IN	X	X
C1:19032708-SNV	A	G	ENSFCAG00000012789	IN	X	X
C1:19038316-SNV	A	G	ENSFCAG00000012789	IN	X	X

C1:19041033-Ins	/	INS A	ENSFCAG00000012789	IN	X	X
C1:19554137-SNV	C	A	PAFAH2	IN	X	X
C1:19972580-Del	/	DEL AGAC	RPS6KA1	IN	X	X
C1:19972581-Del	/	DEL GAC	RPS6KA1	IN	X	X
C1:19975621-SNV	G	A	RPS6KA1	IN	X	X
C1:19987740-SNV	G	C	RPS6KA1	IN	X	X
C1:19987749-SNV	C	T	RPS6KA1	IN	X	X
C1:21555843-SNV	C	A	TRNAU1AP	IN	X	X
C1:21555859-SNV	T	C	TRNAU1AP	IN	X	X
C1:46677815-SNV	C	T	DAB1	IN	X	X
D2:81817432-SNV	A	C	ENSFCAG00000043104	IN	X	X
D2:87015070-SNV	G	A	ENSFCAG00000035983	IN	X	X
D2:87015149-SNV	G	A	ENSFCAG00000035983	IN	X	X
D2:87016885-SNV	G	A	ENSFCAG00000035983	IN	X	X
D2:87018029-Ins	/	INS TGT	ENSFCAG00000035983	IN	X	X
D2:87018030-Ins	/	INS ATAAATAATAATA GGATGTAATAAAT ATAATAAAT	ENSFCAG00000035983	IN	X	X
D2:87018512-SNV	G	A	ENSFCAG00000035983	IN	X	X
D2:87018791-SNV	G	C	ENSFCAG00000035983	IN	X	X
D2:87020508-Del	/	DEL C	ENSFCAG00000035983	IN	X	X
D2:87022889-SNV	G	A	ENSFCAG00000035983	IN	X	X
D2:87023549-Del	/	DEL C	ENSFCAG00000035983	IN	X	X
D2:87025785-Del	/	DEL A	ENSFCAG00000035983	IN	X	X
D2:87033523-SNV	G	A	ENSFCAG00000035983	IN	X	X
D2:87036760-SNV	G	A	ENSFCAG00000035983	IN	X	X
D2:87045571-SNV	G	A	ENSFCAG00000035983	IN	X	X
D2:87052944-SNV	T	C	ENSFCAG00000035983	IN	X	X
D3:70261355-SNV	C	T	ZBTB7C	IN	X	X
D4:93493035-SNV	C	A	RXRA	IN	X	X
D4:95080558-SNV	C	T	CAMSAP1	IN	X	X
D4:95083298-SNV	G	A	CAMSAP1	IN	X	X
D4:95130342-SNV	G	A	KCNT1	IN	X	X
D4:95457799-Del	/	DEL G	RABL6	IN	X	X
E1:32745240-SNV	C	T	PCTP	IN	X	X
E1:32747569-Ins	/	INS T	PCTP	IN	X	X
E1:32748364-Del	/	DEL G	PCTP	IN	X	X
E1:32748528-SNV	T	C	PCTP	IN	X	X
E1:32750222-SNV	G	A	PCTP	IN	X	X
E1:32750714-SNV	T	C	PCTP	IN	X	X

E1:32752923-Ins	/	INS CAAAGGAATCG	PCTP	IN	X	X
E1:32754003-SNV	C	T	PCTP	IN	X	X
E1:32754790-SNV	A	C	PCTP	IN	X	X
E1:32755002-Ins	/	INS C	PCTP	IN	X	X
E1:32755630-SNV	A	G	PCTP	IN	X	X
E1:33055656-SNV	T	A	MMD	IN	X	X
E2:61586217-SNV	C	T	ENSFCAG0000004202 3	IN	X	X
E3:1832850-SNV	C	G	MAD1L1	IN	X	X
E3:2529707-Del	/	DEL TT	CARD11	IN	X	X

*Cat homozygous for the variant

**Variant shared by other cats

INS = Insertion

DEL = Deletion

2.6. References

1. Aberdein, D., Munday, J.S., Gandolfi, B., Dittmer, K.E., Malik, R., Garrick, D.J., Lyons, L.A., and the 99 Lives Consortium 2017. A FAS-ligand variant associated with autoimmune lymphoproliferative syndrome in cats. *Mamm Genome* 28, 47–55. <https://doi.org/10.1007/s00335-016-9668-1>
2. Adams, D., Koike, H., Slama, M., Coelho, T., 2019. Hereditary transthyretin amyloidosis: a model of medical progress for a fatal disease. *Nat Rev Neurol* 15, 387–404. <https://doi.org/10.1038/s41582-019-0210-4>
3. Alhaddad, H., Gandolfi, B., Grahn, R.A., Rah, H.-C., Peterson, C.B., Maggs, D.J., Good, K.L., Pedersen, N.C., Lyons, L.A., 2014. Genome-wide association and linkage analyses localize a progressive retinal atrophy locus in Persian cats. *Mamm Genome* 25, 354–362. <https://doi.org/10.1007/s00335-014-9517-z>
4. Andel, A.C.J., Gruys, E., Kroneman, J., Veerkamp, J., 1988. Amyloid in the horse: A report of nine cases. *Equine Veterinary Journal* 20, 277–285. <https://doi.org/10.1111/j.2042-3306.1988.tb01524.x>
5. Backes, C., Khaleeq, Q.T., Meese, E., Keller, A., 2016. miEAA: microRNA enrichment analysis and annotation. *Nucleic Acids Res* 44, W110–W116. <https://doi.org/10.1093/nar/gkw345>
6. Beatty, J.A., Barrs, V.R., Martin, P.A., Nicoll, R.G., France, M.P., Foster, S.F., Lamb, W.A., Malik, R., 2002. Spontaneous hepatic rupture in six cats with systemic amyloidosis. *J Small Animal Practice* 43, 355–363. <https://doi.org/10.1111/j.1748-5827.2002.tb00086.x>
7. Ben Halima, S., Siegel, G., Rajendran, L., 2016. miR-186 in Alzheimer's disease: a big hope for a small RNA? *J. Neurochem.* 137, 308–311. <https://doi.org/10.1111/jnc.13573>
8. Benson, M.D., Dwulet, F.E., DiBartola, S.P., 1985. Identification and characterization of amyloid protein AA in spontaneous canine amyloidosis. *Lab. Invest.* 52, 448–452.
9. Boo, J.H., Sohn, J.H., Kim, J.E., Song, H., Mook-Jung, I., 2008. Rac1 changes the substrate specificity of γ -secretase between amyloid precursor protein and Notch1. *Biochemical and Biophysical Research Communications* 372, 913–917. <https://doi.org/10.1016/j.bbrc.2008.05.153>
10. Boyce, J.T., DiBartola, S.P., Chew, D.J., Gasper, P.W., 1984. Familial Renal Amyloidosis in Abyssinian Cats. *Vet Pathol* 21, 33–38. <https://doi.org/10.1177/030098588402100106>

11. Brandan, E., Inestrosa, N.C., 1993. Extracellular matrix components and amyloid in neuritic plaques of Alzheimer's disease. *Gen. Pharmacol.* 24, 1063–1068. [https://doi.org/10.1016/0306-3623\(93\)90350-7](https://doi.org/10.1016/0306-3623(93)90350-7)
12. Bronfman, F.C., Soto, C., Tapia, L., Tapia, V., Inestrosa, N.C., 1996. Extracellular matrix regulates the amount of the beta-amyloid precursor protein and its amyloidogenic fragments. *J. Cell. Physiol.* 166, 360–369. [https://doi.org/10.1002/\(SICI\)1097-4652\(199602\)166:2<360::AID-JCP14>3.0.CO;2-F](https://doi.org/10.1002/(SICI)1097-4652(199602)166:2<360::AID-JCP14>3.0.CO;2-F)
13. Chang, W.-S., Wang, Y.-H., Zhu, X.-T., Wu, C.-J., 2017. Genome-Wide Profiling of miRNA and mRNA Expression in Alzheimer's Disease. *Med Sci Monit* 23, 2721–2731. <https://doi.org/10.12659/MSM.905064>
14. Chew, D.J., DiBartola, S.P., Boyce, J.T., Gasper, P.W., 1982. Renal amyloidosis in related Abyssinian cats. *J. Am. Vet. Med. Assoc.* 181, 139–142.
15. Chim, C.S., 2017. Lactate dehydrogenase as a prognostic marker in AL amyloidosis: expected or unexpected? *Br J Haematol* 178, 833–835. <https://doi.org/10.1111/bjh.14835>
16. Choufani, E.B., Sanchowala, V., Ernst, T., Quillen, K., Skinner, M., Wright, D.G., Seldin, D.C., 2001. Acquired factor X deficiency in patients with amyloid light-chain amyloidosis: incidence, bleeding manifestations, and response to high-dose chemotherapy. *Blood* 97, 1885–1887. <https://doi.org/10.1182/blood.v97.6.1885>
17. Clark, E.A., Brugge, J.S., 1995. Integrins and signal transduction pathways: the road taken. *Science* 268, 233–239. <https://doi.org/10.1126/science.7716514>
18. Coccetti, P., Tripodi, F., Tedeschi, G., Nonnis, S., Marin, O., Fantinato, S., Cirulli, C., Vanoni, M., Alberghina, L., 2008. The CK2 phosphorylation of catalytic domain of Cdc34 modulates its activity at the G₁ to S transition in *Saccharomyces cerevisiae*. *Cell Cycle* 7, 1391–1401. <https://doi.org/10.4161/cc.7.10.5825>
19. Comenzo, R.L., Zhou, P., Fleisher, M., Clark, B., Teruya-Feldstein, J., 2006. Seeking confidence in the diagnosis of systemic AL (Ig light-chain) amyloidosis: patients can have both monoclonal gammopathies and hereditary amyloid proteins. *Blood* 107, 3489–3491. <https://doi.org/10.1182/blood-2005-10-4148>
20. Cong, W., Zhang, X.-X., He, J.-J., Li, F.-C., Elsheikha, H.M., Zhu, X.-Q., 2017. Global miRNA expression profiling of domestic cat livers following acute *Toxoplasma gondii* infection. *Oncotarget* 8. <https://doi.org/10.18632/oncotarget.16108>
21. Cox, J., Mann, M., 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26, 1367–1372. <https://doi.org/10.1038/nbt.1511>

22. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 1000 Genomes Project Analysis Group, 2011. The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
23. DiBartola, S.P., Benson, M.D., Dwulet, F.E., Cornacoff, J.B., 1985. Isolation and characterization of amyloid protein AA in the Abyssinian cat. *Lab. Invest.* 52, 485–489.
24. DiBartola, S.P., Tarr, M.J., Webb, D.M., Giger, U., 1990. Familial renal amyloidosis in Chinese Shar Pei dogs. *J. Am. Vet. Med. Assoc.* 197, 483–487.
25. Drummond, E.S., Harvey, A.R., Martins, R.N., 2009. Androgens and Alzheimer's disease: Current Opinion in Endocrinology, Diabetes and Obesity 16, 254–259. <https://doi.org/10.1097/MED.0b013e32832b101f>
26. Du, S., Zhang, Y., Yang, J., Liu, X., Wang, Y., Xu, B., Jia, J., 2019. Curcumin Alleviates β Amyloid-Induced Neurotoxicity in HT22 Cells via Upregulating SOD2. *J Mol Neurosci* 67, 540–549. <https://doi.org/10.1007/s12031-019-01267-2>
27. Dweep, H., Sticht, C., Gretz, N., 2013. In-Silico Algorithms for the Screening of Possible microRNA Binding Sites and Their Interactions. *CG* 14, 127–136. <https://doi.org/10.2174/1389202911314020005>
28. Elitok, O.M., Elitok, B., Unver, O., 2008. Renal Amyloidosis in Cattle with Inflammatory Diseases. *Journal of Veterinary Internal Medicine* 22, 450–455. <https://doi.org/10.1111/j.1939-1676.2008.0059.x>
29. Eskelinen, E.-L., Illert, A.L., Tanaka, Y., Schwarzmann, G., Blanz, J., von Figura, K., Saftig, P., 2002. Role of LAMP-2 in Lysosome Biogenesis and Autophagy. *MBoC* 13, 3355–3368. <https://doi.org/10.1091/mbc.e02-02-0114>
30. Exley, C., 1997. ATP-promoted amyloidosis of an amyloid beta peptide. *Neuroreport* 8, 3411–3414. <https://doi.org/10.1097/00001756-199710200-00043>
31. Fan, Y., Siklenka, K., Arora, S.K., Ribeiro, P., Kimmins, S., Xia, J., 2016. miRNet - dissecting miRNA-target interactions and functional associations through network-based visual analysis. *Nucleic Acids Res* 44, W135–W141. <https://doi.org/10.1093/nar/gkw288>
32. Feng, B., Hu, P., Chen, J., Liu, Q., Li, X., Du, Y., 2015. Analysis of Differentially Expressed Genes Associated With Alzheimer's Disease Based on Bioinformatics Methods. *Am J Alzheimers Dis Other Dement* 30, 746–751. <https://doi.org/10.1177/1533317514537548>
33. Friedländer, M.R., Mackowiak, S.D., Li, N., Chen, W., Rajewsky, N., 2012. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in

- seven animal clades. *Nucleic Acids Research* 40, 37–52. <https://doi.org/10.1093/nar/gkr688>
34. Frost, B., Bardai, F.H., Feany, M.B., 2016. Lamin Dysfunction Mediates Neurodegeneration in Tauopathies. *Curr. Biol.* 26, 129–136. <https://doi.org/10.1016/j.cub.2015.11.039>
35. Furie, B., Greene, E., Furie, B.C., 1977. Syndrome of acquired factor X deficiency and systemic amyloidosis in vivo studies of the metabolic fate of factor X. *N. Engl. J. Med.* 297, 81–85. <https://doi.org/10.1056/NEJM197707142970203>
36. Galli, A., Maffioli, E., Sogne, E., Moretti, S., Di Cairano, E.S., Negri, A., Nonnis, S., Norata, G.D., Bonacina, F., Borghi, F., Podestà, A., Bertuzzi, F., Milani, P., Lenardi, C., Tedeschi, G., Perego, C., 2018. Cluster-assembled zirconia substrates promote long-term differentiation and functioning of human islets of Langerhans. *Sci Rep* 8, 9979. <https://doi.org/10.1038/s41598-018-28019-3>
37. Gandolfi, B., Grahn, R.A., Creighton, E.K., Williams, D.C., Dickinson, P.J., Sturges, B.K., Guo, L.T., Shelton, G.D., Leegwater, P.A.J., Longeri, M., Malik, R., Lyons, L.A., 2015. COLQ variant associated with Devon Rex and Sphynx feline hereditary myopathy. *Anim Genet* 46, 711–715. <https://doi.org/10.1111/age.12350>
38. Gandolfi, B., Gruffydd-Jones, T.J., Malik, R., Cortes, A., Jones, B.R., Helps, C.R., Prinzenberg, E.M., Erhardt, G., Lyons, L.A., 2012. First WNK4-Hypokalemia Animal Model Identified by Genome-Wide Association in Burmese Cats. *PLoS ONE* 7, e53173. <https://doi.org/10.1371/journal.pone.0053173>
39. Garner, M.M., Raymond, J.T., O'Brien, T.D., Nordhausen, R.W., Russell, W.C., 2007. AMYLOIDOSIS IN THE BLACK-FOOTED FERRET (*MUSTELA NIGRIPES*). *Journal of Zoo and Wildlife Medicine* 38, 32–41. <https://doi.org/10.1638/06-023.1>
40. Gastineau, D.A., Gertz, M.A., Daniels, T.M., Kyle, R.A., Bowie, E.J., 1991. Inhibitor of the thrombin time in systemic amyloidosis: a common coagulation abnormality. *Blood* 77, 2637–2640.
41. Goodman, A.B., Pardee, A.B., 2003. Evidence for defective retinoid transport and function in late onset Alzheimer's disease. *Proceedings of the National Academy of Sciences* 100, 2901–2905. <https://doi.org/10.1073/pnas.0437937100>
42. Greipp, P.R., Kyle, R.A., Bowie, E.J.W., 1981. Factor-X deficiency in amyloidosis: A critical review. *Am. J. Hematol.* 11, 443–450. <https://doi.org/10.1002/ajh.2830110414>
43. Griffiths-Jones, S., 2006. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Research* 34, D140–D144. <https://doi.org/10.1093/nar/gkj112>

