Canine oral melanoma genomic and transcriptomic study defines two molecular subgroups with different therapeutical targets

Anais Prouteau¹, Stephanie Mottier¹, Aline Primot¹, Edouard Cadieu¹, Laura Bachelot¹, Nadine Botherel¹, Florian Cabillic², Armel Houel¹, Laurence Cornevin², Camille Kergal¹, Sebastien Corge¹, Jerome Abadie³, Christophe Hitte¹, David Gilot¹, Kerstin Lindblad-Toh⁴,⁵, Catherine Andre¹, Thomas Derrien¹* and Benoit Hedan¹*

¹ University of Rennes, CNRS, IGDR—UMR 6290, F-35000 Rennes, France,
² Laboratoire de cytogénétique et biologie cellulaire, CHU de Rennes, INSERM, INRA, Univ. Rennes 1, Nutrition Metabolisms and Cancer, 35043 Rennes, France,
³ Laboniris, Department of Biology, Pathology and Food Sciences, Oniris, Nantes, France,
⁴ Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA,
⁵ Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Box582, SE-751 24 Uppsala, Sweden,
* T.D. and B.H should be considered co-corresponding authors

Email address: thomas.derrien@univ-rennes1.fr (T. Derrien), benoit.hedan@univ-rennes1.fr (B. Hedan)

*Corresponding author

Abstract: Mucosal melanoma (MM) is a rare and aggressive clinical cancer that occurs mostly in the head, neck, and anogenital regions. Despite recent advances in genetics and the development of revolutionary treatments, such as immunotherapy, the prognosis for MM remains poor. Canine MM shares several clinical, histological, and genetic features with its human counterpart, offering a relevant spontaneous and immunocompetent model to decipher the genetic bases and explore treatment options for human MM. We performed an integrative genomic and transcriptomic analysis of 32 canine MM samples, which allowed us to identify two molecular subgroups differing in microenvironment and structural variant (SV) content. The overexpression of genes related to the microenvironment and T-cell response was associated with tumors harboring a lower content of structural variants, whereas the overexpression of pigmentation-related pathways and oncogenes such as TERT were associated with a high SV burden. To detail the SVs, especially those with focal amplifications, whole-genome sequencing was performed on four canine MM cell lines. We showed that focal amplifications characterized complex chromosomal rearrangements targeting oncogenes such as MDM2 or CDK4 and a recurrently amplified region on canine chromosome 30, comprising the genes TRPML7, GABPB1, USP8, and SPPL2A, were candidate oncogenes for MM. We showed that the copy numbers of these genes were significantly correlated with their expression levels. Finally, we demonstrated that the genes TRPML7, GABPB1, and SPPL2A play a role in cell proliferation; thus, these may be considered new candidate oncogenes for human MM. Our findings suggest the existence of two MM molecular subgroups that may benefit from dedicated therapies, such as immune checkpoint inhibitors or targeted therapies. These results illustrate the relevance of dog models for deciphering genetic mechanisms in spontaneous MM, along with the potential to screen for efficient targeted therapies for rare and aggressive cancers in humans.

Keywords: Mucosal Melanoma, Dog model, oncogenes, immune checkpoint inhibitors, chromosomal rearrangements, CDK4, MDM2
1. Introduction

Melanoma is the deadliest skin cancer in humans, with an incidence of 80,000 cases per year in the USA (1). Mucosal melanoma (MM) is a rare clinical entity that accounts for approximately 1%–2% (2,3) of all melanomas; however, this rate is approximately 22% in certain Asian countries where cutaneous melanoma (CM) occurs less frequently (4). MM is caused by melanocytes that reside in the mucous membranes of the respiratory, gastrointestinal, and urogenital tracts, and occurs mainly in the head and neck (31%–55%), anorectal (17%–24%), and vulvovaginal (18%–40%) regions (3). Compared to CM, MM is highly aggressive and has a less favorable prognosis (2,5,6), with a 5-year survival rate of 20%–35% based on the disease location and stage (2,3,6,7). Treatment options include surgical resection and/or radiation therapy to achieve locoregional control; however, the prognosis remains poor. Some advanced cases of MM may benefit from targeted therapy, such as using KIT inhibitors, or immunotherapy using checkpoint inhibitors (2,3,8–10). However, the response to immunotherapy is highly variable, and patients with MM have a lower response rate than those with CM (3,8).

In recent years, large genomic studies involving whole exome sequencing (WES) and, more recently, whole genome sequencing (WGS) have been conducted to examine the mechanisms of MM tumor initiation, progression, and metastasis, and to find more suitable therapies (11–18). These studies have shown that MM presents a lower mutation burden and has a greater load of structural variants (SVs) than CM (11,12,14,15). In MM SV and CNA frequently involve known oncogenes such as CDK4, MDM2, and TERT (14–16,18). Two recent studies have identified two MM subgroups based on the pattern of complex clustered SVs (14,15). In oral melanomas comprising one such pattern, some of the SVs were linked to a poor outcome (15). Since MM is a rare entity in humans, popular breeds of dog patients with MM can be evaluated to better understand the molecular characteristics of these MM subgroups and to reveal relevant therapeutic targets.

In the last decade, dog models have emerged as being unique, spontaneous, and immunocompetent for human cancers (19–24), particularly for MM (13,22,23). Canine MM is the most frequently occurring oral malignancy in dogs and shares several clinical, biological, and genetic features with its human counterpart. As in humans, canine MM tends to rapidly metastasize to distant organs and responds poorly to chemotherapy. Immunotherapy in dogs has been developed to create a melanoma vaccine; however, the results remain controversial with highly variable responses (23,25,26). Recently, the development of anti PD-L1 (27,28) and anti-PD1 antibodies (26) for canine patients with MM has proven to be effective for MM in pilot studies. Genetically, canine MM mimics human MM with a predominance of somatic CNAs and complex SVs, and a relatively low tumor mutation burden (13,24,29).

