Genome-Wide Analysis of Long Non-Coding RNA Profiles in Canine Oral Melanomas

Christophe Hitte [1,*] , **Céline Le Béguec** [1] , **Edouard Cadieu** [1] , **Valentin Wucher** [2] , **Aline Primot** [1] , **Anaïs Prouteau** [1] , **Nadine Botherel** [1] , **Benoît Hédan** [1] , **Kerstin Lindblad-Toh** [3,4] , **Catherine André** [1] and **Thomas Derrien** [1,*]

[1] University of Rennes, CNRS, IGDR—UMR 6290, F-35000 Rennes, France; celine.lebeguec@gmail.com (C.L.B.); edouard.cadieu@univ-rennes1.fr (E.C.); aline.primot@free.fr (A.P.); anais.prouteau@univ-rennes1.fr (A.P.); nadine.botherel@univ-rennes1.fr (N.B.); benoit.hedan@univ-rennes1.fr (B.H.); catherine.andre@univ-rennes1.fr (C.A.)

[2] Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, 08003 Barcelona, Spain; v.wucher@gmail.com

[3] Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; kersli@broadinstitute.org

[4] Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Box 582, SE-751 24 Uppsala, Sweden

* Correspondence: hitte@univ-rennes1.fr (C.H.); tderrien@univ-rennes1.fr (T.D.); Tel.: +33-2-23-23-47-77 (C.H.); +33-2-23-23-65-34 (T.D.)

Abstract: Mucosal melanomas (MM) are rare aggressive cancers in humans, and one of the most common forms of oral cancers in dogs. Similar biological and histological features are shared between MM in both species, making dogs a powerful model for comparative oncology studies of melanomas. Although exome sequencing recently identified recurrent coding mutations in canine MM, little is known about changes in non-coding gene expression, and more particularly, in canine long non-coding RNAs (lncRNAs), which are commonly dysregulated in human cancers. Here, we sampled a large cohort ($n = 52$) of canine normal/tumor oral MM from three predisposed breeds (poodles, Labrador retrievers, and golden retrievers), and used deep transcriptome sequencing to identify more than 400 differentially expressed (DE) lncRNAs. We further prioritized candidate lncRNAs by comparative genomic analysis to pinpoint 26 dog–human conserved DE lncRNAs, including *SOX21-AS*, *ZEB2-AS*, and *CASC15* lncRNAs. Using unsupervised co-expression network analysis with coding genes, we inferred the potential functions of the DE lncRNAs, suggesting associations with cancer-related genes, cell cycle, and carbohydrate metabolism Gene Ontology (GO) terms. Finally, we exploited our multi-breed design to identify DE lncRNAs within breeds. This study provides a unique transcriptomic resource for studying oral melanoma in dogs, and highlights lncRNAs that may potentially be diagnostic or therapeutic targets for human and veterinary medicine.

Keywords: mucosal melanoma; dogs; transcriptome sequencing; long non-coding RNAs (lncRNAs)

1. Introduction

Mucosal melanomas (MM) are the most frequent form of melanomas in dogs, and they display more aggressive behavior in comparison to cutaneous melanomas. Dogs are spontaneously affected, with specific breeds developing MM with clinical features that are similar to human melanomas [1]. Dog breeds with high melanoma risk have been proposed as relevant natural models for the comparative oncology of melanomas, especially for deciphering their non-UV-dependent pathways, and for developing clinical trials that are based on homologous melanoma subtypes [1,2].

Recently, genomic studies have been conducted, to identify driver genomic alterations that are involved in canine MM [3,4], using exome sequencing to focus on the genetic landscape of somatic mutations in protein-coding genes (messenger RNAs; mRNAs). A consequence of cumulative genetic and epigenetic alterations in coding and non-coding genes is reflected by the study of gene expression, which has not yet been investigated in detail in canine model cancers. Despite the recent identification of thousands of canine long non-coding RNAs (lncRNA) [5,6], little is known about their impact in dog cancers, although they constitute an extensive component of dog genomes [7–9]. In humans, lncRNA expression is recurrently altered in many types of cancers [10–12], including melanomas [13]. From the first annotation of the melanoma-associated lncRNA *SPRY4-IT1* [14] to the recent identification of recurrent amplifications of *SAMMSON*, a dozen of lncRNAs have been functionally validated in cutaneous melanomas [15]. Because lncRNAs are expressed in a tissue-specific manner in both humans [8,16] and dogs [6], they represent a vast and still unexplored repertoire of potential targets and/or biomarkers for comparative oncology approaches.