44. Gruys, E., 1977. Amyloidosis in the bovine kidney. *Vet Res Commun* 1, 265–276. <https://doi.org/10.1007/BF02267657>
45. Guijarro, J.I., Sunde, M., Jones, J.A., Campbell, I.D., Dobson, C.M., 1998. Amyloid fibril formation by an SH3 domain. *Proceedings of the National Academy of Sciences* 95, 4224–4228. <https://doi.org/10.1073/pnas.95.8.4224>
46. Guo, J.-T., Aldrich, C.E., Mason, W.S., Pugh, J.C., 1996. Characterization of serum amyloid A protein mRNA expression and secondary amyloidosis in the domestic duck. *Proceedings of the National Academy of Sciences* 93, 14548–14553. <https://doi.org/10.1073/pnas.93.25.14548>
47. Higaki, S., Muramatsu, M., Matsuda, A., Matsumoto, K., Satoh, J., Michikawa, M., Niida, S., 2018. Defensive effect of microRNA-200b/c against amyloid-beta peptide-induced toxicity in Alzheimer's disease models. *PLoS ONE* 13, e0196929. <https://doi.org/10.1371/journal.pone.0196929>
48. Holub, D., Flodrova, P., Pika, T., Flodr, P., Hajduch, M., Dzubak, P., 2019. Mass Spectrometry Amyloid Typing Is Reproducible across Multiple Organ Sites. *BioMed Research International* 2019, 1–9. <https://doi.org/10.1155/2019/3689091>
49. Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4, 44–57. <https://doi.org/10.1038/nprot.2008.211>
50. Ichii, O., Otsuka, S., Ohta, H., Yabuki, A., Horino, T., Kon, Y., 2014. MicroRNA expression profiling of cat and dog kidneys. *Research in Veterinary Science* 96, 299–303. <https://doi.org/10.1016/j.rvsc.2014.01.003>
51. Jarrett, J.T., Lansbury, P.T., 1992. Amyloid fibril formation requires a chemically discriminating nucleation event: studies of an amyloidogenic sequence from the bacterial protein OsmB. *Biochemistry* 31, 12345–12352. <https://doi.org/10.1021/bi00164a008>
52. Jin, Y., Tu, Q., Liu, M., 2018. MicroRNA-125b regulates Alzheimer's disease through SphK1 regulation. *Mol Med Report*. <https://doi.org/10.3892/mmr.2018.9156>
53. Johnson, K.H., Sletten, K., Werdin, R.E., Westermark, G.T., O'Brien, T.D., Westermark, P., 1989. Amino acid sequence variations in protein AA of cats with high and low incidences of AA amyloidosis. *Comp. Biochem. Physiol., B* 94, 765–768. [https://doi.org/10.1016/0305-0491\(89\)90162-4](https://doi.org/10.1016/0305-0491(89)90162-4)
54. Katara, P., Gautam, B., Kuntal, H., Sharma, V., 2010. Prediction of miRNA targets, affected proteins and their homologs in *Glycine max*. *Bioinformation* 5, 162–165. <https://doi.org/10.6026/97320630005162>

55. Kim, D.Y., Taylor, H.W., Eades, S.C., Cho, D.-Y., 2005. Systemic AL Amyloidosis Associated with Multiple Myeloma in a Horse. *Vet Pathol* 42, 81–84. <https://doi.org/10.1354/vp.42-1-81>
56. Klein, C.J., Vrana, J.A., Theis, J.D., Dyck, P.J., Dyck, P.J.B., Spinner, R.J., Mauermann, M.L., Bergen, H.R., Zeldenrust, S.R., Dogan, A., 2011. Mass Spectrometric–Based Proteomic Analysis of Amyloid Neuropathy Type in Nerve Tissue. *Arch Neurol* 68. <https://doi.org/10.1001/archneurol.2010.261>
57. Kluve-Beckerman, B., Dwulet, F.E., DiBartolna, S.P., Benson, M.D., 1989. Primary structures of dog and cat amyloid A proteins: comparison to human AA. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry* 94, 175–183. [https://doi.org/10.1016/0305-0491\(89\)90030-8](https://doi.org/10.1016/0305-0491(89)90030-8)
58. Kumar, S., Reddy, P.H., 2016. Are circulating microRNAs peripheral biomarkers for Alzheimer’s disease? *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1862, 1617–1627. <https://doi.org/10.1016/j.bbadis.2016.06.001>
59. Kurkinen, K.M.A., Marttinen, M., Turner, L., Natunen, T., Mäkinen, P., Haapalinna, F., Sarajärvi, T., Gabbouj, S., Kurki, M., Paananen, J., Koivisto, A.M., Rauramaa, T., Leinonen, V., Tanila, H., Soininen, H., Lucas, F.R., Haapasalo, A., Hiltunen, M., 2016. SEPT8 modulates β -amyloidogenic processing of APP by affecting the sorting and accumulation of BACE1. *J Cell Sci* 129, 2224–2238. <https://doi.org/10.1242/jcs.185215>
60. Lachmann, H.J., Booth, D.R., Booth, S.E., Bybee, A., Gilbertson, J.A., Gillmore, J.D., Pepys, M.B., Hawkins, P.N., 2002. Misdiagnosis of hereditary amyloidosis as AL (primary) amyloidosis. *N. Engl. J. Med.* 346, 1786–1791. <https://doi.org/10.1056/NEJMoa013354>
61. Laganà, A., Dirksen, W.P., Supsavhad, W., Yilmaz, A.S., Ozer, H.G., Feller, J.D., Vala, K.A., Croce, C.M., Rosol, T.J., 2017. Discovery and characterization of the feline miRNAome. *Sci Rep* 7, 9263. <https://doi.org/10.1038/s41598-017-10164-w>
62. Lavatelli, F., Vrana, J.A., 2011. Proteomic typing of amyloid deposits in systemic amyloidoses. *Amyloid* 18, 177–182. <https://doi.org/10.3109/13506129.2011.630762>
63. Leung, N., Nasr, S.H., Sethi, S., 2012. How I treat amyloidosis: the importance of accurate diagnosis and amyloid typing. *Blood* 120, 3206–3213. <https://doi.org/10.1182/blood-2012-03-413682>
64. Li, Kuo, Zhang, Junling, Ji, Chunxue, Wang, Lixuan, 2016. MiR-144-3p and Its Target Gene β -Amyloid Precursor Protein Regulate 1-Methyl-4-Phenyl-1,2-3,6-Tetrahydropyridine-Induced Mitochondrial Dysfunction. *Molecules and Cells* 39, 543–549. <https://doi.org/10.14348/MOLCELLS.2016.0050>

65. Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., Sandstrom, R., Bernstein, B., Bender, M.A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L.A., Lander, E.S., Dekker, J., 2009. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* 326, 289–293. <https://doi.org/10.1126/science.1181369>
66. Linke, R.P., 2012. On Typing Amyloidosis Using Immunohistochemistry. Detailed Illustrations, Review and a Note on Mass Spectrometry. *Progress in Histochemistry and Cytochemistry* 47, 61–132. <https://doi.org/10.1016/j.proghi.2012.03.001>
67. Liu, L., Komatsu, H., Murray, I.V.J., Axelsen, P.H., 2008. Promotion of amyloid beta protein misfolding and fibrillogenesis by a lipid oxidation product. *J. Mol. Biol.* 377, 1236–1250. <https://doi.org/10.1016/j.jmb.2008.01.057>
68. Maffioli, E., Nonnis, S., Angioni, R., Santagata, F., Calì, B., Zanotti, L., Negri, A., Viola, A., Tedeschi, G., 2017. Proteomic analysis of the secretome of human bone marrow-derived mesenchymal stem cells primed by pro-inflammatory cytokines. *Journal of Proteomics* 166, 115–126. <https://doi.org/10.1016/j.jprot.2017.07.012>
69. Maffioli, E., Schulte, C., Nonnis, S., Grassi Scalvini, F., Piazzoni, C., Lenardi, C., Negri, A., Milani, P., Tedeschi, G., 2018. Proteomic Dissection of Nanotopography-Sensitive Mechanotransductive Signaling Hubs that Foster Neuronal Differentiation in PC12 Cells. *Front. Cell. Neurosci.* 11, 417. <https://doi.org/10.3389/fncel.2017.00417>
70. Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j.* 17, 10. <https://doi.org/10.14806/ej.17.1.200>
71. Martinez, B., Peplow, P., 2019. MicroRNAs as diagnostic and therapeutic tools for Alzheimer's disease: advances and limitations. *Neural Regen Res* 14, 242. <https://doi.org/10.4103/1673-5374.244784>
72. Mauler, D.A., Gandolfi, B., Reinero, C.R., O'Brien, D.P., Spooner, J.L., Lyons, L.A., and 99 Lives Consortium, 2017. Precision Medicine in Cats: Novel Niemann-Pick Type C1 Diagnosed by Whole-Genome Sequencing. *J Vet Intern Med* 31, 539–544. <https://doi.org/10.1111/jvim.14599>
73. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., Shafer, A., Neri, F., Lee, K., Kuttyavin, T., Stehling-Sun, S., Johnson, A.K., Canfield, T.K., Giste, E., Diegel, M., Bates, D., Hansen, R.S., Neph, S., Sabo, P.J., Heimfeld, S., Raubitschek, A., Ziegler, S., Cotsapas, C., Sotoodehnia, N., Glass, I., Sunyaev, S.R., Kaul, R., Stamatoyannopoulos, J.A., 2012. Systematic Localization of Common Disease-

- Associated Variation in Regulatory DNA. *Science* 337, 1190–1195. <https://doi.org/10.1126/science.1222794>
74. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., Cunningham, F., 2016. The Ensembl Variant Effect Predictor. *Genome Biol* 17, 122. <https://doi.org/10.1186/s13059-016-0974-4>
 75. Menotti-Raymond, M., O'Brien, S.J., 1993. Dating the genetic bottleneck of the African cheetah. *Proceedings of the National Academy of Sciences* 90, 3172–3176. <https://doi.org/10.1073/pnas.90.8.3172>
 76. Ménsua, C., Carrasco, L., Bautista, M.J., Biescas, E., Fernández, A., Murphy, C.L., Weiss, D.T., Solomon, A., Luján, L., 2003. Pathology of AA Amyloidosis in Domestic Sheep and Goats. *Vet Pathol* 40, 71–80. <https://doi.org/10.1354/vp.40-1-71>
 77. Merlini, G., Planté-Bordeneuve, V., Judge, D.P., Schmidt, H., Obici, L., Perlini, S., Packman, J., Tripp, T., Grogan, D.R., 2013. Effects of Tafamidis on Transthyretin Stabilization and Clinical Outcomes in Patients with Non-Val30Met Transthyretin Amyloidosis. *J. of Cardiovasc. Trans. Res.* 6, 1011–1020. <https://doi.org/10.1007/s12265-013-9512-x>
 78. Mi, H., Muruganujan, A., Casagrande, J.T., Thomas, P.D., 2013. Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc* 8, 1551–1566. <https://doi.org/10.1038/nprot.2013.092>
 79. Moszyńska, A., Gebert, M., Collawn, J.F., Bartoszewski, R., 2017. SNPs in microRNA target sites and their potential role in human disease. *Open Biol.* 7, 170019. <https://doi.org/10.1098/rsob.170019>
 80. Muchtar, E., Dispenzieri, A., Lacy, M.Q., Buadi, F.K., Kapoor, P., Hayman, S.R., Gonsalves, W., Warsame, R., Kourelis, T.V., Chakraborty, R., Russell, S., Lust, J.A., Lin, Y., Go, R.S., Zeldenrust, S., Dingli, D., Leung, N., Rajkumar, S.V., Kyle, R.A., Kumar, S.K., Gertz, M.A., 2017. Elevation of serum lactate dehydrogenase in AL amyloidosis reflects tissue damage and is an adverse prognostic marker in patients not eligible for stem cell transplantation. *Br J Haematol* 178, 888–895. <https://doi.org/10.1111/bjh.14830>
 81. Murphy, C.L., Wang, S., Kestler, D., Larsen, C., Benson, D., Weiss, D.T., Solomon, A., 2010. Leukocyte chemotactic factor 2 (LECT2)-associated renal amyloidosis: a case series. *Am. J. Kidney Dis.* 56, 1100–1107. <https://doi.org/10.1053/j.ajkd.2010.08.013>
 82. Murphy, C.L., Wang, S., Williams, T., Weiss, D.T., Solomon, A., 2006. Characterization of Systemic Amyloid Deposits by Mass Spectrometry, in: *Methods*

- in Enzymology. Elsevier, pp. 48–62. [https://doi.org/10.1016/S0076-6879\(06\)12004-2](https://doi.org/10.1016/S0076-6879(06)12004-2)
83. Narfström, K., 1983. Hereditary progressive retinal atrophy in the Abyssinian cat. *Journal of Heredity* 74, 273–276. <https://doi.org/10.1093/oxfordjournals.jhered.a109782>
 84. Niewold, T.A., van der Linde-Sipman, J.S., Murphy, C., Tooten, P.C., Gruys, E., 1999. Familial amyloidosis in cats: Siamese and Abyssinian AA proteins differ in primary sequence and pattern of deposition. *Amyloid* 6, 205–209.
 85. Oh, A., Pearce, J.W., Gandolfi, B., Creighton, E.K., Suedmeyer, W.K., Selig, M., Bosiack, A.P., Castaner, L.J., Whiting, R.E.H., Belknap, E.B., Lyons, L.A., and the 99 Lives Consortium, 2017. Early-Onset Progressive Retinal Atrophy Associated with an IQCB1 Variant in African Black-Footed Cats (*Felis nigripes*). *Sci Rep* 7, 43918. <https://doi.org/10.1038/srep43918>
 86. Østevik, L., Gunnes, G., de Souza, G.A., Wien, T.N., Sørby, R., 2014. Nasal and ocular amyloidosis in a 15-year-old horse. *Acta Vet Scand* 56, 50. <https://doi.org/10.1186/s13028-014-0050-6>
 87. Ovelgönne, J.H., Landman, W.J., Gruys, E., Gielkens, A.L., Peeters, B.P., 2001. Identical amyloid precursor proteins in two breeds of chickens which differ in susceptibility to develop amyloid arthropathy. *Amyloid* 8, 41–51.
 88. Paltrinieri, S., Sironi, G., Giori, L., Faverzani, S., Longeri, M., 2015. Changes in Serum and Urine SAA Concentrations and Qualitative and Quantitative Proteinuria in Abyssinian Cats with Familial Amyloidosis: A Five-year Longitudinal Study (2009-2014). *J Vet Intern Med* 29, 505–512. <https://doi.org/10.1111/jvim.12561>
 89. Papendick, R.E., Munson, L., O'Brien, T.D., Johnson, K.H., 1997. Systemic AA Amyloidosis in Captive Cheetahs (*Acinonyx jubatus*). *Vet Pathol* 34, 549–556. <https://doi.org/10.1177/030098589703400602>
 90. Patel, N., Hoang, D., Miller, N., Ansaloni, S., Huang, Q., Rogers, J.T., Lee, J.C., Saunders, A.J., 2008. MicroRNAs can regulate human APP levels. *Mol Neurodegeneration* 3, 10. <https://doi.org/10.1186/1750-1326-3-10>
 91. Pathak, B.K., Mondal, S., Banerjee, S., Ghosh, A.N., Barat, C., 2017. Sequestration of Ribosome during Protein Aggregate Formation: Contribution of ribosomal RNA. *Sci Rep* 7, 42017. <https://doi.org/10.1038/srep42017>
 92. Pedersen, N.C., 1992. Inflammatory Oral Cavity Diseases of the Cat. *Veterinary Clinics of North America: Small Animal Practice* 22, 1323–1345. [https://doi.org/10.1016/S0195-5616\(92\)50130-9](https://doi.org/10.1016/S0195-5616(92)50130-9)