In this study, we propose to examine the genetic basis of MM in a spontaneous canine model in specific breeds that are frequently affected by MM, by combining genomic and transcriptomic sequencing experiments together with functional validations. Unsupervised clustering analysis of RNaseq data from 32 canine MM samples highlighted the existence of two molecular subgroups of MM with differential expressions: one characterized by immune and microenvironment signatures, and the second by the overexpression of oncogenes such as MDM2, CDK4, and TERT, which was characterized by the enrichment of complex chromosomal rearrangements. The effect on cell proliferation of genes, which is frequently altered through chromosomal rearrangements, was explored in both canine and human MM cell lines. The results of the study suggest the existence of two MM molecular subgroups that could benefit from distinct therapies and could lead to further translational studies in human and veterinary oncology.

2. Material and methods

2.1. Sample collection

Blood and tissue biopsy samples from oral melanoma-affected dogs were collected for this study through Cani-DNA (http://dog-genetics.genouest.org) and the Canine Comparative Oncology Genomics Consortium (CCOGC) Biological Resource Centers BRC (www.cccogc.org) (61 and 18 cases, respectively) (Supplementary Table 1). Oral melanomas were confirmed based on the histopathological analysis (JA) with adequate immunostaining (expression of the S100, MelanA, PNL2 markers). The DNA and RNA were extracted as previously described (21,30). Blood and tissue sampling was performed by veterinarians on privately owned dogs during the course of the health follow-up, with the owner’s consent.

2.2. Canine Cell lines

The Bear (accession numbers CVCL_OD14) and CML10 (accession number CVCL_IZ11) cell lines were obtained from Dr. Modiano (Colorado University). Cell lines were cultured at 37°C in RPMI 1640 medium (Gibco, Amarillo, TX, USA) supplemented with 10% fetal bovine serum and 0.2% prrimocin (Invitrogen, San Diego, CA, USA). Five other cell lines (Dog-OralMel-18249, Dog-OralMel-18395, Dog-OralMel-18333, Dog-OralMel-18848, and Dog-OralMel-18657) were developed in the laboratory and were obtained from fresh canine oral melanoma samples (Supplementary Table 1). These cell lines were cultured in DMEM/F-12 medium (Gibco) supplemented with 10% fetal bovine serum and 0.2% primocin, and after ten passages, RPMI 1640 was used. All the cell lines were tested for mycoplasma using the MycoAlertTM Plus kit (Lonza, Rockland, ME, USA) and were mycoplasma-free.

2.3. Fluorescence in situ hybridization (FISH)

FISH analyses were performed on chromosome preparations generated from the cell lines using the following conventional techniques: colcemid arrest, hypotonic
treatment, and methanol/glacial acetic acid fixation, as described previously (31). The following bacterial artificial chromosome (BAC) clones were used: CH82-199H02 and CH82-179B09 for the MDM2 region, CH82-213B06 and CH82-204K11 for the CDK4 region, CH82-99P23 and CH82-60016 for the CFA 7 region (58.4Mb to 58.9Mb), CH82-1E17 and CH82-4015 for the region containing GABBP1, and USP8 and TRPM7 amplifications (https://bacpacresources.org/). These BAC clones were labeled using green-dUTP (Abbott Molecular, Des Plaines, IL, USA) and Cy3-dCTP (Amersham Biosciences, Chalfont, UK). The slides were analyzed by an experienced cytogeneticist (FC) using a fluorescence microscope (Axioskop2, Axio Imager Z2, Zeiss, Göttingen, Germany) and Isis imaging software (Metasystems, Altlussheim, Germany). At least 100 non-overlapping tumor nuclei were examined in our study.

2.4. RNAseq clustering and signature analysis.

For the 32 canine MM samples, polyadenylated RNAs were extracted, sequenced, and analyzed as described previously (30). All RNAseq fastq files are available in SRA under the BioProject PRJNA749900. Briefly, the pipeline used the “canFam3.1-plus” annotation as the reference annotation (32), and the canFam3.1 assembly version as the reference genome (33). Based on the protocol described by Djebali et al. (34), the FASTQ reads were aligned to the transcriptome and genome using the STAR program (v2.5.0a) (35). Finally, the gene expression levels were estimated as raw counts (unnormalized) using the RSEM program (v1.2.25) (36) for each sample individually, and subsequently merged to obtain a matrix expression file (with genes in rows and samples in columns). This matrix of expression was then normalized and transformed across all samples in order to stabilize the variance using the DESeq2 (v1.22.2) (37) function vstcounts with the option “blind = TRUE”.

To select the genes that would provide the most information for clustering analysis, we used 6,000 of most variable genes, that is, those with the highest median absolute deviation. Then, the nonnegative matrix factorization (NMF) algorithm from the NMF R package (v0.22.0) (38) was applied for different values of k clusters/ranks (k = 2–6). To determine the best k cluster, the nmdEstimateRank function was used (Supplementary Figure 1), and the consensus matrix method (Silhouette) identified k = 2 as the optimal number of clusters. Finally, we employed the NMF function extractFeatures to obtain cluster-specific gene signatures for the two clusters (Supplementary Table 2). These gene lists were then used as the input for the gprofiler2 (v0.1.9) R program (39) with the “gost” function, with the “organism” set to “c familiiarisis” to perform a gene set enrichment analysis over multiple databases (Gene Ontology, KEGG, CORUM).

Finally, heatmaps were plotted using the complexHeatmap software (v2.3.1) (40) to integrate and visualize multiple sources of information (gene signatures, sample clustering, SV content and oncogene mutational status) (https://github.com/tderrien/Prouteau_et_al).

2.5 Whole Genome Sequencing (WGS)

DNA from four canine oral melanoma cell lines and from the blood samples of two affected dogs were extracted for WGS, as previously described (21). The WGS was performed using the BGI sequencing platform (BGI, Shenzhen, China), using BGISEQ-500 short-read sequencing as previously described (41). Briefly, 1,000 ng of genomic DNA was quantified using a Qubit 3.0 fluorometer (Life Technologies, Paisley, UK), and was sheared using an E220 Covaris instrument (Covaris Inc., Woburn, MA, USA). Size selection was subsequently performed using a Vahsm DNA Clean beads kit (Vazyme, Nanjing, CHINA) to an average size of 200–400 bp. The selected fragments were end repaired and 3’ adenylated, and BGISEQ-500 platform-specific adaptors were ligated to the A-tailed fragments. The ligated fragments were purified and then amplified using PCR. Finally, circularization was performed to generate single-stranded DNA circles. After quantification, the libraries were loaded onto a sequencing flow cell and processed for 100 bp paired-end sequencing on the BGISEQ-500 platform.