Here, we analyzed a large cohort of canine MM transcriptomes from three breeds, sampled from the oral cavity (*n* = 52). We quantified differential gene expression by controlling for cell heterogeneity, using a signature-based method, and we assessed transcriptional networks by using co-expression analysis. We showed that lncRNA expression profiles discriminate between normal and tumor samples, and identified a significant amount of deregulation for 400 lncRNAs. Gene-set enrichment analyses were performed using co-expression networks of lncRNA:mRNA, to acquire associated GO enrichments for all-breed and breed-specific DE lncRNAs. Furthermore, we conducted dog-human orthologous relationship analyses to identify conserved lncRNAs, with potential interest in human melanomas. Altogether, this study provides an in-depth characterization of lncRNAs that are dysregulated in canine oral melanomas, and prioritizes potential biomarker lncRNAs by investigating their conservation and co-expression networks. Our findings provide a novel transcriptomic resource with detailed sample characterization for the comparative oncology of melanomas in dogs and humans.

2. Materials and Methods

2.1. Canine RNA Samples: Extraction and Sequencing

In total, 39 dogs from three breeds (GRET: golden retrievers, LABR: Labrador retrievers, and PODL: poodles) were sampled with either tumor-only (*n* = 26) or matched tumor/normal samples (*n* = 13 × 2) (totaling 52 samples) from the two biobanks, "Cani-DNA_BRC", which is part of the CRB-Anim infrastructure, and the Canine Comparative Oncology and Genomics Consortium (CCOGC). Samples were collected in the course of the health management of the dogs, by DVM veterinarians, with the owner's consent, and the diagnosis was performed through histopathological analyses (CNRS ethical board, France (35-238-13)). Material was collected at surgery, then stored in RNAlater, and the diagnosis of mucosal melanoma was evaluated by specialized veterinarians after histological examinations of the samples.

For all of the 52 samples, RNAs were extracted by using RNA II NucleoSpin Kits according to the manufacturer's instructions (Macherey-Nagel, Hoerdt, France) then polyadenylated RNAs (polyA+) were selected and sequenced at the BROAD sequencing platform, in a paired-end and stranded fashion, using HiSeq-2000 Illumina technology (BROAD Genomics Platform, Cambridge, MA, USA), at a mean depth of 107.4 million reads per sample. The RNA-Seq data is available in European Nucleotide Archive.

We used the "canFam3.1-plus" annotation [5,6] containing 10,444 lncRNAs and 21,810 protein-coding genes as the reference annotation, and the canFam3.1 assembly version as the genome reference [16]. Based on the protocol described in Djebali et al. [17], FASTQ reads were aligned, both on the transcriptome and on the genome, using the STAR program (v2.5.0a) [18]. Finally, gene and isoform expression levels were estimated in both normalized (TPM: transcripts per million) and un-normalized

(raw count as required by DE tools) with the RSEM program (v1.2.25) [19] for each sample individually, and then merged in order to obtain a matrix expression file, with genes in rows and samples in columns.

2.2. Analysis by DESeq2 using a Multi-Factor Design including Cell-Type Heterogeneity

The matrix of reads counts, including lncRNAs and mRNA genes, was used by DESeq2 (v1.22.2) [20] to compute differentially expressed genes. Given the cellular heterogeneity between healthy and tumor samples, we used the xCell program (v1.12) [21] to compute cell-type enrichment from our gene expression data, based on the reference signature set of 64 immune and stroma human cell types (Supplementary Materials Figure S1). The cell-type enrichment scores for keratinocytes, melanocytes, and skeletal muscle cells were then included in the DESeq2 design, in order to specifically control for DE genes involved in the tumor condition, and not in the differentiation between cell types (e.g., keratinocytes versus melanocytes). To control for other covariates, we included breed and sex information in our design, resulting in the following DESeq2 formula: design = ~sex + cell_types + breed + condition, with the condition here being the status of the sample (with normal tissue being a control for cell type heterogeneity versus tumoral tissue). To take into account low gene counts (which are especially the case for lncRNAs), we used the recently developed lfcShrink method with the type = apeglm option [22] to better estimate log-fold changes for poorly expressed genes. To test whether the log-fold change linked to oral melanoma was different between breeds, we added an interaction to the design, such as breed:condition.

2.3. Identification of Human Orthologous lncRNAs

For each canine lncRNA gene belonging to the "canFam3.1-plus" annotation, we projected all of its exons onto the canine genome, resulting in one representative "meta-transcript" sequence per gene. These sequences were then mapped onto the human genome assembly version GRCh38, using minimap2 [23] with the following parameters -ax splice -t16, and only primary alignments being retained in the case of multiple mappings. Based on the CIGAR field, sequence identity was defined as the number of matching bases over the number of alignment columns. Finally, human orthologous coordinates were compared with the GENCODE (v29) annotation of the lncRNA exons [8,24] using the bedtools [25] intersect program (after BAM to BED12 file format conversion) with the following parameters: -s -split, in order to assign orthologous relationships.