93. Picken, M.M., 2010. Amyloidosis-where are we now and where are we heading? Arch. Pathol. Lab. Med. 134, 545–551. <https://doi.org/10.1043/1543-2165-134.4.545>
94. Picken, M.M., 2007. New insights into systemic amyloidosis: the importance of diagnosis of specific type. Curr. Opin. Nephrol. Hypertens. 16, 196–203. <https://doi.org/10.1097/MNH.0b013e3280bdc0db>
95. Platz, S.J., Breuer, W., Geisel, O., Linke, R.P., Hermanns, W., 1997. Identification of λ light chain amyloid in eight canine and two feline extramedullary plasmacytomas. Journal of Comparative Pathology 116, 45–54. [https://doi.org/10.1016/S0021-9975\(97\)80042-4](https://doi.org/10.1016/S0021-9975(97)80042-4)
96. Pontén, F., Jirström, K., Uhlen, M., 2008. The Human Protein Atlas—a tool for pathology. J. Pathol. 216, 387–393. <https://doi.org/10.1002/path.2440>
97. Pontius, J.U., Mullikin, J.C., Smith, D.R., Agencourt Sequencing Team, Lindblad-Toh, K., Gnerre, S., Clamp, M., Chang, J., Stephens, R., Neelam, B., Volfovsky, N., Schaffer, A.A., Agarwala, R., Narfstrom, K., Murphy, W.J., Giger, U., Roca, A.L., Antunes, A., Menotti-Raymond, M., Yuhki, N., Pecon-Slaterry, J., Johnson, W.E., Bourque, G., Tesler, G., NISC Comparative Sequencing Program, O'Brien, S.J., 2007. Initial sequence and comparative analysis of the cat genome. Genome Research 17, 1675–1689. <https://doi.org/10.1101/gr.6380007>
98. Puchtler, H., Sweat, F., Levine, M., 1962. ON THE BINDING OF CONGO RED BY AMYLOID. J Histochem Cytochem. 10, 355–364. <https://doi.org/10.1177/10.3.355>
99. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., Sham, P.C., 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics 81, 559–575. <https://doi.org/10.1086/519795>
100. Qian, Q., Zhang, J., He, F.-P., Bao, W.-X., Zheng, T.-T., Zhou, D.-M., Pan, H.-Y., Zhang, H., Zhang, X.-Q., He, X., Sun, B.-G., Luo, B.-Y., Chen, C., Peng, G.-P., 2019. Down-regulated expression of microRNA-338-5p contributes to neuropathology in Alzheimer's disease. FASEB J. 33, 4404–4417. <https://doi.org/10.1096/fj.201801846R>
101. Rideout, B.A., Montali, R.J., Wallace, R.S., Bush, M., Phillips, L.G., Antonovych, T.T., Sabnis, S.G., 1989. Renal Medullary Amyloidosis in Dorcas Gazelles. Vet Pathol 26, 129–135. <https://doi.org/10.1177/030098588902600205>
102. Riffo-Campos, Á., Riquelme, I., Brebi-Mieville, P., 2016. Tools for Sequence-Based miRNA Target Prediction: What to Choose? IJMS 17, 1987. <https://doi.org/10.3390/ijms17121987>

103. Rowczenio, D., Dogan, A., Theis, J.D., Vrana, J.A., Lachmann, H.J., Wechalekar, A.D., Gilbertson, J.A., Hunt, T., Gibbs, S.D.J., Sattianayagam, P.T., Pinney, J.H., Hawkins, P.N., Gillmore, J.D., 2011. Amyloidogenicity and Clinical Phenotype Associated with Five Novel Mutations in Apolipoprotein A-I. *The American Journal of Pathology* 179, 1978–1987. <https://doi.org/10.1016/j.ajpath.2011.06.024>
104. Saito, S., Ando, Y., Nakamura, M., Ueda, M., Kim, J., Ishima, Y., Akaike, T., Otagiri, M., 2005. Effect of nitric oxide in amyloid fibril formation on transthyretin-related amyloidosis. *Biochemistry* 44, 11122–11129. <https://doi.org/10.1021/bi050327i>
105. Schroeder, F.M., Miller, F.J., Nelson, J.A., Rankin, R.S., 1978. Gastrointestinal angiographic findings in systemic amyloidosis. *AJR Am J Roentgenol* 131, 143–146. <https://doi.org/10.2214/ajr.131.1.143>
106. Schulte, C., Ripamonti, M., Maffioli, E., Cappelluti, M.A., Nonnis, S., Puricelli, L., Lamanna, J., Piazzoni, C., Podestà, A., Lenardi, C., Tedeschi, G., Malgaroli, A., Milani, P., 2016. Scale Invariant Disordered Nanotopography Promotes Hippocampal Neuron Development and Maturation with Involvement of Mechanotransductive Pathways. *Front. Cell. Neurosci.* 10. <https://doi.org/10.3389/fncel.2016.00267>
107. Schulze, C., Brüggmann, M., Böer, M., Brandt, H.-P., Pohlenz, J., Linke, R.P., 1998. Generalized AA-amyloidosis in Siberian Tigers (*Panthera tigris altaica*) with Predominant Renal Medullary Amyloid Deposition. *Vet Pathol* 35, 70–74. <https://doi.org/10.1177/030098589803500108>
108. Segev, G., Cowgill, L.D., Jessen, S., Berkowitz, A., Mohr, C.F., Aroch, I., 2012. Renal Amyloidosis in Dogs: A Retrospective Study of 91 Cases with Comparison of the Disease between Shar-Pei and Non-Shar-Pei Dogs. *J Vet Intern Med* 26, 259–268. <https://doi.org/10.1111/j.1939-1676.2011.00878.x>
109. Sethi, S., Theis, J.D., Leung, N., Dispenzieri, A., Nasr, S.H., Fidler, M.E., Cornell, L.D., Gamez, J.D., Vrana, J.A., Dogan, A., 2010. Mass spectrometry-based proteomic diagnosis of renal immunoglobulin heavy chain amyloidosis. *Clin J Am Soc Nephrol* 5, 2180–2187. <https://doi.org/10.2215/CJN.02890310>
110. Sethi, S., Theis, J.D., Shiller, S.M., Nast, C.C., Harrison, D., Rennke, H.G., Vrana, J.A., Dogan, A., 2012a. Medullary amyloidosis associated with apolipoprotein A-IV deposition. *Kidney Int.* 81, 201–206. <https://doi.org/10.1038/ki.2011.316>
111. Sethi, S., Vrana, J.A., Theis, J.D., Leung, N., Sethi, A., Nasr, S.H., Fervenza, F.C., Cornell, L.D., Fidler, M.E., Dogan, A., 2012b. Laser microdissection and mass spectrometry-based proteomics aids the diagnosis and typing of renal amyloidosis. *Kidney International* 82, 226–234. <https://doi.org/10.1038/ki.2012.108>

112. Sierksma, A., Lu, A., Salta, E., Vanden Eynden, E., Callaerts-Vegh, Z., D'Hooge, R., Blum, D., Buée, L., Fiers, M., De Strooper, B., 2018. Deregulation of neuronal miRNAs induced by amyloid- β or TAU pathology. *Mol Neurodegeneration* 13, 54. <https://doi.org/10.1186/s13024-018-0285-1>
113. Simons, J.P., Al-Shawi, R., Ellmerich, S., Speck, I., Aslam, S., Hutchinson, W.L., Mangione, P.P., Disterer, P., Gilbertson, J.A., Hunt, T., Millar, D.J., Minogue, S., Bodin, K., Pepys, M.B., Hawkins, P.N., 2013. Pathogenetic mechanisms of amyloid A amyloidosis. *Proceedings of the National Academy of Sciences* 110, 16115–16120. <https://doi.org/10.1073/pnas.1306621110>
114. Sipe, J.D., Benson, M.D., Buxbaum, J.N., Ikeda, S., Merlini, G., Saraiva, M.J.M., Westermark, P., 2016. Amyloid fibril proteins and amyloidosis: chemical identification and clinical classification International Society of Amyloidosis 2016 Nomenclature Guidelines. *Amyloid* 23, 209–213. <https://doi.org/10.1080/13506129.2016.1257986>
115. Sipe, J.D., Benson, M.D., Buxbaum, J.N., Ikeda, S., Merlini, G., Saraiva, M.J.M., Westermark, P., 2014. Nomenclature 2014: Amyloid fibril proteins and clinical classification of the amyloidosis. *Amyloid* 21, 221–224. <https://doi.org/10.3109/13506129.2014.964858>
116. Sipe, J.D., Benson, M.D., Buxbaum, J.N., Ikeda, S., Merlini, G., Saraiva, M.J.M., Westermark, P., 2012. Amyloid fibril protein nomenclature: 2012 recommendations from the Nomenclature Committee of the International Society of Amyloidosis. *Amyloid* 19, 167–170. <https://doi.org/10.3109/13506129.2012.734345>
117. Solomon, A., Murphy, C.L., Westermark, P., 2008. Unreliability of immunohistochemistry for typing amyloid deposits. *Arch. Pathol. Lab. Med.* 132, 14; author reply 14-15. [https://doi.org/10.1043/1543-2165\(2008\)132\[14b:IR\]2.0.CO;2](https://doi.org/10.1043/1543-2165(2008)132[14b:IR]2.0.CO;2)
118. Sucker, C., Hetzel, G., Grabensee, B., Stocksclaeder, M., Scharf, R., 2006. Amyloidosis and Bleeding: Pathophysiology, Diagnosis, and Therapy. *American Journal of Kidney Diseases* 47, 947–955. <https://doi.org/10.1053/j.ajkd.2006.03.036>
119. Sun, J., Wang, J., Wang, S., Yuan, D., Birame, B.M., Li, Z., Yi, B., Liu, W., 2014. MicroRNA profile analysis of a feline kidney cell line before and after infection with mink enteritis virus. *Gene* 539, 224–229. <https://doi.org/10.1016/j.gene.2014.01.074>
120. Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., Kuhn, M., Bork, P., Jensen, L.J., von Mering, C., 2015. STRING v10: protein–protein interaction networks,

- integrated over the tree of life. *Nucleic Acids Research* 43, D447–D452.
<https://doi.org/10.1093/nar/gku1003>
121. Tak, Y.G., Farnham, P.J., 2015. Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. *Epigenetics & Chromatin* 8, 57.
<https://doi.org/10.1186/s13072-015-0050-4>
 122. Tedeschi, G., Pagliato, L., Negroni, M., Montorfano, G., Corsetto, P., Nonnis, S., Negri, A., Rizzo, A.M., 2011. Protein pattern of *Xenopus laevis* embryos grown in simulated microgravity. *Cell Biol. Int.* 35, 249–258.
<https://doi.org/10.1042/CBI20100248>
 123. Terio, K.A., O'Brien, T., Lamberski, N., Famula, T.R., Munson, L., 2008. Amyloidosis in Black-footed Cats (*Felis nigripes*). *Vet Pathol* 45, 393–400.
<https://doi.org/10.1354/vp.45-3-393>
 124. van der Linde-Sipman, J.S., Niewold, T.A., Tooten, P.C.J., de Neijls-Backer, M., Gruys, E., 1997. Generalized AA-amyloidosis in Siamese and Oriental cats. *Veterinary Immunology and Immunopathology* 56, 1–10.
[https://doi.org/10.1016/S0165-2427\(96\)05717-0](https://doi.org/10.1016/S0165-2427(96)05717-0)
 125. van Rossum, M., van Asten, F.J., Rofina, J., Lenstra, J.A., Benson, M.D., Gruys, E., 2004. Analysis of cDNA sequences of feline SAAs. *Amyloid* 11, 38–43.
 126. Vernocchi, V., Morselli, M.G., Varesi, S., Nonnis, S., Maffioli, E., Negri, A., Tedeschi, G., Luvoni, G.C., 2014. Sperm ubiquitination in epididymal feline semen. *Theriogenology* 82, 636–642.
<https://doi.org/10.1016/j.theriogenology.2014.06.002>
 127. Villa, C., Fenoglio, C., De Riz, M., Clerici, F., Marcone, A., Benussi, L., Ghidoni, R., Gallone, S., Cortini, F., Serpente, M., Cantoni, C., Fumagalli, G., Boneschi, F.M., Cappa, S., Binetti, G., Franceschi, M., Rainero, I., Giordana, M.T., Mariani, C., Bresolin, N., Scarpini, E., Galimberti, D., 2011. Role of *hnRNP-A1* and miR-590-3p in Neuronal Death: Genetics and Expression Analysis in Patients with Alzheimer Disease and Frontotemporal Lobar Degeneration. *Rejuvenation Research* 14, 275–281. <https://doi.org/10.1089/rej.2010.1123>
 128. Vizcaíno, J.A., Csordas, A., del-Toro, N., Dianes, J.A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., Xu, Q.-W., Wang, R., Hermjakob, H., 2016. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* 44, D447–456. <https://doi.org/10.1093/nar/gkv1145>
 129. Vrana, J.A., Gamez, J.D., Madden, B.J., Theis, J.D., Bergen, H.R., Dogan, A., 2009. Classification of amyloidosis by laser microdissection and mass

- spectrometry-based proteomic analysis in clinical biopsy specimens. *Blood* 114, 4957–4959. <https://doi.org/10.1182/blood-2009-07-230722>
130. Wang, P.-L., Niidome, T., Akaike, A., Kihara, T., Sugimoto, H., 2009. Rac1 inhibition negatively regulates transcriptional activity of the amyloid precursor protein gene. *J. Neurosci. Res.* 87, 2105–2114. <https://doi.org/10.1002/jnr.22039>
131. Warren, M., 2018. The approach to predictive medicine that is taking genomics research by storm. *Nature* 562, 181–183. <https://doi.org/10.1038/d41586-018-06956-3>
132. Washburn, M.P., 2011. Driving biochemical discovery with quantitative proteomics. *Trends in Biochemical Sciences* 36, 170–177. <https://doi.org/10.1016/j.tibs.2010.09.001>
133. Wasik, A.A., Dash, S.N., Lehtonen, S., 2019. Septins in kidney: A territory little explored. *Cytoskeleton* 76, 154–162. <https://doi.org/10.1002/cm.21477>
134. Weber, K., Rostert, N., Bauersachs, S., Wess, G., 2015. Serum microRNA profiles in cats with hypertrophic cardiomyopathy. *Mol Cell Biochem* 402, 171–180. <https://doi.org/10.1007/s11010-014-2324-8>
135. Weids, A.J., Grant, C.M., 2014. The yeast peroxiredoxin Tsa1 protects against protein-aggregate-induced oxidative stress. *Journal of Cell Science* 127, 1327–1335. <https://doi.org/10.1242/jcs.144022>
136. Weng, L., Spencer, B.H., SoohHoo, P.T., Connors, L.H., O'Hara, C.J., Seldin, D.C., 2011. Dysregulation of miRNAs in AL amyloidosis. *Amyloid* 18, 128–135. <https://doi.org/10.3109/13506129.2011.588977>
137. Williams, J.H., Van Wilpe, E., Momberg, M., 2005. Renal medullary AA amyloidosis, hepatocyte dissociation and multinucleated hepatocytes in a 14-year-old free-ranging lioness (*Panthera leo*). *J S Afr Vet Assoc* 76, 90–98. <https://doi.org/10.4102/jsava.v76i2.404>
138. Witkowski, A., Chan, G.K.L., Boatz, J.C., Li, N.J., Inoue, A.P., Wong, J.C., van der Wel, P.C.A., Cavigliolo, G., 2018. Methionine oxidized apolipoprotein A-I at the crossroads of HDL biogenesis and amyloid formation. *The FASEB Journal* 32, 3149–3165. <https://doi.org/10.1096/fj.201701127R>
139. Zhang, B., Une, Y., Ge, F., Fu, X., Qian, J., Zhang, P., Sawashita, J., Higuchi, K., Mori, M., 2008. Characterization of the Cheetah Serum Amyloid A1 Gene: Critical Role and Functional Polymorphism of a Cis-Acting Element. *Journal of Heredity* 99, 355–363. <https://doi.org/10.1093/jhered/esn015>
140. Zhang, W., Zhang, J., Xu, C., Li, N., Liu, H., Ma, J., Zhu, Y., Xie, H., 2012. LFQuant: A label-free fast quantitative analysis tool for high-resolution LC-MS/MS

- proteomics data. Proteomics 12, 3475–3484.
<https://doi.org/10.1002/pmic.201200017>
141. Zhou, Q., Luo, L., Wang, X., Li, X., 2019. Relationship between single nucleotide polymorphisms in the 3'UTR of amyloid precursor protein and risk of Alzheimer's disease and its mechanism. Biosci. Rep. 39, BSR20182485.
<https://doi.org/10.1042/BSR20182485>
142. Zhu, W., Smith, J.W., Huang, C.-M., 2010. Mass Spectrometry-Based Label-Free Quantitative Proteomics. Journal of Biomedicine and Biotechnology 2010, 1–6.
<https://doi.org/10.1155/2010/840518>

CHAPTER 3.