2.6 Low pass Sequencing

Low-pass sequencing on 42 formalin-fixed paraffin-embedded tissue (FFPE) of melanoma was performed by the Psmagen company (Rockville, MD, USA). Briefly, the input DNA quality was verified using gel electrophoresis and the quantity was measured using the Picogreen assay (Thermo Scientific, Waltham, MA, USA). 2ng of DNA was prepared in 30ul of buffer and used for library construction using the Nextera DNA Flex Library Kit (Illumina, San Diego, CA, USA) according to the manufacturer’s guidelines. The size of the final DNA libraries was then validated using the TapeStation D1000 ScreenTape (Agilent, Santa Clara, CA, USA), and D1000 reagents (Agilent, Santa Clara, CA, USA). The quantity was measured using the Picogreen assay (Thermo Scientific, Waltham, MA, USA), and the molar concentration was calculated using both sources. The libraries were normalized to 2nM, pooled in equimolar volume, and then loaded on the flow cell from Novaseq S4 300cycle kit (Illumina, San Diego, CA, USA), and the XP-4lane kit (Illumina, San Diego, CA, USA). The prepared flow cell and SBS cartridge from Novaseq S4 300cycle kit (Illumina, San Diego, CA, USA) were inserted into the Novaseq 6000 system and sequenced using 151-10-10-151 running parameters, reaching a mean depth of 0.48X.

2.7 Mapping

After sequencing, the raw reads were filtered (adapter sequences, contamination, and low-quality reads were removed) according to the manufacturer’s guidelines. Sequence data were then aligned to the dog reference genome (assembly version canFam3.1) using the BWA-MEM (version 0.7.17) (H. Li and Durbin 2009), and PCR duplicate reads were removed using Picard tools (version 2.18.23) (http://broadinstitute.github.io/picard/). The read data were processed according to the GATK best
practices; specifically, the base quality score recalibration was assessed using GATK4 (version 4.0.12).

2.8 Somatic variant calling

The MuTect2 tool from the GATK4 software was then used to call somatic SNVs and short INDELS against a panel of normal (PON), comprising matched normal samples, and against germline variants from 722 dog genomes (https://data.broadinstitute.org/vgb/Ostrander_VCFs/722g.990_SNP_INDEL_chrAll.vcf.gz). Variant annotation was performed using the VEP program (43) with the Ensembl “Canis familiaris” annotation (v. 95).

2.9 Somatic CNV calling

The somatic copy-number variants (CNVs) were determined using the “R” Package DNAcopy (44). The canine genome was split into bins with a window size of 10 kb, or based on the exome target regions for WGS and WES. The number of reads was normalized according to the total number of reads per sample. The log2 ratio with the corresponding germlinal DNA, when available (or with a pool of three normal DNA for BEAR or CML10 cell lines), was estimated for each window. Several windows were merged (segmentation) into a larger segment with a significance threshold of $1 \times 10^4$, and the copy number of segments was estimated.

2.10 SV calling

Based on the aligned bam files from the BWA-MEM software, four structural variants were identified: breakends (BND), deletions (DEL), duplications (DUP), and inversions (INV) using the Delly software (45) (version v0.8.3). Only somatic SVs with the flags “PASS” and “PRECISE” were retained. Finally, circular representations of the SVs and the gain/loss from the DNA copies along the canine chromosomes were measured using the circlize R package (46) (version 0.4.10) (https://github.com/tderrien/Prouteau_et_al).

2.11 Modeling chromothripsis events

To assess the patterns of chromothripsis, we tested for the enrichment of chromosome-specific SVs by modeling the association between SV breakpoints and chromosome length using hypergeometric distributions (or a hypermetric distribution). More precisely, we defined “x” as the number of SV breakpoints on a specific chromosome out of a total of “k” detected breakpoints, where “m” was the chromosome size and “N” was the size of the canine genome. The computed $p$-value was the probability “x” SVs in the total number of SVs, calculated using $\text{hyperp}(x=m, m=m, n=N-m, k=k)$ in R. Statistical significance was set at $p<10^{-3}$. The clustering of the SV-breakpoints was tested as previously described (Korbel and Campbell, 2013). Regions of chromothripsis were detected using the Shatterproof software (48), using the copy-number and BND information along with the standard parameters.

2.12 Classification of dogs with “low” versus “high” content of structural variants (SV)

High-low SV classification was defined using the copy-number status of segments resulting from the WES data. SV samples were defined as “high” if they had an abundance of copy-number states oscillating on at least one chromosome (copy-number states $\geq 10$ states) (14), with at least one major amplification (copy number $\geq 90$ percentile of whole genome amplifications). This classification was verified using manual curation.

2.13 Targeted copy number and expression analysis

High-throughput quantitative PCR (qPCR) was performed using the SmartChip system (Takara Bio, Kusatsu, JAPAN) on 47 oral melanoma samples to detect focal amplifications on tumor DNA on CFA 10 and CFA 30. We used primer pairs targeting two genes on CFA 10 (MDM2 and CDK4) and four genes on CFA 30 (TRPM7, USP8, SPPL2A, and GABPB1) (Supplementary Table 3). A primer pair targeting a region of CFA 9 was used as an internal control, and each experiment was performed using DNA from an unaffected dog as an external control. The qPCR was performed on tumor DNA samples after pre-amplification using the SYBR green PCR master mix (Thermo Fisher Scientific, Waltham, MA, USA) on the 7900HT Fast Real-Time PCR System (Applied Biosystems, Waltham, MA, USA), using standard procedures. Each sample was measured three times, and the relative amounts of the sequences were determined using the $\Delta\Delta$Ct method.

RNA was extracted from the 47 tumor samples and reverse transcription was performed on 1 µg of RNA from tumorous or healthy tissue using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Waltham, MA, USA) according to the manufacturer’s instructions. RT-qPCR was performed on the same target genes. The gene SAP130 was used as the control housekeeping gene, and a pool of data from oral non-affected tissues of four dogs was used as the external control. Each sample was measured three times, and the relative amounts of the sequences were determined using the $\Delta\Delta$Ct method.

2.14 Western blot

Cellular protein extracts from cell lines were prepared using a cell lysis buffer containing 18 mmol/L Tris-Cl pH 7.5, 135 mmol/L NaCl, 0.9 mmol/L EDTA, 0.9 mmol/L EGTA and supplemented with 1 mmol/L PMSF, 1× EDTA-free cocktail protease inhibitor (Roche Diagnostics, Basel, SWISS), 30 mmol/L sodium fluoride, 40 mmol/L glycerol phosphate, 1 mmol/L sodium orthovanadate, and 0.5% Triton X-100.

Protein concentrations were determined using the BCA protein assay (Sigma-Aldrich) using bovine serum albumin as a standard. Protein samples were denatured for 10 min at 95°C, and equal amounts of cell proteins (50 µg) were subjected to 10% SDS-PAGE and transferred onto nitrocellulose membranes (Amersham-GEH Life, Buckinghamshire, UK). The membranes were probed with
the appropriate antibodies. The primary antibodies used were: anti-CDK4 (559693, BD Pharmingen, San Diego, CA, USA), anti-MDM2 (clone 2A10, ref MABE281, MerckMillipore, Darmstadt, GERMANY), and anti–ERK (sc-94, Santa Cruz Biotechnology, Dallas, TX, USA). Horseradish peroxidase-conjugated secondary antibodies were purchased from Jackson ImmunoResearch (West Grove, PA, USA). Signals were detected using the LAS-4000 Imager (Fuji Photo Film, Tokyo, JAPAN).

2.15 Transfection with specific and control siRNAs and cell proliferation assays

The siRNAs (80 nM) for CDK4, GABPB1, TRPM7, USP8, SPPL2A, and control siRNA (IDT) were transfected into the Bear cell line using Lipofectamine 2000 (Invitrogen, Waltham, MA, USA) by following the manufacturer’s recommendations (Supplementary Tables 1 and 3). CDK4 siRNA was used as the positive control. The siRNAs (10 nM) for GABPB1, TRPM7, and SPPL2A and control siRNA (IDT) were transfected into two human cell lines (HMV-2/CVCL_1282 -Merck-, WM3211/CVCL_6797 -Rockland-) using Lipofectamine 2000 (Invitrogen, Waltham, MA, USA) following the manufacturer’s recommendations (Supplementary Tables 1 and 3). Cells were seeded in 6-well plates in 1 ml of RPMI 1640 medium ( Gibco, Amarillo, TX, USA) supplemented with 10% fetal bovine serum and 0.2% primocin (Invivogen, Waltham, MA, USA) at a density of 1 × 10^6 cells, and incubated at 37°C. The medium was changed 6h post transfection to avoid toxicity.

Cell proliferation was evaluated in 96-well plates, 72h after transfection with an initial density of 30,000 cells per well, using a methylene blue colorimetric assay. Briefly, the cells were fixed for 30 min in 90% ethanol, removed, dried, and subsequently stained for 30 min using a 1% methylene blue dye in borate buffer. The fixed cells were washed 4–5 times using tap water, and 100 µl of 0.1 N HCl was added to each well. Cell density was analyzed using a spectrophotometer at 620 nm.

3. Results

3.1 Transcriptome landscape of canine oral melanoma highlights two distinct subgroups

We first investigated whether the expression profiles of the 32 canine MM tumors could predict the prognostic subtypes of MM (Supplementary Table 1). Using unsupervised clustering based on the nonnegative matrix factorization (NMF) algorithm (see Methods), we observed that the canine tumor samples could be separated into two distinct clusters (group 1, 12 samples; group 2, 20 samples) based on the underlying structure of their expression data (Figure 1 A). Since NMF organizes both the samples and genes, it also allows us to define a list of genes that represent the signatures of the two clusters (Supplementary Table 2). Using gene set enrichment analysis with the gprofiler tool (39), we showed that the 384 genes with the group 1 signature (Supplementary Table 3) were significantly associated with immune-related functional gene sets (Biological Process Term “Immune Response” p=1.6×10^−34, “cytokine-mediated signaling pathway” p=2.05×10^−15) (Supplementary Table 4). This suggests that the microenvironment differs between the two groups, with the MM in group 1 presenting a higher degree of immune cell infiltration. This was confirmed by the overexpression of genes encoding the components of the T-cell receptor (TCR) complex, such as CD3E, CD3D, CD3G, and CD247 (Supplementary Figure 2). Moreover, in this group, genes related to cytotoxic functions (such as granzyme B), the IFN-γ pathway (such as IRF1), or to the presence of an ongoing immune response and a cytokine-rich microenvironment (such as IL18, CCL3, CCL4, CLL13, and CCL22) (49) were significantly overexpressed (Supplementary Figure 2). The ratio between immune cell types with an immunostimulatory or immunosuppressive function seemed to be more important than the absolute number of immune cell types in determining the antitumorigenic vs. protumorigenic role of the microenvironment (49). Therefore, we determined the expression ratio of proinflammatory cytoktes (IFN-γ, IL-1α, IL-1β, and IL-2), versus immunosuppressive molecules (IL-10, IL-11, IL-4, and TGFβ1) in both groups. This ratio was significantly increased in group 1, suggesting a relative increase in pro-immunogenic responses in this group (p=0.03, Student’s t-test; Figure 1 B). This profile is concordant with that of “hot immune” tumors (51), and the overexpression of immune checkpoint genes such as CTLA4 (log2fold change=2.78; adjusted p=1.22×10^-5), TIM3 (log2fold change=1.6; adjusted p-value=2.2×10^-5), LAG3 (log2fold change=1.3; adjusted p-value=5.8×10^-3), or immunomodulating cytokines such as IL2RG. This suggests that this MM group may benefit from treatment with immune checkpoint inhibitors (50,51). In addition, several studies have involved the phenotype switching of melanoma cells as an escape route to CM-targeted therapies using BRAF inhibitors (52,53). Under the control of the microenvironment or intrinsic cell factors, melanoma cells can acquire a resistance to targeted therapies by switching from a proliferative to an invasive state (54,55). These changes to an aggressive phenotype are associated with dedifferentiation (from a melanocytic/transitory to a neural crest-like/undifferentiated phenotype) and an epithelial-to-mesenchymal-like (EMT-like) transition that promotes metastasis (56). Several genes associated with the acquisition of an invasive-dedifferentiated-EMT-like phenotype (54,57) were also overexpressed in group 1 (Supplementary Table 2).