*FIRST GENOME – WIDE CNV MAPPING
IN FELIS CATUS USING NEXT
GENERATION SEQUENCING DATA*

RESEARCH ARTICLE

Open Access



First genome-wide CNV mapping in *FELIS CATUS* using next generation sequencing data

F. Genova¹, M. Longeri¹, L. A. Lyons², A. Bagnato¹, the 99Lives Consortium and M. G. Strillacci^{1*} 

Abstract

Background: Copy Number Variations (CNVs) have becoming very significant variants, representing a major source of genomic variation. CNVs involvement in phenotypic expression and different diseases has been widely demonstrated in humans as well as in many domestic animals. However, genome wide investigation on these structural variations is still missing in *Felis catus*. The present work is the first CNV mapping from a large data set of Next Generation Sequencing (NGS) data in the domestic cat, performed within the 99 Lives Consortium.

Results: Reads have been mapped on the reference assembly_6.2 by Maverix Biomics. CNV detection with cn.MOPS and CNVnator detected 592 CNVs. These CNVs were used to obtain 154 CNV Regions (CNVRs) with BedTools, including 62 singletons. CNVRs covered 0.26% of the total cat genome with 129 losses, 19 gains and 6 complexes. Cluster Analysis and Principal Component Analysis of the detected CNVRs showed that breeds tend to cluster together as well as cats sharing the same geographical origins. The 46 genes identified within the CNVRs were annotated.

Conclusion: This study has improved the genomic characterization of 14 cat breeds and has provided CNVs information that can be used for studies of traits in cats. It can be considered a sound starting point for genomic CNVs identification in this species.

Keywords: CNV, CNVR, Cn.MOPS, CNVnator, NGS, *Felis catus*, Cat breeds

Background

Short Tandem Repeats and Single Nucleotide Variants (Single Nucleotide Polymorphism – SNPs, as they are known) have been widely used in the study of the genome diversity and inherited diseases for a long time. Other structural and more complex variants - like Copy Number Variants (CNV) - were also known [1], and the recent advances in genome technologies, especially the development of array platforms and next generation sequencing, has allowed more global analyses of CNVs at a genome-wide level. These variants consist of changes in copy number of DNA sequences in comparison to a reference genome, including duplications (gain state) and deletions (loss state). The absence of variations is defined as normal state [2]. Conventionally, CNVs are defined as 1 Kb to several Mb in length, therefore being

distinct from the smaller structural variants as indels and are more variable than SNPs, which vary by a single base pair [3].

An early study of the completed human genome for large-scale copy-number variants (LCVs) identified approximately 200 polymorphisms within the genomes of 20 normal individuals [4]. In another study, the regions including LCVs overlapped with genes involved in human disease syndromes and cancer, such as *CMT4B2* gene (Charcot-Marie-Tooth disease type 4B2) and *DPY*, *LRP12*, *FOG2* genes (squamous cell carcinoma) [5]. CNVs not only influence human phenotypes but are also related to genome evolution. The location and frequencies of the human and chimpanzee CNVs have been compared and duplications and deletions of genes with functions related to cell proliferation and inflammatory response have been found. In particular, a loss of the chimpanzee *TBC1D3* gene, compared to the eight paralog copies of this gene in the human reference individual, may reflect a positive selection and adaptive

* Correspondence: maria.strillacci@unimi.it

¹Department of Veterinary Medicine, University of Milan, 20122 Milan, Italy
Full list of author information is available at the end of the article



© The Author(s). 2018 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

phenotypic differentiation during primate evolution [6]. These studies used array-based comparative genomic hybridization methods (array CGH, aCGH), which lacked sensitivity to detect a wide range of CNVs [7].

The introduction of SNP high density genotyping and Next Generation Sequencing (NGS) approaches, together with the development of new bioinformatics tools, has led to better strategies for CNV detection [8, 9]. The 1000 Genomes Project Consortium (2010), using 179 low coverage whole genome human sequences, identified 28,025 structural variants, suggesting that CNVs represent 4.8–9.5% of the human genome [9]. NGS is also an alternative tool for genotyping CNVs associated to diseases, such as inherited kidney diseases [10].

Genome-wide studies to detect CNVs have been extended to domestic animals. CNVs have been shown to affect phenotypes such as pigmentation, morphology and production traits [11, 12]. CNVs have also been suggested to be responsible for complex disease traits such as periodic fever syndrome in Shar-Pei dogs [13] and canine squamous cell carcinoma of the digit [14].

NGS has supported the discovery of CNVs in animals too, providing higher sensitivity and allowing their identification at a genome-wide level in cattle [15], chickens [16], mice [17] and dogs [18]. Recently, CNV detection with NGS data has been used for association studies on production traits in livestock, such as fatty acids dynamics in beef cattle [19].

The genome assembly of the domestic cat is based on approximately 3× Sanger sequencing and ~ 14× short-read Illumina-based NGS. This assembly_6.2 and the re-sequencing of additional breed individuals have identified a wealth of genetic variation within the cat genome which led in turn to the development of a successful DNA 63 K array [20]. However, an analysis of large structural variants as CNV in the domestic cat across breeds and individuals has not been attempted. New cat genomes are now available from the 99 Lives cat genome sequencing project, which has produced high quality NGS data of 30× coverage, using similar techniques and technologies [21–24]. In such a context, the present study is the first genome-wide CNV detection in the domestic cat and is based on the sequence data from the cat 99 Lives Project. Ultimately this work is meant to assist the evaluation of cat breeds and to be used for the association of CNVs to breed-specific phenotypes, including disease phenotypes.

Materials and methods

Samples

Forty-two whole genome sequences representing 14 different cat breeds were available from the 99 Lives Cat Genome Sequencing Initiative (<http://felinegenetics.missouri.edu/99lives>) [21]. The genome sequences were

produced as previously described [22] and sequencing data are available at NCBI BioProject PRJNA308208. The represented breeds, the number of individuals per breed and additional sequencing information are reported in Table 1.

The genome data for all the cats were produced by Illumina-based short-read technology using PCR-free libraries of ~ 350 bp and/or ~ 550 bp. Most genomes are at ~30× coverage, with 100–150 bp paired end reads. All the reads were mapped to the cat reference assembly_6.2 by Maverix Biomics (<http://www.ncbi.nlm.nih.gov/assembly/320798>) [25].

CNVs detection

Two programs using Read Depth (RD) based methods, cn.MOPS [26] and CNVnator [27], were employed to detect CNVs in the cat genomic data. Cn.MOPS and CNVnator software support the readily available bam files for the analysis and are considered the most suitable tools for Illumina sequencing data [28].

Cn.MOPS

The CNVs were firstly identified using the R “cn.MOPS” routine. The output was filtered to exclude false calls following the software manual indications, so only losses with median in the expected log fold change < -1 and gains with median > 0.6 were considered, as described in details by the authors [26]. Window Length (WL) was set at 5 Kb.

CNVnator

The second CNV detection method utilized CNVnator. The filtering was carried out considering only CNVs with size ≥ 1 Kb, zero mapping quality (q_0) < 0.5 and P_{val1} < 0.001. Since the sequences used for the analysis had ~20× - 30× coverage, the WL was set at 100 bp as suggested by the authors [27].

Consensus mapping of CNV, CNVR and data analysis

To reduce the false positive calls from each of the detection method, the CNVs obtained from cn.MOPS and CNVnator were compared using the -intersectBed command of Bedtools software [29]. Only CNVs that overlap for at least 80% (consensus_CNVs) were considered true calls and included in further analyses. When a different loss/gain state was identified by the two software for a specific CNV, a visual inspection of the sequence read depth was carried out for that CNV to identify the true state. Particularly, samtool -view option (<http://samtools.sourceforge.net/>) was used to extract the CNV sequence, including 50 Kbp of the flanking regions, and the read depth was visually inspected with the Golden Helix GenomeBrowse. This allowed to identify false and true positive calls and assign the true CNVs state.

Table 1 List of the main cat breeds used in the study and number of individuals per breed

Breed	Sample/ Breed	Library	Run Accession	Location
ABYSSINIAN (4)	ABY_1	~ 350 bp	SRR5373742	USA
	ABY_3		SRR6997541	Italy
	ABY_4		Under release	USA
BENGAL (2)	BEN_1	~ 550 bp	SRR5366704	USA
	BEN_2		Under release	Switzerland
BIRMAN (6)	BIR_1	~ 350 bp	SRR5055405	Asia
	BIR_2		Under release	USA
	BIR_3		Under release	Sweden
	BIR_4		Under release	Sweden
	BIR_5		Under release	Sweden
	BIR_6		Under release	Sweden
BRITISH SHORT HAIR (2)	BSH_1	~ 550 bp	SRR5358834	Western
	BSH_2		SRR5358833	Western
BURMESE (5)	BUR_1	~ 550 bp	SRR5373736	Western
	BUR_2		SRR5055402	Asia
	BUR_3		SRR5055400	Asia
	BUR_4		Under release	USA
	BUR_5		Under release	USA
DEVON REX (2)	REX_1	~ 550 bp	SRR5373726	Western
	REX_2		SRR5373735	Western
EGYPTIAN (1)	EGY_1	~ 350 bp	Under release	USA
MAINE COON (2)	MCO_1		Under release	Switzerland
	MCO_2		Under release	USA
NAPOLEON (3)	NAP_1	~ 550 bp	SRR5373738	Western
	NAP_2		SRR5373737	Western
	NAP_3		SRR5373734	Western
ORIENTAL SHORT HAIR (8)	OSH_1	~ 550 bp	SRR5358555	Asia
	OSH_2		SRR5358556	Asia
	OSH_3		SRR5358554	Asia
	OSH_4		SRR5358559	Asia
	OSH_5		SRR5358558	Asia
	OSH_6		Under release	USA
	OSH_7		Under release	USA
	OSH_8		SRR5358557	Asia
PERSIAN (1)	PER_1		SRR5055403	Western
RAGDOLL (2)	RAG_1		SRR5055399	Western
	RAG_2		SRR5055396	Western
SPHYNX (1)	SPH_1	~ 350 bp	Under release	USA
SIAMESE (4)	SIA_1	~ 550 bp	Under release	Asia
	SIA_2		SRR5363128	Asia
	SIA_3		SRR5363127	Asia
	SIA_4		SRR5363129	Asia

CNV regions (CNVRs) for all the cats were obtained by merging consensus_CNVs using the `-mergeBed` command of Bedtools [29]. In order to validate the CNVRs identified and to exclude any possible false positive call, all the identified CNVRs were validated by the visual inspection. The sequence read depths of all the cats pertaining to a CNVR, and thus showing a CNV in that region, were compared to the ones of two cats showing CNV normal state for that region. Only losses or gains with clear boundaries were considered true CNV calls, contributing to CNVRs. All singleton CNV calls were also visually validated to exclude false positive calls and then considered as CNVRs. All the CNVRs boundaries were validated and re-assigned with visual inspection: for both sides, the initial descending bp and the initial ascending bp positions were identified as boundaries for losses and gains respectively.

Genes within the CNVRs were identified using the annotation of the NCBI *Felis catus* assembly_6.2 gene dataset and the Bedtools `-intersectBed` command was used to catalogue genes in the corresponding regions. Gene Ontology Terms (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were performed with the DAVID Bioinformatic Database (<https://david.ncifcrf.gov/tools.jsp>). GeneCards database (www.genecards.org) was consulted to obtain information on the function of the identified genes. Imprinted Gene database (www.geneimprint.com) was used to identify imprinted genes among the ones located within the identified CNVRs.

Cat population analyses

Structure and genomic diversity among all the cats were examined using Principal Component Analysis (PCA) and a clustering analysis, grouping the individuals according to their CNVR similarities [30]. A scoring matrix of the CNVRs was developed by encoding a value of '0' or '1' according to the presence or absence of any mapped CNV in the corresponding CNVR. The matrix considered only the 92 CNVRs shared by two or more individuals. A hierarchical agglomerative clustering, based on Unweighted Pair-Group Average method (UPGMA), was applied to the scoring matrix using the *pvclust* function from the *pvclust* R package [31]. Multiscale bootstrap resampling was performed to calculate the Approximately Unbiased *P*-value (AU-P) using 10,000 bootstraps to assess the robustness of branches. The AU-P and Bootstrap Probability value (BP-P) are presented for each node, as well as the edge numbers. The PCA was performed with Past3 software using the same matrix of the clustering analysis [32].

Results

CNVs detection with cn.MOPS

After filtering, 2282 CNVs were identified using cn.MOPS software. A mean of 23 CNVs/cat was observed. The Siamese cat sample SIA_4 showed 585 CNVs and was

excluded as an outlier. Therefore, 1697 CNVs were considered for downstream analyses.

CNVs detection with CNVnator

After editing, CNVnator detected 285,533 CNVs. A mean of 5827 CNVs/cat was reported and the Siamese cat sample SIA_4 was also confirmed as an outlier. The Siamese cat sample SIA_2 was the cat with the highest number of CNVs (9413) while 4405 CNVs were counted in the Birman cat sample BIR_2, which was the one with the lowest number of CNVs. After the SIA_2 outlier exclusion, 234,484 CNVs were considered in downstream analyses.

Consensus mapping of CNV, CNVR and data analysis

Comparing cn.MOPS and CNVnator detections, a total of 999 consensus_CNVs was obtained, representing 59% of the 1697 CNVs identified by cn.MOPS. Out of the 147 consensus_CNVs with different state, 78 were identified by visual inspection as true call and their state was assigned according to their read depth respect to the flanking regions (Additional file 1: Table S1).

The CNVRs were then identified using the 930 remaining consensus_CNVs. Each of the 389 identified CNVR (including 269 singletons) was visually validated to identify false positive calls, including all detected singletons. An example for the CNVR validation is shown in Additional file 2: Figure S1 at chrD1:10,624,001-10,645,000 where the comparison with the reference genome and two normal state cats, ABY_1 and BEN_1, clearly allows to identify true CNV calls. A total of 154 CNVRs were, containing 589 validated consensus_CNVs, were confirmed. Among those, 62 were identified as singleton. Table 2 summarizes the statistics of consensus_CNVs found in each breed.

The size of the singleton regions (62 CNVRs) ranged from 5 Kb to 283 Kb, while the remaining 92 CNVRs ranged in size from 5 Kb to 529 Kb. The 154 CNVRs and are graphically represented in Fig. 1 and reported in Additional file 3: Table S2 and Additional file 4: Table S3.