In contrast, melanin metabolic processes and pigmentation pathways were identified in the second group (n=20 samples) Biological Process Term “melanin biosynthetic process” p=7.6×10^-4, “melanocyte differentiation” p=9.9×10^-7, “pigmentation” p= 1.8×10^-5 (Supplementary Table 4). Gene signatures from group 2 corresponded to a proliferative and differentiated phenotype of melanoma (DCT, MLANA, TYR).
This was reflected by the overexpression of the melanocyte-specific transcription factor MITF (log2fold change=2.2, adjusted p-value=1.0×10^{-5}) (Figure 1C), which is known to control the proliferation, migration, and invasion of melanoma cells in CMs. Interestingly, even if these CMs are highly proliferative, they are highly sensitive to targeted therapies, such as that using BRAF inhibitors (54). This group presented with the overexpression of well-recognized cancer driver genes such as TERT (log2fold change=1.5, adjusted p-value=0.028), KIT (log2fold change=2.1, adjusted p-value=5.0×10^{-5}) or the oncogene POM121 (log2fold change =0.67, adjusted p-value=0.011), which has recently been linked to poor prognosis in human oral MM (15).

The SVs involving MDM2 and CDK4 genes are hallmarks of human and canine MM (14–16,18), and a recent study in humans pointed out two different MM subgroups based on SV profiling (14). We used the data for WES that was previously performed on a subset of canine MM (Hitte et al, in prep) to analyze the SV profiles of the cohort in the present study. Based on the distribution of SV features (number, intensity, clustering), canine tumors were classified as having “low SV” (n=12) or “high SV” (n=20) (Figure 1A). Interestingly, the two subgroups defined by the transcriptomic analysis significantly overlapped with the two groups defined according to the SV status (exact Fisher test p=9.5×10^{-3}). The first transcriptome subgroup, that is, “group 1” in this study, which was characterized by an immune signature, comprised tumors with “low SV,” while the second transcriptome subgroup contained more “high SV” tumors. These results are concordant with previous studies in human cancers showing that chromosomal instability and CNAs are associated with a decrease in cytotoxic immune cell infiltration (59) and a resistance to checkpoint inhibitors (50,59,60). Our findings suggest that oral melanomas behave similarly and may evolve through two different paths driven by SV somatic changes that would lead to marked differences in therapeutic options: the first group containing “hot immune” tumors with a lower numbers of SVs would likely benefit from immunotherapy, while the second group, which contains “cold immune” tumors with a higher number of SVs content would respond better to therapies targeting amplified oncogenes, such as treatment using CDK4/6 inhibitors.

---

**Figure 1**: Transcriptomic analysis of 32 canine MM samples. A. NMF clustering of expression data identified two subgroups that overlap with SV content. The first subgroup is characterized by the expression of immune response and cytokine mediated signaling pathways and a low SV content, whereas the second subgroup is characterized by the expression of melanin metabolic processes and pigmentation pathways and is enriched in SV as well as focal amplifications of MDM2, CDK4, or CFA30:17Mb region. B. Expression ratio of proinflammatory cytokines (IFN-γ, IL-1A, IL-1B, and IL-2) and immunosuppressive molecules (IL-10, IL-11, and TGFB1) significantly differs between the two groups (p=0.03, Student test). C. Expression of immune checkpoint genes or known oncogenes according to the transcriptomic classification. The following immune checkpoint genes CTLA4, LAG3, TIM3 and IL2RG are overexpressed in group 1 while the oncogenes MITF, TERT, KIT and POM121 are overexpressed in group 2.
Given that SVs have been associated with resistance to immunotherapy in human cutaneous melanoma (50) and with a poor outcome in human (15) and canine oral MM (24), we focused on the characterization of these SVs in canine MM as models for therapies in human oral MM

3.2 “High SV” MM are characterized by focal amplifications and numerous chromosomal translocations

To refine the annotation of SVs, we performed WGS on the four canine MM cell lines. This allowed us to catalog a total of 5,906 SVs (DEL, DUP, INV, insertions, and BND) on the four cell lines (Supplementary Table 5). Three canine cell lines showed strong focal amplifications of CFA 30 and/or CFA 10 as recurrently observed in canine MM (13,24,29,61,62) (Figure 2). The Bear cell line showed focal amplifications of the CFA 30, targeting the 16-17 Mb region, (up to 67 copies) and of the CFA 10 targeting MDM2 and CDK4 region (up to 46 copies). The Dog-OralMel-18333 cell line was characterized by focal amplifications targeting the same region of CFA 30 (up to 27 copies) and CFA 7 (up to 13 copies), while the Dog-OralMel-18395 cell line presented strong focal amplifications on CFA 10 (up to 41 copies) targeting MDM2 and CDK4. In contrast, the last cell line, CML10, did not show any of these alterations.

Interestingly, when analyzing the SV distribution in the tumor genome, we found that the regions harboring focal amplifications comprised significant amounts of several types of SV, such as INVs, DELs, and insertions (Supplementary Table 6 and Figure 2). In particular, these focal amplifications were found to display inter-or intra-chromosomal translocations with significant breakpoint clustering, similar to what is seen in human MM, particularly, oral MM, which presents inter-and intra-chromosomal translocations between/within HSA 5 (TERT) and HSA 12 (CDK4) (14,16). Thus, for the three canine MM cell lines harboring strong focal amplifications, the amplified regions presenting breakpoint clustering (see Methods) were detected as potential chromothripsis regions using the Shatterproof tool (Supplementary Table 6). For the three cell lines harboring these focal amplifications, inter-or intra-chromosomal translocations were explored using FISH. It showed large genomic regions with repeated amplifications and fusions of CFA 30 and CFA 10 for the Bear, CFA 30 and CFA 7 for the Dog-OralMel-18333, and intra CFA10 for the Dog-OralMel-18395 cell lines (Figure 2 C,F,I). Although we could not find the presence of double-minute chromosomes to explain the highly elevated copy numbers using FISH analysis, the three cell lines harboring strong focal
amplifications presented features of chromothripsis as defined by Korbel et al (47), that is, breakpoint clustering and irregular oscillating copy-number states.

Thus, the results of this study suggest that focal amplifications of the CDK4 and MDM2 genes, and those in the 16-17 Mb region of CFA 30 in canine MM reflect major complex inter-and intra-chromosomal rearrangements. These massive DNA rearrangements clustered with high amplifications and deletions, result in long derivative chromosomes arising from cataclysmic events compatible with chromothripsis (14,47). These recurrent focal amplifications are expected to drive the oncogenic process of MM.

3.3 Exploring candidate oncogenes on CFA 30

3.3.1 Correlation between copy numbers (CNA) and expression of candidate oncogenes

Chromothripsis-like events often involve the amplification of oncogenes to promote tumor progression (63,64). In canine MM, these recurrent amplifications target well-recognized oncogenes such as CDK4, which are involved in the early phase of the cell cycle, and MDM2, whose protein inhibits p53, as well as several candidate oncogenes localized on CFA 30 amplifications, which remain unidentified in human MM. Based on the intersection of genomic intervals defining recurrent CFA30 amplifications in the WES data from 32 dogs we highlighted the minimal recurrently amplified region, containing four genes that may be relevant candidate oncogenes: TRPM7 is involved in the PI3K/AKT and MAPK pathways, as well as in EMT in ovarian and lung cancers (65–68); GABPB1 encodes the β subunit of a transcription factor and is linked to cell proliferation and poor outcome in renal cell carcinoma (69), USP8 has been linked to tumorigenesis and poor prognosis in several cancers (70,71), and SPPL2A plays a role in innate and adaptive immunity (72).

We hypothesized that the tumor advantages conferred by these focal amplifications were associated with the overexpression of the oncogenes targeted by these amplifications. Thus, we first checked whether the expression levels of the candidate oncogenes of CFA 30 (TRPM7, USP8, GABPB1, and SPPL2A) were correlated with the copy number of these genes. With this aim, we analyzed the copy numbers and gene expression of the candidate oncogenes in 47 novel canine MM samples by performing qPCR and RT-qPCR. A significant positive correlation between gene expression and copy number was found for all four CFA 30 candidate genes, with SPPL2A having the strongest correlation (Spearman coefficient: 0.76, \(p = 6.9 \times 10^{-9} \)). Concerning oncogenes located on CFA 10, the expression levels of MDM2 and CDK4 were positively correlated with DNA copy numbers (Spearman coefficient: 0.68, \(p = 8.9 \times 10^{-7} \) and Spearman coefficient: 0.65, \(p = 1 \times 10^{-6} \), respectively) (Figure 3 A). Finally, western blot analyses using cross-reactive available antibodies showed that the MDM2 and CDK4 focal amplifications were linked to a higher amount of protein (Figure 3 B).

Thus, amplifications and chromosomal rearrangements of CFA 10 and CFA 30 led to a clear overexpression of the targeted genes and protein levels, suggesting the acquisition of new cell capacities in connection with the proliferative and aggressive features of the MM cells.

3.3.2 Effect of candidate oncogenes on MM cell proliferation

To evaluate the effect of the altered expression of the CFA30 genes on cancer cell proliferation, we evaluated the proliferation capability of canine MM cancer cells after silencing the TRPM7, SPPL2A, GABPB1, and USP8 genes using colorimetric cell proliferation assays. Expression analysis showed that the siRNA experiments reduced the mRNA expression levels of the four oncogenes by 50%–72% (Supplementary Table 7), and the silencing of the TRPM7, SPPL2A, and GABPB1 genes significantly decreased cell proliferation from 18%–28% (Figure 3 C). These results suggest that the combined amplification of the CFA 30 candidate oncogenes might have a cumulative effect on MM cell proliferation.
3.3.3 Effect of candidate oncogenes on survival of canine MM

Recent studies have pointed out that complex clustered SVs in human MM or specific recurrent SVs (i.e., focal amplifications of CFA30) in canine MM are linked to poor outcomes (15,24). To explore the impact of the SV profiles on canine MM, the SV profiles were determined using low pass sequencing on 42 canine MM FFPE samples with available survival data (part of a previously published cohort, Prouteau et al.) (24). In this cohort, the minimal region of CFA 30 amplification was delimited between 16.2 Mb and 16.8 Mb. This region was observed to be amplified in 64.3% of cases and contained GABPB1, USP8, TRPM7, and SPPL2A genes. While high SV content in canine MM is associated with a poor outcome (one-sided log-rank p-value=0.025), the high SV is less significantly associated with a poor outcome than are the amplification of CFA30 candidate oncogenes are more significantly associated with a poor outcome (one-sided log-rank p-value=0.0012) (Supplementary Figure 3).

3.4 Exploring the value of CFA 30 oncogenes in human melanomas

The canine model of MM allowed us to identify new oncogenes; thus, we explored the involvement of these canine CFA 30 oncogenes in human cancers, more specifically, in melanomas, according to published data. Since TRPM7 and USP8 CRISPR inactivation has an impact on the majority of human cell lines (703/990 and 960/990 for TRPM7 and USP8, respectively), these genes are considered to be common, essential genes (https://depmap.org/). Moreover, a high expression of TRPM7, USP8, GABPB1, or SPPL2A is proven to be associated with shorter survival times in cutaneous melanomas (Log-rank TRPM7 p-value=0.005, Log-rank USP8 p-value=0.017, Log-rank GABPB1 p-value=0.019, Log-rank GABPB1 p-value=0.022, https://www.proteinatlas.org).

Human MM and acral melanomas are characterized by a higher SV content and a lower tumor mutational burden than UV-induced cutaneous melanomas; thus, they are more likely to present strong focal amplifications of the corresponding genes. In the Cancer Genome Atlas, we found one case of acral melanoma (TCGA-ER-A19T-01) harboring a strong focal amplification on chromosome 15 (orthologous to canine chromosome 30) encompassing the candidate oncogenes SPPL2A, USP8, TRPM7, and GABPB1. This amplification of 29 copies of these genes was associated with high gene overexpression (16.8, 22.7, and 30.2-fold for TRPM7, USP8, and SPPL2A, respectively). In another cohort of 34 patients with acral melanomas, we found one tumor with an intrachromosomal rearrangement on chromosome 15 targeting the genes TRPM7 and MYO5A (73). Further, two recent studies involving MM cases that performed WGS on 65 and 67 tumor samples showed that chromosome 15 frequently had a deletion in the proximal part, and more rarely, amplifications targeting the candidate orthologous region identified in the present study (14,15). Moreover, in a cohort of 65 patients with only oral melanoma, TRPM7 was one of the 48 significantly mutated genes (15).