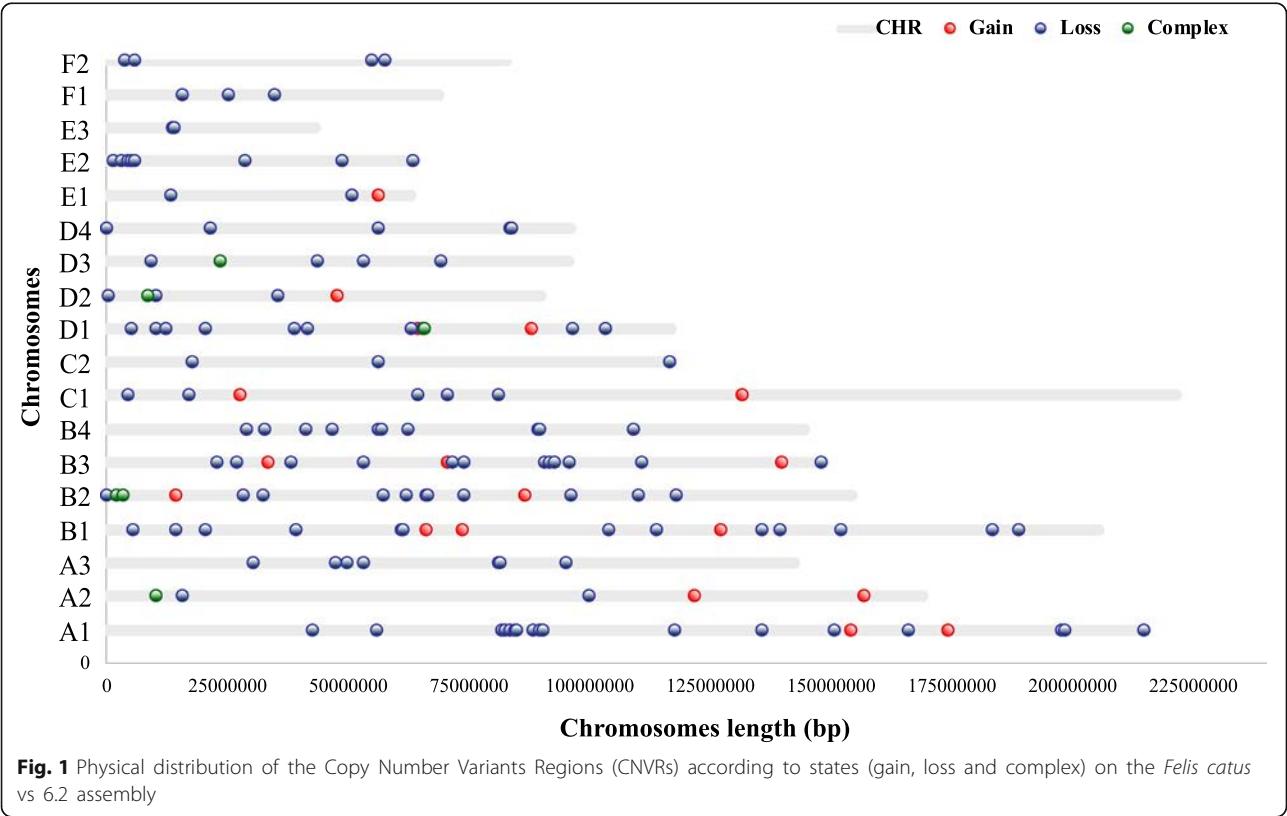
Considering the length of the cat autosomes mapped with *Felis catus* vs 6.2 assembly (about 2.2 Gb), the CNVRs covered about 0.26% of the cat genome. A total of 129 loss, 19 gain and 6 complex regions were identified. The statistics and the contribution of each breed in relation to singletons are reported in Fig. 2. All breeds show to have at least one singleton CNVR with the NAP showing the largest number.

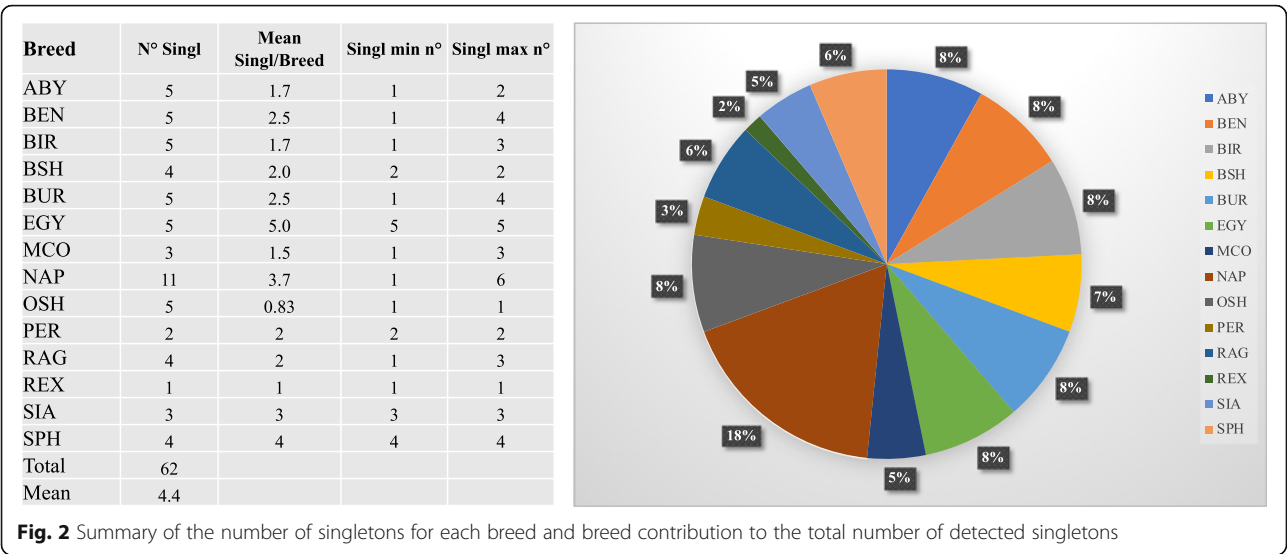
The contribution of every individual to each CNVR is reported in Additional file 4: Table S3. Two CNVRs were in common to the 66% of the cats: the first on chrD4:83,618,990-83,634,627, shared by 28 cats of 10 breeds and the second on chromosome chrA1:89,919,879-89,940,219, shared by 27 cats of 13 breeds. The CNVRs identified in at least 10 individuals were 13,

Table 2 Descriptive statistics of validated copy number variant (consensus_CNVs) identified for each breed

Breed	N. of samples	Tot N. of CNV (*)	Min N. of CNV per sample	Max N. of CNV per sample	Tot N. Losses	Tot N. gains	Tot N. of Chr with CNV
ABY	3	46 (15.3)	13	17	42	4	13
BEN	2	29 (14.5)	12	17	25	4	14
BIR	6	95 (15.8)	11	20	88	7	16
BSH	2	19 (9.5)	9	10	17	2	7
BUR	5	78 (15.6)	10	20	67	11	16
EGY	1	19	19	19	18	1	14
MCO	2	24 (12)	12	12	23	1	9
NAP	3	53 (17.6)	16	20	47	6	14
OSH	8	136 (17)	10	28	117	19	17
PER	1	11	11	11	10	1	7
RAG	2	17 (9)	4	14	16	1	9
REX	2	22 (11)	10	12	19	3	12
SIA	3	21 (7)	4	12	17	4	11
SPH	1	19	19	19	15	4	11
Total	41	589			521	68	

(*) average number of CNV per Breed





while 26 CNVRs were present in only two cats. A total of 16 CNVRs were found in only one breed (ABY, BIR, BSH, BUR, MCO, NAP and OSH) as reported in Table 3.

In Fig. 3 the distribution of the CNVRs across the genome is reported together with the proportion of coverage

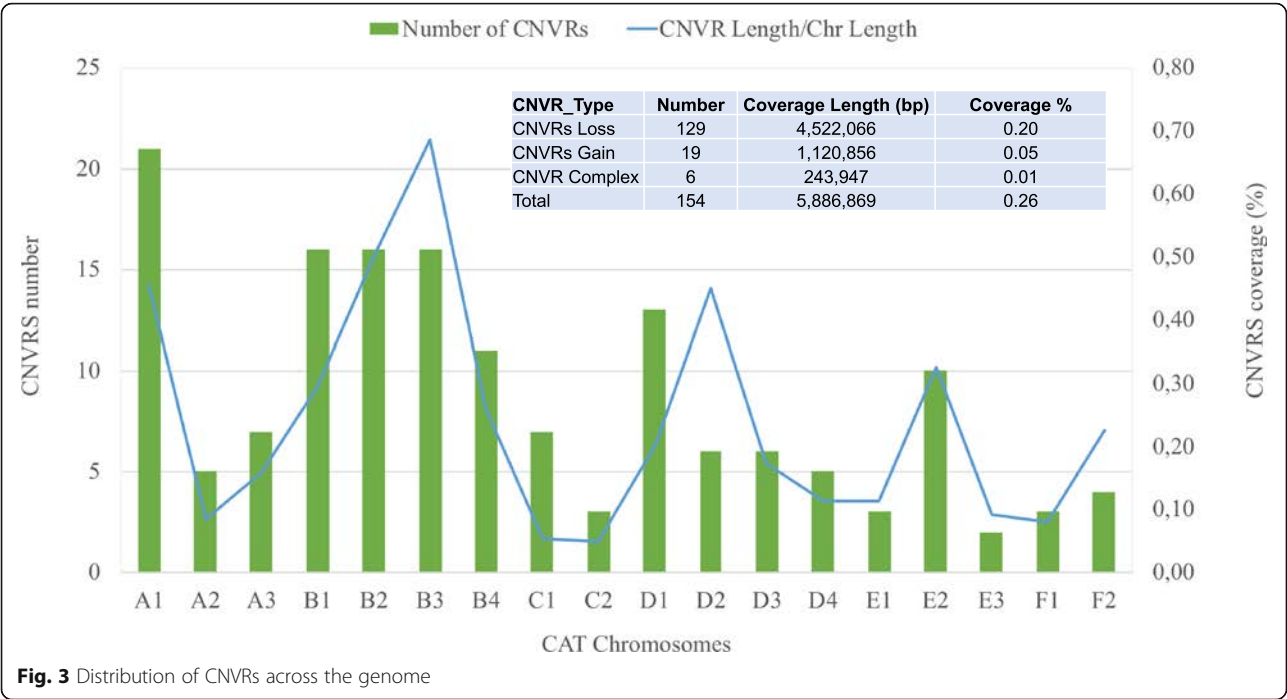
within each chromosome. The number of CNVR per chromosome spans from 2 (chrE1) to 21 (chrA1) while the proportion of the total CNVR per chromosome spans from 0.05% (chrC2) to 0.69% (chrB3).

In Fig. 4 the distribution of CNVRs by five size classes is reported. Only one CNVR, a gain, showed a size large

Table 3 CNVRs found in only one breed

Chr	Start	End	Length	Breed	Samples (*)	State	Genes	Function
chrA1	117,723,730	117,738,524	14,794	ABY	2 (67)	Loss	PCDHB6, PCDHB10, PCDHB12	Neural cadherin-like cell adesion protein
chrB2	74,110,332	74,120,349	10,017	BIR	2 (25)	Loss	SH3BGR2	SH3 Domain Binding Glutamate Rich Protein Like 2
chrD1	103,586,099	103,601,179	15,080	BIR	3 (60)	Loss	LOC101085660	Olfactory Receptor 5G3-like
chrD4	83,882,234	83,893,715	11,481	BIR	5 (83)	loss		
chrE1	56,306,407	56,332,265	25,858	BIR	3 (60)	Loss		
chrB2	96,292,366	96,391,191	98,825	BSH	2 (25)	Loss		
chrB3	70,859,943	71,389,394	529,451	BUR	2 (25)	Gain	AVEN CHRM5 EMC7 RYS3	Apoptosis and Autophagy Pathways Muscarinic receptor Membrane Protein binding carbohydrates Ryanodine receptor for calcium release
chrB4	109,221,798	109,251,945	30,147	BUR	3 (60)	Loss		
chrE2	28,876,777	28,888,767	11,990	BUR	3 (60)	Loss		
chrA1	82,527,688	82,544,767	17,079	MCO	2 (100)	Loss		
chrC2	116,748,111	116,766,145	18,034	NAP	2 (25)	Loss	PLSCR4	Phospholipids migration by calcium ions binding
chrA2	121,965,379	121,983,429	18,050	OSH	2 (25)	Gain	ZNRF2	Maintenance of neural transmission
chrB1	139,721,970	139,739,225	17,255	OSH	2 (25)	Loss	CB1H4orf22	Cilia and flagella associated protein 299
chrB2	110,152,121	110,360,880	208,759	OSH	2 (25)	Loss		
chrB4	89,793,534	89,841,761	48,227	OSH	2 (25)	Loss	FAM19A2	TAFa family, regulators of immune and nervous cells
chrC2	17,974,294	17,988,686	14,392	OSH	2 (25)	Loss		

*Number of samples defining the CNVR and proportion on total number of cats per breed (%)



than 500 Kb (chrB3:70,859,943-71,389,394, 520 Kb), while 67 are smaller than 20 Kb.

Considering the assembly_6.2 annotation, 46 genes were located within the CNVRs (Additional file 3: Table S2 and Additional file 4: Table S3) and 13 mapped in CNVRs identified in only one breed (Table 3).

All the 46 identified genes were submitted to the David Bioinformatic Database but for only 25 genes the GO terms and KEGG metabolic pathways were available as reported in Additional file 5: Table S4.

The GO terms and the KEGG pathway clusters, resulting from the DAVID classification database, are reported in Table 4. Only two clusters have a *P*-Value lower than 0.05 and correspond to a biological process term (G-protein coupled receptor signaling pathway) and a KEGG pathway (Olfactory transduction).

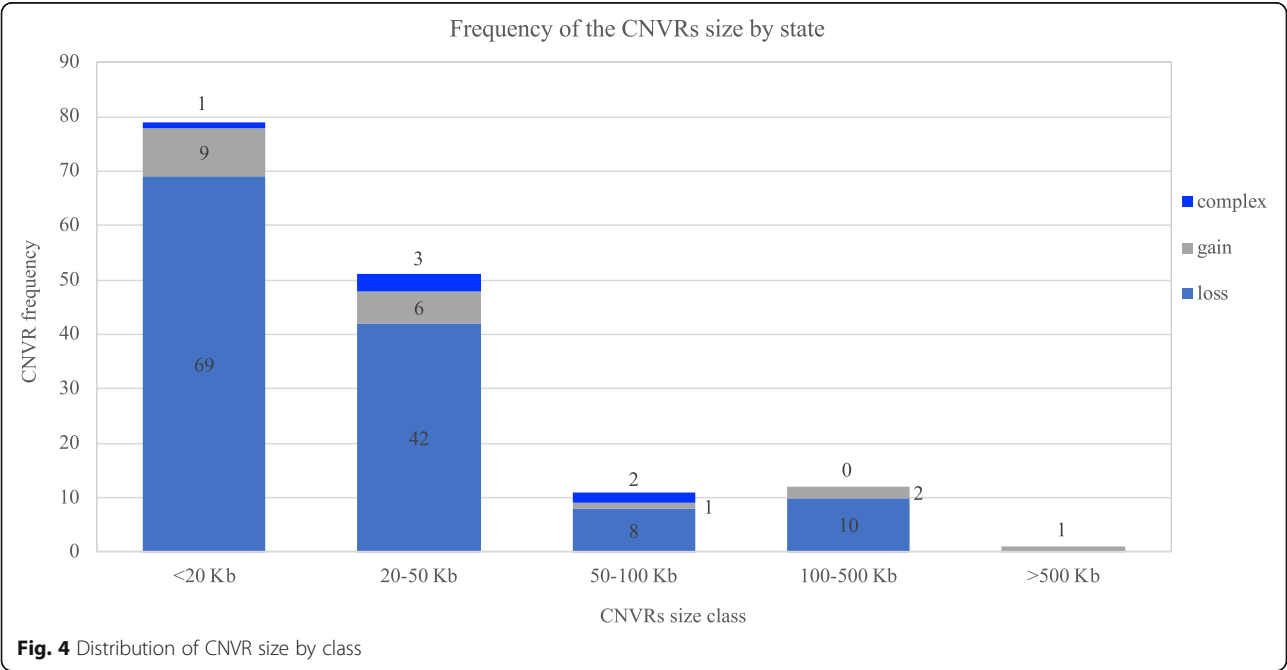


Table 4 Genes clusters according to DAVID database classification

Category	Term	Count	P-Value	Genes
GOTERM_BP	GO:0007186~G-protein coupled receptor signaling pathway	3	9.14E-02	LOC101101252, LOC101084174, LOC101083150
GOTERM_MF	GO:0005509~calcium ion binding	3	1.35E-01	MICU1, ANXA10, RYR3
GOTERM_MF	GO:0004984~olfactory receptor activity	3	1.27E-01	LOC101101252, LOC101084174, LOC101083150
GOTERM_MF	GO:0004930~G-protein coupled receptor activity	3	1.82E-01	LOC101101252, LOC101084174, LOC101083150
GOTERM_CC	GO:0016021~integral component of membrane	5	6.06E-01	ANTXRL, LOC101101252, RYR3, LOC101084174, LOC101083150
GOTERM_CC	GO:0005886~plasma membrane	3	6.28E-01	LOC101101252, LOC101084174, LOC101083150
GOTERM_CC	GO:0005737~cytoplasm	4	6.07E-01	ELP4, SYDE2, PAX6, ARNTL2
KEGG_PATHWAY	fca04740:Olfactory transduction	8	1.10E-05	LOC101095519, LOC101089503, LOC101101252, LOC101084174, LOC101089105, LOC101083150, LOC101083405, LOC101086964