Finally, to confirm the importance of these genes in non-UV-induced melanomas, we evaluated the proliferation capability of two human melanoma cell lines, one mucosal melanoma cell line (HMV-2/CVCL_1282), and one acral melanoma cell line (WM3211/CVCL_6797). Using siRNA to induce silencing of the TRPM7, SPPL2A, and GABPB1...
genes, we showed that their expression was decreased by 61%–69% (Supplementary Table 8). The silencing of TRPM7, SPPL2A, and GABPB1 significantly decreased cell proliferation from 10.8%–27.4% (Supplementary Figure 4).

These findings confirm that the candidate genes of the canine CFA30 16.2-16.8 Mb region / Human CFA15 50.3-50.9 Mb region are also involved in a subset of human non-UV-induced mucosal and acral melanomas. The results of this study suggest that the therapeutic target potential of these genes in human MM, especially those of the oral cavity, should be explored.

Discussion
In the last decade, dog models have been established as relevant models for clinical and genetic studies of human MM (13,24,74–77). The present work provided a better understanding of the genomic and transcriptomic profiles associated with canine MM by identifying two molecular subgroups of canine patients differing in their transcriptomic profiles and SV content. Thus, these findings suggest different tumor microenvironments for each subgroup that each requires different types of therapies. In addition, this work allowed the identification of new candidate oncogenes for MM, which should be of interest to human oncology.

The first molecular subgroup contained a majority of tumors with a relatively low SV content, with overexpression of genes related to the microenvironment, particularly, the T-cell response and cytotoxic functions. Thus, the likely presence of an effective T-cell infiltrate in tumors in this group, in addition to the overexpression of immune checkpoint proteins such as CTLA-4, TIM3, and LAG3 favors a response to immunotherapy, especially that using checkpoint inhibitors (50,51,59,78). While several studies have linked high tumor mutational burden to a better response to immunotherapy (60,79,80), two recent publications have shown that CNA/SV levels are also predictive of a response to immunotherapy in different cancer types. Among these cancers, cutaneous melanomas harboring a high level of CNAs have the poorest response rate to immunotherapy (50,59). Ock et al. also suggested that these alterations (CNA/SV) are stronger predictors of the response to immune checkpoints than tumor mutational burden (50). Following these studies, we suggest that the first molecular subgroup with low SV that was identified in this study may respond to immunotherapy, while the other subgroup, with higher SV rates, may respond better to targeted therapies (e.g., anti-CDK4). This hypothesis is further reinforced by the results of a previous study by Ock et al, who observed that the differentially expressed genes between the two groups are enriched in genes associated with response to immunotherapy in human cancers according to (50). We found similarities in terms of the expression modifications (up or downregulation) in the genes differing in expression between the two canine MM subgroups and the immunotherapy responders and non-responders in the study by Ock et al. (p=1.0×10^{-11}) (50) (Supplementary Table 9).

The association of different SV profiles between the two molecular subgroups of MM could reflect the involvement of different DNA repair mechanism alterations and thus, different tumorigenesis pathways. The second subgroup was characterized by the presence of recurrent focal high amplifications and numerous chromosomal rearrangements (“high SV”). This chromosomal instability has already been described in human melanomas and is associated with poor outcomes in cutaneous and oral melanomas (15,81). This aggressiveness was also observed in canine oral melanoma harboring the focal amplification on CFA 30 (median survival time of 159 days vs. 317 days for MM with CFA 30 amplification or not CFA30 amplification respectively, p=5×10^{-5}) (24), and we demonstrated that dogs with MM carrying a high SV burden had a poorer prognosis (one-sided p=0.0255) than those with MM carrying a lower SV burden.

Focal amplification in melanomas results from complex genomic catastrophes previously described as chromothripsis in humans (14,15,81) and dogs (29); however, most events are too complex to confidently assign to one particular type of mutational mechanisms (14). The WGS of four canine MM cell lines showed that the focal amplifications had complex rearrangements reminiscent of chromothripsis, with chained or clustered breakpoints localized to a subset of chromosomes in regions that also contained copy-number oscillations. The formation of double minutes is expected in chromothripsis to explain the high amplification of driver oncogenes. The FISH analysis in this study revealed large, homogeneously stained derivative chromosomes from CFA 10 and/or CFA 30, but no double-minute chromosomes. Similar patterns of high MDM2/CDK4 amplification with large derivative chromosomes were observed in human glioblastoma, and were assigned to chromothripsis, with aggregation of double minute chromosomes (64). In human acute myeloid leukemia, such catastrophic events have been linked to somatic TP53 mutations and, thus, p53 dysfunction (82). In dogs, approximately 50% of MM harbor an amplification of the TP53-inhibitor MDM2 (24), and this alteration is mutually exclusive with TP53 mutations (29). Similarly, inactivated p53 mutations (11–15,17,83–86) and MDM2 amplification (13–16) have been observed in human mucosal and acral melanoma, suggesting that p53 pathway dysregulation may be crucial in non-UV-induced melanoma development (29).

In human oral melanomas, recurrent interchromosomal translocations between HSA 5 (TERT) and HSA 12 (CDK4) were observed in 39.6% of cases (15). Similarly, over 50% of canine oral MM cases harbored recurrent focal amplification of CDK4/MDM2 oncogenes. However, unlike human MM, these amplifications did not co-occur with TERT DNA amplification, even though TERT RNA was significantly over-expressed in the “high SV” cases (log2ratio=2.54, p-corrected=6.7×10^{-7}). These results suggest that focal
amplifications of the CDK4/MDM2 oncogenes in human and canine MM drive the same oncogenic pathway activation by distinct mechanisms. Promoter enhancements TERT expression in human cutaneous melanoma (87), and these mutations create a de novo binding site for the transcription factor GABP heterotetramer, activating TERT transcription (88,89). In dogs, GABPB1 is overexpressed due to CFA 30 amplification; no TERT promoter mutations were observed in the four canine cell lines characterized by WGS. Similarly, Hendricks et al. (2018) did not find any TERT mutations in five canine MM cases analyzed using WGS to explain TERT overexpression. Other mechanisms could be involved in gene overexpression, such as epigenetic regulation or activation of the NFKB or JAK/STAT pathways (87,90,91).

Regarding the focal amplifications of CFA 30 (16-17 Mb region), the corresponding amplification of HSA 15 has already been observed in few human melanomas (73,81) (https://www.cancer.gov/tcga), and was associated with overexpression of the amplified genes (https://www.cancer.gov/tcga). Nevertheless, based on the high recurrence of this region amplifications in canine MM (~50% (24)), the CFA 30 region should contain important oncogenes for non-UV-induced melanoma development. The following lines of evidence support this hypothesis: i) in dogs, the degree of expression of candidate oncogenes of CFA 30 is correlated with the copy number, ii) these genes appear to be essential in a majority of human cell lines (TRPM7 and USP8), and (iii) they are associated with a shorter survival time in cutaneous melanomas. In the present study, we showed that TRPM7, SPPL2A, and GABPB1 are involved in the proliferation of canine and human non-UV-induced melanomas. We hypothesized that genes included in the CFA 30 amplifications have a cumulative effects on the tumoral transformation of melanocytes, and further studies are required to explore their roles, such as that in tumor proliferation, adhesion, and migration (92). TRPM7 appears to be an important gene for non-UV-induced melanomas as it was recently found to be involved in several other human cancers (92), was significantly mutated in human oral melanomas (15), and was also involved in intrachromosomal translocation in acral melanoma in a previous study (72).

Conclusion

Domestic dogs are increasingly being considered as a genetic system for the study of human cancers with underlying genetic components, owing to the strong breed-related predisposition to cancers. In this study, we identified two molecular subgroups of canine MM, differing in their SV content and gene expression profiles (immune microenvironment vs. pigmentation pathway and oncogenes), which may correspond to different therapeutic options. Moreover, this model allowed us to identify several oncogenes that are relevant for MM development, particularly TRPM7, which was recently proven to be involved in rare human non-UV-induced melanomas. These results suggest that it is important to permit the sub-setting of tumors to improve tumor stratification for clinical trials, which will likely have differing clinical outcomes, and to provide the optimum therapy depending on this classification for both canine and human medicine.

Acknowledgements:

The authors thank the veterinarians, who collected samples from melanoma affected dogs, together with their clinical data, and who provided anatomopathological diagnoses, especially Olivier Albaric, Laetitia Dorso and Florian Chocteau (Laboniris, Oniris, Ecole Nationale Vétérinaire de Nantes, France), Marie-Odile SEMIN (LAPV, Amboise, France), Caroline Laprie (Vei-Histo, Marseille, France), Marie Lagadic (Idexx Alfort, France), and Frédérique Degorce-Rubiale (LAPVSO, Toulouse, France). We also thank the dog owners for general health data and samples of their dogs.

Melanoma data were gathered through the Cani-DNA BRC (http://dog-genetics.genouest.org), which is part of the CRB-Anim PIA1 infra-structure, ANR-11-INBS-0003. Authors thank A. Fautrel (Platform H2P2 Biosit Rennes, France) for support with FFPE samples. Authors especially thank Dr. Patrick Devauchelle and Dr. Pauline de Fornel, (MICEN-Vet, Cretail, France) as well as Dr. Didier Lanore (Bordeaux) for their advice on veterinary oncology, and Marie-Dominique Galibert (IGDR, Rennes, France) for her interest and follow-up. We also warmly thank Clotilde de Brito, Laetitia Lagouette, Annabelle Garand, Anne Sophie Guillery and Frédérique Allais (IGDR, Rennes, France) for their input on canine oral melanomas, and Stéphane Dréano (IGDR, Rennes, France) for Sanger sequencing. Authors also thank Dr. Jaime Modiano (Colorado University) who provided the 2 oral melanoma cell lines (Bear -CVCL_OD14 and CML10 CVCL_IZ11).

We are most grateful to Biogenouest Genomics, the Human & Environmental Genomics core facility of Rennes (Biosis/OSUR) for its technical support, the GenOuest Bioinformatic core facility (https://www.genouest.org) for storing sequencing data, hosting the Cani-DNA website, and for providing the computing infrastructure.

Supports

This study was supported by CNRS and in the frame of the French Plan Cancer 2009-2013, by Aviesan/INSERM MTS 2012-08. A. Prouteau PhD (salary & direct costs) was funded through the FHU CAMIn (CHU-Univ Rennes-Brittany region & Ligue Nationale contre le cancer)

The sample collection performed through the French Cani-DNA BRC was funded by the CRB-Anim infrastructure ANR-11-INBS-0003 in the frame of the ‘Investing for the Future Program (PIA1).
Appendix

Supplementary Figure 1: Estimation of the best K cluster using the silhouette method.

Supplementary Figure 2: Expression of genes linked to infiltration of the immune cells according to the clustering groups. The following genes are overexpressed in group 1, reflecting significant immune cell infiltration in this group: CD3E, CD3D, CD3G, CD247, IRF1, IL18, CCL3, CCL4, CCL13 and CCL22.
Supplementary Figure 3: Kaplan Meier Survival Curves for death from MM according to SV profiles (A) (one sided log-rank \( p = 0.025 \), MDM2 amplification (B) (one sided log-rank \( p = 0.39 \), CDK4 amplification (C) (one sided log-rank \( p = 0.34 \) and CHR30 amplification (D) (one sided log-rank \( p = 0.0012 \)).

Supplementary Figure 4: Cell proliferation assay on two human non-UV induced melanoma cell lines (HMV2 and WM3211) showing the effect of the knockdown (siRNA) of TRPM7, GABPB1 and SPPL2A genes on cell survival.
References


43. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS,


65. Liu K, Xu S-H, Chen Z, Zeng Q-X, Li Z-J, Chen Z-M.


84. Sheng X, Kong Y, Li Y, Zhang Q, Si L, Cui C, et al. GNAQ and GNA11 mutations occur in 9.5% of mucosal melanoma and are associated with poor