*CC cellular component, MF molecular function

Cat population analyses

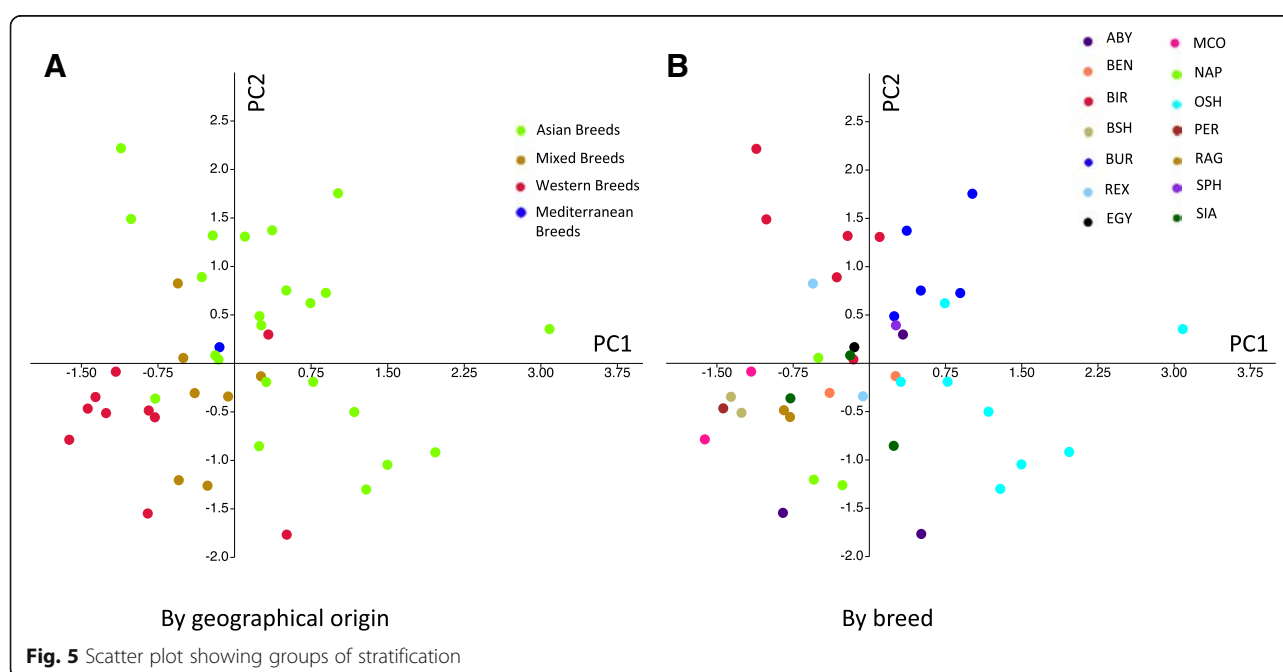
Both the PCA and the cluster analysis depicted a similar population stratification based on different breeds (Figs. 5, 6).

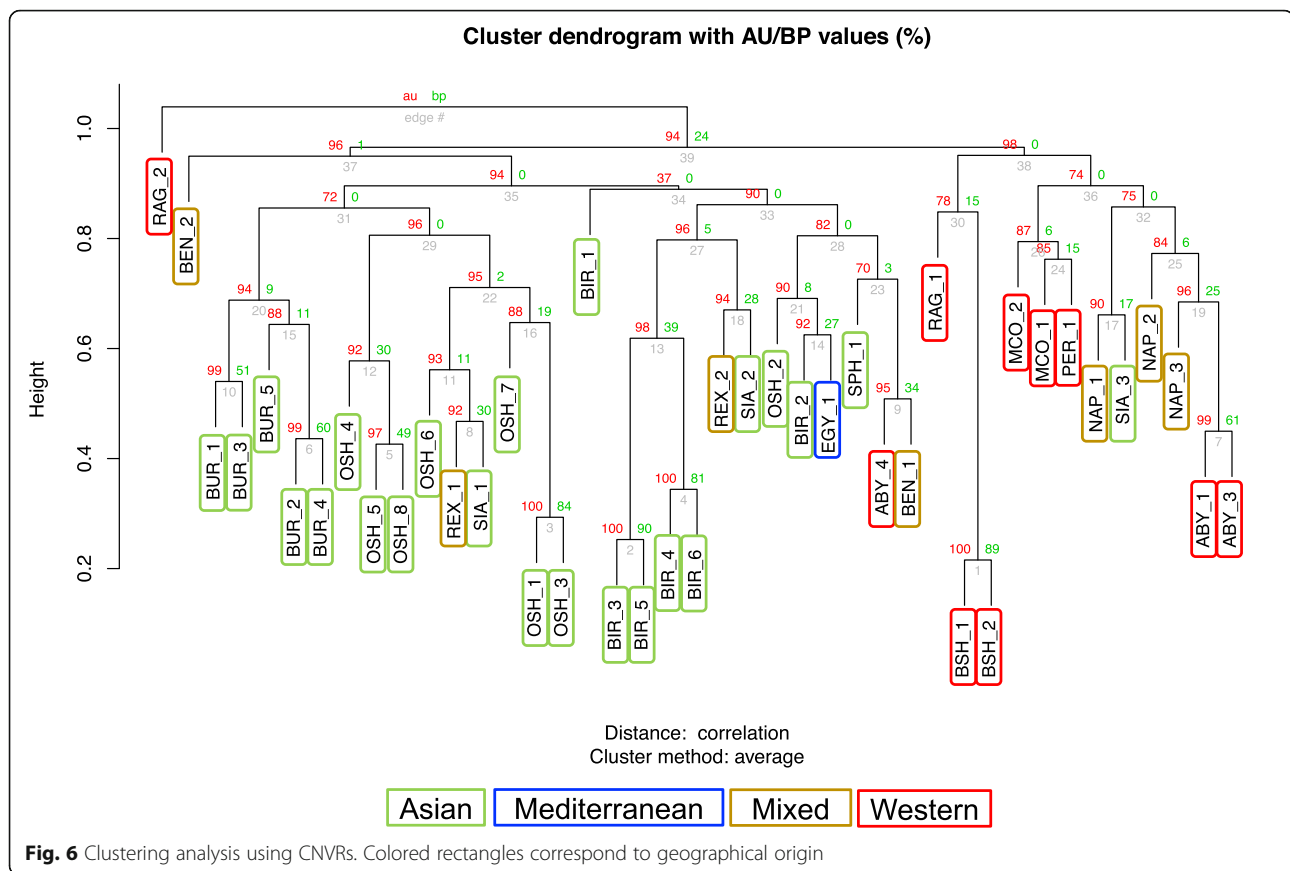
Figure 5 shows the population stratification based on geographical origin (A) and on breeds (B). The principal components explained the 9.62 and 8.63% of the variance for PC1 and PC2 respectively. The breeds were grouped according to their geographical origin as follows: i) Asian Breeds - SIA, OSH, BUR, BIR and SPH; ii) Western Breeds - ABY, MCO, BSH, PER and RAG; iii) Mediterranean breed - EGY; iv) Mixed breeds - BEN, REX and NAP.

The clustering in Fig. 5-a shows two main groups of individuals represented by Asian breeds (green) and Western breeds (red). In Fig. 5-b a clear separate clustering is shown for BIR, BUR and OSH that are part of the Asian breeds group in Fig. 5-a. Among the Western breeds the BSH the PER and the RAG cluster very closely (Fig. 5-a).

The CNVR clustering tree also presented close associations for individuals belonging the same breed and for breeds with similar origins (Fig. 6).

The breeds clustered well according to the geographical origin, except for very few cases (RAG_2 and ABY_4 for the Western breeds; SIA_3 for the Asian breeds). For





breeds with only one representative, the distribution was based on their geographical origins: Persian (PER_1) within the Western breed group while Egyptian (EGY_1) and Sphynx (SPH_1) within the Eastern breed group. Additionally, cats pertaining to the same breeds tend to cluster together as, e.g. the Burmese cats. Clustering with AU-P values > 90 were identified for several groupings of breeds, as for groups of individuals sharing the same geographical origins.

Discussion

Genome-wide mapping of CNVs has allowed new insights into genomic variation. Different techniques based on different approaches have been developed with the aim of more efficient and accurate CNV detection [28]. While aCGH has been used to detect CNVs for several years, the advent of NGS technologies has improved the global detection and analysis of CNV data. In humans, as in domestic animals, different CNV analyses have already been performed and have demonstrated the strong relationship between the presence of CNVs and phenotypic diversity, evolution and the onset of several diseases [33, 12]. Although the examination of repetitive elements was carried out in cats [25] and a preliminary annotation of the whole genome reference sequence

(*Felis catus* assembly_6.2) has revealed that repetitive elements comprise about 55.7% of the whole genome, no formal studies focusing on CNVs have been performed [34].

This is the first CNV detection in the domestic cat and also the first one using NGS data. CNVnator and cn.MOPS software were selected to identify the CNVs. Both packages are based on the Read Depth (RD) method, which has been suggested as the most suitable for Illumina sequencing data [28]. Compared to other methods, the RD method does not require a reference sample and is based on coverage of the sequencing data. This represents an important advantage as, during the detection, the software compares reads within the same sample and also among all the samples. Moreover, the RD method uses a negative binomial distribution statistical model [28] and has the advantage of setting the WL depending on the number of reads. The difference between the two software programs is based on the RD method pipelines described by [35]. Cn.MOPS is based on a Bayesian approach that measures the depth of coverage at each genomic position, across multiple samples, thus increasing the statistical power and decreasing the false discovery rate in the detection. CNVs and noise identification are achieved using mixture components

and Poisson distribution [26]. In contrast, CNVnator uses the mean-shift approach [36] and CNV detection is made more accurate through GC and multiple-bandwidth partitioning corrections [35]. The significant difference in the number of detected CNVs between the two packages is based on the different approach for defining a single CNV. In cn.MOPS, copy number of adjacent windows are compared and those with the same copy number sequence are joined into one single CNV. The length of the final CNV is given by the sum of the two adjacent copy number sequences. The more adjacent segments with high or low copy number call are joined, the higher will be the confidence in the detections. CNVnator does not use this overlapping approach, which explains the large discrepancy in the number of CNVs detected by the two packages. Ten or more CNVs detected with CNVnator correspond to one CNV detected with cn.MOPS. For this reason, it was decided to obtain CNVRs after the selection of CNVs in common to the two software.

The number of analyzed individuals is comparable or superior to those used in previous studies on horses (16 individuals [37]), pigs (20 individuals [38]) and dogs (34 individuals [39]). The number and the mean lengths of CNVs are in accordance to those found in dogs, where 1748 CNVs with a mean length of 194,559 bp were identified [40], and in other studies using cn.MOPS [41] and CNVnator software [17].

The present CNV detection revealed a higher number of loss regions compared to gain and complex regions. This could be explained by the difference within breeds of the same species, as already reported for Holstein cattle, where the number of loss regions is consistently higher than in other cattle breeds [42, 43]. Moreover, even though CNVnator is still the most used software for CNV detection, it is known to have a major deficiency in terms of detecting gain regions [27]. This could also explain the identification of 147 CNVs with different state when the output of the two software packages was compared. Nevertheless, the visual inspection approach used in this study to validate all the CNVRs, allows to overcome the problem of having different calling state between the cn.MOPS and CNVnator software. Additionally, it further reduces the false positive calls that may occur when using only one of the software for the CNV detection. In fact, the visual inspection leads to a full validation of the called CNV, improving the true calls obtained by comparing the results of the two calling algorithms. Visual inspection was also used to re-assign the boundaries at each single CNVR, allowing to further refine the genome proportion covered by CNVR. The singletons detected in this study represent the 41% of the total CNVRs, a lower proportion compared to those reported in other studies [44–46].

The proportion of validated non singleton CNVRs after visual inspection was 77%. This proportion of true calls is in accordance to the findings of [47] who identified a concordance for the 80% of mapped CNV using two sequencing data runs.

Grouping the individuals according to their CNVR similarities, both the PCA and the clustering analyses showed cats belonging to the same breed tended to cluster together as well as cats sharing the same geographical origin. The Western and Eastern breeds were distinct in both of the analyses and resembled the results obtained in previous studies of cat phylogenies [20, 48]. However, the Devon Rex cats, a breed developed in the United Kingdom [49], neither cluster nor have apparent correct historical origins, which are suggested as Eastern in this study. Depending on breeding associations, the genetic contributions of some cat breeds can be varied as different associations allow for different outcrosses for a given breed. Thus, some of the variation in the breed and individual associations may be due to historical breeding differences.

The Bengal breed is a hybrid, developed from crossing spotted cats from India, Egyptian Maus or Abyssinians with the Asian leopard cat (*Prionailurus bengalensis*). Thus, the convoluted genetics of an individual cat of the Bengal breed could easily result in placement nearly anywhere within the cat genetic spectrum. As previously stated [50], the Abyssinian, which is one of the oldest cat breeds, has mysterious origins that are not clearly defined.

The gene annotation performed in this study showed that 19% of the CNVRs harbor genes. This proportion is lower than the ones found in studies developed in other species [8, 40, 46]. This is likely due to the more incomplete information in the cat gene annotation, compared to other species.

The *LOC101085660*, *LOC101095519*, *LOC101101252* and *LOC101089105* genes encode for olfactory receptors and were found in CNVRs of several breeds. As previously shown in humans [51], it may be related to significant variability in olfactory capabilities, an important sensorial attribute in predators.

The *RYR3* gene in the Burmese cat was found differentially expressed in adipose tissue in cats during winter/short days and summer/long days [52].

The *SYDE2*, *PAX6*, *ELP4* and the *CHRM5* genes, have been found in genomic regions that have been recognized to be under positive selection in cats during their domestication [53]. Interestingly, the *PAX6* and the *ELP4* genes were found by the same authors as genes underlying segmental duplication in the domestic cat genome, as occurring in this study.

The *PTCHD3* gene was found in the CNVR (complex) at chrB2:2,291,441–2,347,122 in 5 breeds (BEN, BIR,

OSH, RAG, SPH). This gene has been previously found in a rare CNV in humans [54] where the CNV homozygous deletion was not associated to an abnormal phenotype. More recently, [55] have found this gene as associated with diabetes in humans and [56] have classified the gene as a potential imprinted gene. The *PTCHD3* gene is not yet included in the Imprinted Genes database (www.geneimprint.com), as no other gene found in the CNVRs here reported. The presence of imprinted genes in CNVRs has already been shown in other species such as in cattle [57]. The regulation of the gene expression is mainly determined by the genetic imprinting and it could be interesting to further investigate this aspect in cat too.

Conclusions

The CNV calling performed in this study represents the first effort for the detection of genomic structural variation in the domestic cat. The clustering among the cat breeds that was possible to obtain in this study using CNVRs, complement findings of other studies based on other type of markers, leading to a closer insight of common and proprietary functional aspects of each population. Further studies based on further resequencing and on novel NGS technologies, might disclose other insights on CNV in *Felis catus* species and could complement the results obtained with the mapping performed in this study. Since CNV are well known to be related to gene expression regulation, also in complex diseases, this first mapping is meant to be the first information on a class of genomic variants that can be related to recorded phenotypes in cat populations.

Additional files

Additional file 1: Table S1. (XLSX). Consensus CNVs. List of the consensus list of CNVs detected by both cn.mops and CNVnator software and state confirmed with visual inspection with GenomeBrowse (XLSX 34 kb)

Additional file 2: Figures S1. (PDF). Visual inspection of CNVR at chrD1:10624094–10,643,050. (TIF 2999 kb)

Additional file 3: Table S2. (XLSX). Singleton CNVRs and list of annotated genes. (XLSX 14 kb)

Additional file 4: Table S3. (XLSX). List of CNVRs. List of the CNVRs and annotated genes. Singletons were excluded. (XLSX 20 kb)

Additional file 5: Table S4. (DOCX). David annotation of genes in all CNVRs. List of the annotation gene names GO terms (BP: biological process; CC: cellular component; MF: molecular function), and KEGG pathways. (DOCX 22 kb)

Abbreviations

ABY: Abyssinian breed; aCGH: Array-Based Comparative Genomic Hybridization; AU-P: Approximately Unbiased *P*-value; BEN: Bengal breed; BIR: Birman breed; BP-P: Bootstrap Probability value; BSH: British Short Hair breed; BUR: Burmese breed; CNV: Copy Number Variants; CNVR: Copy Number Variants Region; EGY: Egyptian breed; GO: Gene Ontology Terms; KEGG: Kyoto Encyclopedia of Genes and Genomes; LCVs: Large-Scale Copy Number Variants; MCO: Main Coon breed; NAP: Napoleon breed; NGS: Next Generation Sequencing; OSH: Oriental Short Hair breed; PER: Persian breed;

RAG: Ragdoll breed; RD: Read Depth; REX: Devon Rex breed; SIA: Siamese breed; SNP: Single Nucleotide Polymorphism; SPH: Sphynx breed; UPGMA: Unweighted Pair-Group Average

Acknowledgements

MFR of the 99Lives Consortium acknowledges support of state of Iowa and hatch funding.

EIG of the 99Lives Consortium acknowledges University of Massachusetts Medical School for providing funding.

EO of the 99Lives Consortium acknowledges the Intramural Program of the National Human Genome Research Institute at the National Institutes of Health, USA for providing funding.

99 lives consortium

Barbara Gandolfi¹, Danielle Aberdein², Paulo C. Alves^{3,4}, Gregory S. Barsh^{5,6}, Holly C. Beale⁷, Tomas F. Bergström⁸, Adam R. Boyko⁹, Jeffrey A. Brockman¹⁰, Marta G. Castelhamo¹¹, Patricia P. Chan⁷, Brian W. Davis¹², Ottmar Distl¹³, Nicholas H. Dodman¹⁴, N. Matthew Ellinwood¹⁵, Jonathan E. Fogle¹⁶, Oliver P. Forman¹⁷, Dorian J. Garrick^{2,17}, Jens Häggström¹⁸, Christopher R. Helps¹⁹, Marjo K. Hytönen²⁰, Daniel M. Ibrahim²¹, Maria Kaukonen²⁰, Christopher B. Kaelin^{5,6}, Emilie Leclerc²², Teri L. Lear²³, Tosso Leeb²⁴, Hannes Lohi²⁰, Dario G. Lupiáñez²¹, Mark A. Magnuson^{25,26}, Richard Malik²⁷, Michael J. Montague²⁸, John S. Munday², William J. Murphy²⁹, Elaine A. Ostrander¹², Niels C. Pedersen³⁰, Simon M. Petersen-Jones³¹, Max F. Rothschild¹⁵, Beth Shapiro³², Joshua A. Stern³⁰, William F. Swanson³³, Karen A. Terio³⁴, Rory J. Todhunter¹¹, Yu Ueda³⁰, Wesley C. Warren²⁸, Elizabeth A. Wilcox¹¹, Julia H. Wildschutte³⁵, Edward I. Ginns³⁶.

¹Department of Veterinary Medicine and Surgery, College of Veterinary Medicine, University of Missouri, Columbia, Missouri, 65211 USA.

²School of Veterinary Science, Massey University, Palmerston North 4410 New Zealand.

³CIBIO/InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos/InBIO Associate Lab & Faculdade de Ciências, Universidade do Porto, Campus e Vairão, 4485–661 Vila do Conde, Portugal.

⁴Wildlife Biology Program, University of Montana, Missoula, Montana, 59812 USA.

⁵HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, 35806 USA.

⁶Department of Genetics, Stanford University, Stanford, California, 94305 USA.

⁷Maverix Biomics, Inc., San Mateo, California, 94402 USA.

⁸Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, 750 07 Uppsala, Sweden.

⁹Department of Biomedical Sciences and Cornell Veterinary Biobank, College of Veterinary Medicine, Cornell University, Ithaca, New York, 14853 USA.

¹⁰Hill's Pet Nutrition Inc., Topeka, Kansas 66601 USA.

¹¹Department of Clinical Sciences and Cornell Veterinary Biobank, College of Veterinary Medicine, Cornell University, Ithaca, New York, 14853 USA.

¹²Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892 USA.

¹³Institute for Animal Breeding and Genetics, University of Veterinary Medicine Hannover (Foundation), 30559, Hannover, Germany.

¹⁴Department of Clinical Sciences, Cummings School of Veterinary Medicine, Tufts University, Grafton, MA, 01536 USA.

¹⁵Department of Animal Science, College of Agriculture and Life Sciences, Iowa State University, Ames, Iowa, 50011 USA.

¹⁶College of Veterinary Medicine, North Carolina State University, Raleigh, NC 27607.

¹⁷WALTHAM Centre for Pet Nutrition, Freeby Lane, Waltham on the Wolds, Leicestershire, LE14 4RT UK.

¹⁸Department of Clinical Sciences, Faculty of Veterinary Medicine and Animal Science, Swedish University of Agricultural Sciences, Uppsala, SE-750 07 Sweden.

¹⁹Langford Veterinary Services, University of Bristol, Langford, Bristol, BS40 5DU UK.

²⁰Department of Veterinary Biosciences and Research Programs Unit, Molecular Neurology, University of Helsinki and Folkhälsan Research Center, Helsinki 00014 Finland.

²¹Max Planck Institute for Molecular Genetics, RG Development and Disease, D-14195, Berlin, Germany.

²²Diana Petfood, SPF, ZA du Gohelis, 56250 Elven, France.

²³Department of Veterinary Science, University of Kentucky - Lexington, Lexington, KY, 40506 USA (In memoriam).

- ²⁴Vetsuisse Faculty, Institute of Genetics, University of Bern, 3001 Bern, Switzerland.
- ²⁵Department of Cell and Developmental Biology, Vanderbilt Center for Stem Cell Biology, Vanderbilt University School of Medicine, Vanderbilt University Program in Developmental Biology, Nashville, Tennessee, 37,240 USA.
- ²⁶Department of Molecular Physiology and Biophysics, Vanderbilt University School of Medicine, Nashville, Tennessee, 37,240 USA.
- ²⁷Centre for Veterinary Education, University of Sydney, Sydney, NSW, 2006 Australia.
- ²⁸The McDonnell Genome Institute, Washington University School of Medicine, St. Louis, Missouri, 63,108 USA.
- ²⁹Department of Veterinary Integrative Biosciences, College of Veterinary Medicine, Texas A&M University, College Station, Texas, 77,845 USA.
- ³⁰Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California at Davis, Davis, California, 95,616 USA.
- ³¹Small Animal Clinical Sciences, College of Veterinary Medicine, Michigan State University, East Lansing, Michigan, 48,824 USA.
- ³²Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, Santa Cruz, California 95,064 USA.
- ³³Center for Conservation and Research of Endangered Wildlife (CREW), Cincinnati Zoo & Botanical Garden, Cincinnati, Ohio, 45,220 USA.
- ³⁴Zoological Pathology Program, University of Illinois, Brookfield, Illinois 60,513 USA.
- ³⁵Bowling Green State University, Department of Biological Sciences, Bowling Green, Ohio 43,403 USA.
- ³⁶Department of Psychiatry, University of Massachusetts Medical School, Worcester, MA, 01655 USA.

Funding

Funding for the 99 Lives Cat Genome Sequencing Initiative has been provided by the Winn Feline Foundation and the George Sydney and Phyllis Redman Miller Trust (MT-13-010), the National Geographic Society Education Foundation (2P-14), the University of Missouri Gilbreath-McLorn Endowment (LAL), the Cornell Feline Health Center (ARB, MGC, RJT), contributions from the 99 Lives Consortium participants, donations from Associazione Nazionale Felina Italiana (ANFI), Zoetis, Orivet Genetic Pet Care, Langford Veterinary Services, the World Cat Federation, and public donations. Internal funds from the Università degli Studi di Milano, Italy.

Availability of data and materials

All the data used in this study were fully available without restriction. The data are fully accessible in NCBI database according to accession number as specified in Table 1. Data with no accession number are under release and can be accessed in the meantime freely upon request to the 99Lives Consortium Coordinator.

Authors' contributions

FG, MGS and AB analyzed the data and wrote the manuscript, ML, LAL, AB and MGS revised critically the results and the manuscript, 99L provided the genomic data, MGS conceived the idea and coordinated the research. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable as all the whole genome sequence data were made available by the 99Lives Consortium.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Veterinary Medicine, University of Milan, 20122 Milan, Italy.

²Department of Veterinary Medicine and Surgery, College of Veterinary Medicine, University of Missouri, Columbia, MO 65211, USA.

Received: 12 July 2018 Accepted: 21 November 2018

Published online: 10 December 2018

References

- Bridges CB. The bar "gene" a duplication. *Science*. 1936;83(2148):210–1.
- Henrichsen CN, Chaignat E, Reymond A. Copy number variants, diseases and gene expression. *Hum Mol Genet*. 2009;18(R1):R1–8.
- Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, et al. The diploid genome sequence of an individual human. *PLoS Biol*. 2007;5(10):e254.
- Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, et al. Large-scale copy number polymorphism in the human genome. *Science*. 2004;305(5683):525–8.
- Iafate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, et al. Detection of large-scale variation in the human genome. *Nat Genet*. 2004;36(9):949–51.
- Perry GH, Yang F, Marques-Bonet T, Murphy C, Fitzgerald T, Lee AS, et al. Copy number variation and evolution in humans and chimpanzees. *Genome Res*. 2008;18(11):1698–710.
- Pinto D, Darvishi K, Shi X, Rajan D, Rigler D, Fitzgerald T, et al. Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nat Biotechnol*. 2011;29(6):512–20.
- Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. *Nature*. 2006;444:444–54.
- Zarrei M, MacDonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. *Nat Rev Genet*. 2015;16(3):172–83.
- Nagano C, Nozu K, Morisada N, Yazawa M, Ichikawa D, Numasawa K, et al. Detection of copy number variations by pair analysis using next-generation sequencing data in inherited kidney diseases. *Clin Exp Nephrol*. 2018;22(4):881–8.
- Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet*. 1998;20(2):207–11.
- Clop A, Vidal O, Amills M. Copy number variation in the genomes of domestic animals. *Anim Genet*. 2012;43(5):503–17.
- Olsson M, Meadows JRS, Truvé K, Pielberg GR, Puppo F, Mauceli E, et al. A novel unstable duplication upstream of HAS2 predisposes to a breed-defining skin phenotype and a periodic fever syndrome in Chinese Shar-Pei dogs. *PLoS Genet*. 2011;7(3):e1001332.
- Karyadi DM, Karlins E, Decker B, vonHoldt BM, Carpintero-Ramirez G, Parker HG, et al. A copy number variant at the KITLG locus likely confers risk for canine squamous cell carcinoma of the digit. *PLoS Genet*. 2013;9(3):e1003409.
- Bickhart DM, Hou Y, Schroeder SG, Alkan C, Cardone MF, Matukumalli LK, et al. Copy number variation of individual cattle genomes using next-generation sequencing. *Genome Res*. 2012;22(4):778–90.
- Sohrabi SS, Mohammadabadi M, Wu D, Esmailzadeh A. Detection of breed-specific copy number variations in domestic chicken genome. *Genome*. 2018;61(1):7–14.
- Harr B, Karakoc E, Neme R, Teschke M, Pfeifle C, Pezer Ž, et al. Genomic resources for wild populations of the house mouse, *Mus musculus* and its close relative *Mus spretus*. *Sci Data*. 2016;13(3):160075.
- Freedman AH, Gronau I, Schweizer RM, Ortega-Del Vecchyo D, Han E, Silva PM, et al. Genome sequencing highlights the dynamic early history of dogs. *PLoS Genet*. 2014;10(1):e1004016.
- de Lemos MVA, Peripolli E, Berton MP, Feitosa FLB, Olivieri BF, Stafuzza NB, et al. Association study between copy number variation and beef fatty acid profile of Nelore cattle. *J Appl Genet*. 2018;59(2):203–23.
- Gandolfi B, Alhaddad H, Abdi M, Bach LH, Creighton EK, Davis BW, et al. Applications and efficiencies of the first cat 63K DNA array. *Sci Rep*. 2018;8:7024.
- Mauler DA, Gandolfi B, Reinero CR, O'Brien DP, Spooner JL, Lyons LA, et al. Precision medicine in cats: novel Niemann-pick type C1 diagnosed by whole-genome sequencing. *J Vet Intern Med*. 2017;31(2):539–44.
- Gandolfi B, Grahn RA, Creighton EK, Williams DC, Dickinson PJ, Sturges BK, et al. COLQ variant associated with Devon rex and Sphynx feline hereditary myopathy. *Anim Genet*. 2015;46(6):711–5.
- Aberdein D, Munday JS, Gandolfi B, Dittmer KE, Malik R, Garrick DJ, et al. A FAS-ligand variant associated with autoimmune lymphoproliferative syndrome in cats. *Mamm Genome*. 2017;28(1–2):47–55.
- Oh A, Pearce JW, Gandolfi B, Creighton EK, Suedmeyer WK, Selig M, et al. Early-onset progressive retinal atrophy associated with an IQCB1 variant in African black-footed cats (*Felis nigripes*). *Sci Rep*. 2017;7:43918.

25. Pontius JU, Mullikin JC, Smith DR. Agencourt sequencing team, Lindblad-Toh K, Gnerre set al. Initial sequence and comparative analysis of the cat genome. *Genome Res.* 2007;17(11):1675–89.
26. Klambauer G, Schwarzbauer K, Mayr A, Clevert DA, Mitterecker A, Bodenhofer U, et al. Cn.MOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. *Nucleic Acids Res.* 2012;40(9):e69.
27. Abyzov A, Urban AE, Snyder M, Gerstein M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* 2011;21(6):974–84.
28. Pirooznia M, Goes FS, Zandi PP. Whole-genome CNV analysis: advances in computational approaches. *Front Genet.* 2015;6:138.
29. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26(6):841–2.
30. Tian M, Wang Y, Gu X, Feng C, Fang S, Hu X, et al. Copy number variants in locally raised Chinese chicken genomes determined using array comparative genomic hybridization. *BMC Genomics.* 2013;14:262.
31. Ryota S, Hidetoshi S. Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics.* 2006;22(12):1540–2.
32. Hammer Ø, Harper DAT, Ryan PDPAST. Paleontological statistics software package for education and data analysis. *Palaeontol Electron.* 2001;4(1):9.
33. Shastry BS. Copy number variation and susceptibility to human disorders. *Mol Med Rep.* 2009;2(2):143–7.
34. Tamazian G, Simonov S, Dobrynin P, Makunin A, Logachev A, Komissarov A, et al. Annotated features of domestic cat – *Felis catus* genome. *Gigascience.* 2014;3:13.
35. Miller CA, Hampton O, Coarfa C, Milosavljevic A. ReadDepth: a parallel R package for detecting copy number alterations from short sequencing reads. *PLoS One.* 2011;6(1):e16327.
36. Wang LY, Abyzov A, Korbel JO, Snyder M, Gerstein MMSB. A mean-shift-based approach for the analysis of structural variation in the genome. *Genome Res.* 2009;19(1):106–17.
37. Doan R, Cohen N, Harrington J, Veazey K, Juras R, Cothran G, et al. Identification of copy number variants in horses. *Genome Res.* 2012;22(5):899–907.
38. Wang J, Jiang J, Wang H, Kang H, Zhang Q, Liu JF. Improved detection and characterization of copy number variations among diverse pig breeds by array CGH. *G3 (Bethesda).* 2015;5(6):1253–61.
39. Serres-Armero A, Povolotskaya IS, Quilez J, Ramirez O, Santpere G, Kuderna LFK, et al. Similar genomic proportions of copy number variation within gray wolves and modern dog breeds inferred from whole genome sequencing. *BMC Genomics.* 2017;18(1):977.
40. Molin AM, Berglund J, Webster MT, Lindblad-Toh K. Genome-wide copy number variant discovery in dogs using the CanineHD genotyping array. *BMC Genomics.* 2014;19(15):210.
41. McNew SM, Beck D, Sadler-Riggelman I, Knutie SA, Koop JAH, Clayton DH, et al. Epigenetic variation between urban and rural populations of Darwin's finches. *BMC Evol Biol.* 2017;17(1):183.
42. Jiang L, Jiang J, Yang J, Liu X, Wang J, Wang H, et al. Genome-wide detection of copy number variations using high-density SNP genotyping platforms in Holsteins. *BMC Genomics.* 2013;14:131.
43. Gao Y, Jiang J, Yang S, Hou Y, Liu GE, Zhang S, et al. CNV discovery for milk composition traits in dairy cattle using whole genome resequencing. *BMC Genomics.* 2017;18(1):265.
44. Yi G, Qu L, Liu J, Yan Y, Xu G, Yang N. Genome-wide patterns of copy number variation in the diversified chicken genomes using next-generation sequencing. *BMC Genomics.* 2014;15:962.
45. Han R, Yang P, Tian Y, Wang D, Zhang Z, Wang L, et al. Identification and functional characterization of copy number variations in diverse chicken breeds. *BMC Genomics.* 2014;15:934.
46. Prinsen RTMM, Strillacci MG, Schiavini F, Santus E, Rossoni A, Maurer V, et al. A genome-wide scan of copy number variants using high-density SNPs in Brown Swiss dairy cattle. *Livest Sci.* 2016;191:153–60.
47. Letaief R, Rebours E, Grohs C, Meersseman C, Fritz S, Trouilh L, et al. Identification of copy number variation in French dairy and beef breeds using next-generation sequencing. *Genet Sel Evol.* 2017;49:77.
48. Lipinski MJ, Froenicke L, Baysac KC, Billings NC, Leutenegger CM, Levy AM, et al. The ascent of cat breeds: genetic evaluations of breeds and worldwide random bred populations. *Genomics.* 2008;91(1):12–21.
49. Robinson R. The rex mutants of the domestic cat. *Genetica.* 1971;42:466.
50. Robinson R. The domestic cat, *Felis catus*. In: King RC, editor. *Handbook of genetics*. Boston: Springer; 1975.
51. Young JM, Endicott RM, Parghi SS, Walker M, Kidd JM, Trask BJ. Extensive copy-number variation of the human olfactory receptor gene family. *Am J Hum Genet.* 2008;83(2):228–42.
52. Mori A, Kappen KL, Dilger AC, Swanson KS. Effect of photoperiod on the feline adipose transcriptome as assessed by RNA sequencing. *BMC Vet Res.* 2014;10:146.
53. Montague MJ, Li G, Gandolfi B, Khan R, Aken BL, Searle SM, et al. Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. *Proc Natl Acad Sci U S A.* 2014;111(48):17230–5.
54. Seno MMG, Kwan BYM, Lee-Ng KKM, Moessner R, Lionel AC, Marshall CR, Scherer SW. Human PTCHD3 nulls: rare copy number and sequence variants suggest a non-essential gene. *BMC Medical Genetics.* 2011;12:45.
55. Yamada Y, Sakuma J, Takeuchi I, Yasukochi Y, Kato K, Oguri M, et al. Identification of five genetic variants as novel determinants of type 2 diabetes mellitus in Japanese by exome-wide association studies. *Oncotarget.* 2017;8(46):80492–505.
56. Bonaldi A, Kashiwabara A, de Araújo ÉS, Pereira LV, Paschoal AR, Andozia MB. Mining novel candidate imprinted genes using genome-wide methylation screening and literature review. *Epigenomes.* 2017;1(2):13.
57. da Silva JM, Giachetto PF, da Silva LO, Cintra LC, Paiva SR, Yamagishi MEB, et al. Genome-wide copy number variation (CNV) detection in Nelore cattle reveals highly frequent variants in genome regions harboring QTLs affecting production traits. *BMC Genomics.* 2016;17:454.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions



CHAPTER 4.

CONCLUSIONS

CHAPTER 4. CONCLUSIONS

4.1. Summary of results

The whole genome cat sequences available through the 99 Lives Initiative have played a pivotal role in the development of the whole PhD project and can be considered a powerful source for genomics studies in cats.

The sequencing data were firstly used to carry out a large-scale genetic study, investigating the genetic architecture of amyloidosis in domestic animals, specifically in the Abyssinian cat breed. The use of different genome assembly versions led to obtain different data with different significance, but the main finding was represented by the characterization of feline amyloidosis as a non-monogenic trait with different mutations involved in the onset of disease. Moreover, the several variants underlying the pathogenic pathway in the Abyssinians were not shared with the Siamese breed.

In the first study on Abyssinians, using the vs 6.2 of the cat genome, the single locus statistical analysis allowed to identify six amyloidosis associated variants among the 51 coding variants detected through the WGS, with three being extremely significant. The application of the MALR to consider all the possible combination of the genotypes in the genetic profiles at the six associated loci, allowed to preliminary subdivide cats in different risk classes, according to the risk prediction probability. MALR showed that for all six loci, the presence of the mutated allele increased the risk of disease.

Anyway, the major concern was to clarify the meaning of the variants and the related risk classes to the breeders: Abyssinian cats could not be put under selective pressure for six mutations as it could be very dangerous. The preliminary screening of the six variants performed on unknown Abyssinians, showed that the rare alleles were present in the population, although with lower frequencies compared to the ones present in the affected cats. This could further support the hypothesis of multiple variants involved in the disease with different impacts: a cat could carry just one variant with a high impact and be affected, or it could carry three variants with

a moderate impact and be healthy. The main issue was represented by the assignment of the thresholds that identified the different risk classes. In fact, few cats were discordant between risk profile and disease status and many individuals resulted to be borderline in this scenario, making difficult the interpretation of the data for a risk prediction.

Concerning the Siamese study, only the variants detection with the WGS data was performed. The Siamese population was composed by too few samples to proceed with the genotyping of the variants and with the statistical analysis. To overcome this problem, a worldwide call was made to retrieve as many Siamese samples as possible to proceed with the investigations. By the way, an important result was obtained, as the different variants underlying Siamese amyloidosis resulted to be unique for the breed and not shared with Abyssinians, showing that the same disease is characterized by different backgrounds when the two breeds are considered.

Overall, the results obtained during the first study on both the breeds had two big limitations. The first and the most important one was represented by the version of the genome assembly that was used for the analysis. The version 6.2 was the first official cat genome assembly, obtained with a light genome coverage (12x). This means that a lot of gaps were still present, and they could have affected the variant calls. In fact, incorrect reads mapping can result in false negative or false positive variant calls. The other main issue was represented by the method that was chosen for variants detection, as only exonic variants with a high and/or moderate impact were considered, under the hypothesis of a monogenic disorder. This means that variants falling into intronic or intergenic regions and interfering with some regulatory regions or binding sites, were not taken into account. Moreover, the presence of discordances between some cat's phenotypes and risk profiles led to consider the possible presence of other factors involved in the development of the disease, such as environmental factors or the diet.

In the light of all these considerations, with the advent of the cat genome vs 9.0, all the WGS analysis for both the breeds were repeated using the new assembly. This time, the attempt to identify amyloidosis related variants was not only improved

by the realignment of the reads to the version 9.0 of the genome assembly, but it was also a result of the increased number of cats that joined the 99 Lives Initiative. Almost 70 whole genome cat sequences were added to the database, bringing to 197 the number of cats that could be used as controls for variants exclusion.

As it was also important to point out that other factors and genes could be involved in the onset of amyloidosis, variant detection was performed including also intronic and intergenic regions of the genome, and two other approaches were considered to support the genomic analysis. Both a proteomic and miRNA analysis were carried out in the Abyssinian breed due to the availability of samples while, for Siamese breed, only the variant detection was performed.

One of the main findings of the genomics analysis in both the breeds, was that the majority of the variants mapped in intronic and intergenic regions and those mapping within exonic regions did not match the variants found in the first study with the old version of the genome assembly. This could be addressed to the change of the assembly, to the increased number of cats in the WGS analysis or to the different criteria used to filter the variants. As said before, the gaps present in the version 6.2 of the cat genome could have led to both false positive and false negative variant calls. At the same time, the new version could have shed a light to the presence of new variants. The number of cats used for the analysis is extremely important. A variant that was unique for the affected cats in the first study could be shared, for example, by other cats among those recently added, becoming therefore not anymore significant. Finally, the study performed with the vs 9.0 of the genome assembly, used more stringent filters as it considered only those variants that were present in both the affected cats and it was based on the genotype frequencies. Instead, the first detection with vs 6.2 considered also variants that were present in at least one of the two affected cats and it filtered the variants based on their allele frequencies.

Overall, the analysis with the new version has to be considered the more accurate, especially because the intergenic and intronic regions were evaluated as well and the variants found in the 3'UTR regions resulted to be very important in the reconstruction of the amyloidosis pathway.

The integration of the genomics data with both proteomic and miRNA analysis was fundamental. In particular, miRNAs helped in matching all the results together as they turned out to be the common feature and the link between genomics and proteomics. What emerges from this study is that the affected cats are characterized by the presence of different downregulated miRNAs that are already reported in human literature to be involved in AD. These miRNAs target a great part of the genes coding for proteins that result to be expressed only in the affected individuals. They also target the genes that result to be mutated in the 3'UTR region from the genomic analysis, with one miRNA having his binding site exactly in the site of mutation of one gene. The match among these data helped in giving new insights in the reconstruction of the feline amyloidosis pathway and to highlight how this disorder resembles many of the mechanisms that have been described in the Alzheimer disease. This also brings to conclude that amyloidosis, as AD, is a very complex disorder with different factors and genes involved. The study performed through the integration of genomics, proteomics and miRNAomics allowed to collect a huge amount of data that can be stored and used later in time when the cat annotation will be improved. There are miRNAs that resulted to be exclusive feline miRNAs, having no match with other species, and their role has to be further investigated and clarified. At the same time, there are all the intronic variants that, with an improved annotation of the cat genome, could be addressed to regulatory regions or protein binding sites that could better reconstruct the mechanism that underlie the disease.

The other study conducted simultaneously using the NGS data of the 99 Lives Initiative, was the detection of CNVs using the whole genome sequences of 41 cats belonging to 14 different breeds. At the beginning, this work aimed to the identification of CNVs potentially associated to the amyloidosis phenotype in the Abyssinian and Siamese cats. Anyway, no correlation was found and no CNVs present exclusively in the affected cats were identified. However, through this study it was possible to demonstrate that individuals of different breeds can be grouped according to their CNVR similarities, and cats belonging to the same breed tend to cluster together as well as cats sharing the same geographical origins.

It has to be considered that this study represents the first effort for the identification of genomic structural variation in the domestic cat. The partially incomplete information related to gene annotation, as well as the identification of CNVs with different state when the output of the two software packages was compared, could be addressed again to the use of the version 6.2 of the cat genome assembly. Improvement in CNV detection could be achieved using the version 9.0 of the genome assembly, increasing the number of individuals per breed, and adding new breeds. By the way, this work represents the first CNV mapping in the *Felis catus* and could be used as a starting point for further investigation on structural variants in the domestic cat.

4.2. Concluding Remarks

The sequencing data available through the 99 Lives Initiative have demonstrated to be an important source for genomics analyses in cats, providing data that can be used for different kind of scientific researches. Through this PhD project, it was shown how these WGS data can be applied for different fields of study. First of all, the 99 Lives database can be used as a powerful tool to detect disease associated variants, as it was done for the feline amyloidosis. Several variants have been detected using the whole genome sequences of just two affected cats and comparing them to all the other cats within the database. Moreover, besides the studies concerning the associations between genomic variants and disease phenotypes, the sequencing data have been used to conduct a descriptive study, the first CNV mapping in the *Felis catus* species, which led to the characterization of 14 different cat breeds based on the presence of different CNV regions.

Finally, it has to be considered that the database includes also different wild felids, which represent an extremely valuable source for evolutionary studies and conservation genetics.

Genomics is moving forward and precision medicine represents the new frontier in medicine and health management. The 99 Lives initiative has given to cats the chance to be included in this brand-new kind of approach, bringing genomics into feline health-care too.

Acknowledgements

These three years of PhD have been among the most challenging and inspiring of my life.

I sincerely thank my Supervisor, Prof. Maria Longeri, for giving me the opportunity to get in touch with the world of research. I would like to express my gratitude for sharing her precious experience and for guiding me in the research path with valuable advices.

A special acknowledgment has to be addressed to my Co-supervisor, Prof. Leslie Lyons. She allowed me to experience research in a different and intriguing reality, on the other side of the ocean. Her availability and her teachings have been fundamental for writing this thesis.

My acknowledgements also go to all the people who contributed in the realization of this work.

Thanks to Paolo and Caterina who helped me in the managing of the samples and in all the laboratory practices.

Thanks to Maria Pia. The year spent together working on CNVs allowed me to discover the smart researcher and the reliable friend you are. Thanks also to Prof. Alessandro Bagnato, to my colleague Erica and to Cristina, who shared their time and knowledge during these years.

Thanks to the Biochemistry group of the Department of Veterinary Medicine at the University of Milan for their collaboration and their precious work. In particular, thanks to Prof. Gabriella Tedeschi and to Simona Nonnis, whose contribution was essential.

Thanks to the Bioinformatics group of the Institute of Biomedical Technologies at the National Research Council (CNR) of Milan. They introduced me to new concepts and new tools that will be extremely useful.

Thanks to Dr. Lyons' Feline Genetics and Comparative Medicine Laboratory. A special thanks to Dr. Barbara Gandolfi who spent great part of her time teaching me and pushing me to always do my best. She is not just a colleague, but she is a true

friend. Thanks also to Nicholas Gustafson, Reuben Buckley and Thomas Juba for their help and their support during my American days.

These three years were important not only for my professional growth, but also because they allowed me to meet amazing people. I thank my fellow labmates for the stimulating discussions, for the endless afternoons we worked together before deadlines, and for all the fun we had in the last three years.

Thanks to Alessandra and Valentina, my “beautiful minds”, whose support, affection and endless laughs lightened all the daily efforts. They are true friends and real partners in crime. They will always have a special place in my heart.

I am also grateful to all the lunches shared with Claudia, Elisa and Eleonora. You girls are amazing, and I could not imagine my days without your jokes and your humor. Even in the darkest of times, you have always found a way to get a smile back on my face.

Thanks also to Prof. Chiara Bazzocchi for her availability and for the hospitality she offered me in her laboratory.

Last but not least, I would like to thank my family, my parents and my friends for the endless support, for the strength they gave me in all the difficult moments and for relying on me before I could do it myself. I feel tremendously lucky to have all of you in my life.