2.4. Weighted Gene Coexpression Network Analysis

A weighted gene coexpression network analysis (WGCNA) was carried out on the 52 RNA-Seq reads, using the R package WGCNA 1.66 [26]. The program utilizes a similarity measure to summarize the relationship between all pairs of genes, using expression data normalized as TPM to create a correlation matrix. We used the signed WGCNA coexpression measure. To identify coexpression modules, we used the 'soft-thresholding procedure'. WGCNA utilizes a similarity measure to summarize the relationship between all pairs of gene expression data across the data set, to create a correlation matrix. Co-expression modules are defined as branches of a cluster tree, using a dynamic branch-cutting approach. Therefore, co-expression modules are clusters of co-expressed genes identified by hierarchical cluster analysis.

Constructing a weighted gene network entails the use of a soft-threshold score that assigns a connection weight to each gene pair. The co-expression similarity is raised against the soft thresholding power, in order to calculate adjacency. For soft thresholding, we used the two adjacency functions that convert the co-expression measure to a connection weight. First, the scale-free fit index is a function of the soft-thresholding power. Second, the mean connectivity is a function of the soft-thresholding power. We set the soft-threshold to 7, to avoid the selection of an arbitrary cut-off. The weighted separation of co-expression was achieved by the transformation of the correlation matrix in an adjacency matrix, using default values. Gene profiles that had a low expression and/or did not vary sufficiently across each of the data sets were eliminated. A total of 3,830 genes met these criteria. We performed principal

component analysis, and used the first principal component (module eigengene; ME) to summarize the standardized module expression data.

To assess the potential associations between coexpressed gene modules and the melanoma condition, a single-column vector of clinical data for each breed and for all breeds considered together was defined and utilized. An association analysis was performed by using the module-trait WGCNA method to perform correlation analysis of the ME with clinical traits. Correlations and the corresponding *p*-values allowed for an inspection of the most significant associations.

Intramodular analysis was performed to identify genes with high gene significance and module membership measures, as recommended by WCGNA procedures. Genes with high significance (>0.5) for each variable, as well as high module membership (>0.5) in interesting modules were extracted.

2.5. Gene Set Enrichment Analysis

We conducted Gene Set Enrichment Analysis using the GSEA webserver [27], to construct meaningful annotation from the GO of genes (mRNAs), defined a priori by the WGCNA modules. The ontology that was used covered the domain of biological processes (BP).

3. Results

### 3.1. Whole-Transcriptome Sequencing of Oral Melanomas

We sampled 39 oral melanomas from three breeds (16 golden retrievers, 13 Labrador retrievers and 10 poodles) that were classified with respect to their oral melanoma locations, which included the tongue for 26% of the annotated cases, followed by the maxilla (18%) (Supplementary Materials Table S1).

Combining healthy and tumor samples, 5.7 billion sequencing reads were generated, with an average of 107 million reads per sample (Supplementary Materials Table S2). After quality control and trimming of the adapters, between 89% and 96% of the reads could be mapped onto both the canine genome assembly (canFam3.1) and the "canFam3.1-plus" annotation [5,6], using the state-of-the-art bioinformatic protocol described in Djebali et al. [17]. Amongst the lncRNA genes, we focused on long intergenic ncRNAs (lincRNA; $n = 5651$) and antisense lncRNAs (antisense; $n = 4793$), thus removing sense intronic lncRNAs which may correspond to the misannotation of coding alternative isoforms, and observed that 59.0% and 58.5% respectively could be considered as being expressed, using a soft filter of 10 reads in total per gene. In comparison, 87.6% of the total number of protein coding genes ($n = 21,810$) were retained, using the same threshold.

3.2. Analysis of Differentially Expressed Genes (DEG) in Mucosal Melanomas

We first performed quality control of the samples by using a PCA with all gene counts, as normalized by the DESeq2 program (size factors normalization) (Figure 1a). This revealed a clear distinction of the samples, with the first principal component distinguishing the control from the tumor samples in the three breeds. A similar distribution was observed when taking into account only the lncRNA-normalized counts, although the percentage of the explained variance was slightly lower (Supplementary Materials Figure S2). We next used DESeq2 to identify differentially expressed genes (both lncRNAs and mRNAs) by controlling for specific covariates: breed, sex, and cell-type heterogeneity between the samples (see Methods). For the latter, expression data was incorporated into the xCell program [21], and samples were then clustered according to their enrichment within the 64 cell-type signatures used by the program (Supplementary Materials Figure S1). Control samples were found to be enriched in keratinocyte-like and skeletal muscle cells, while tumor samples tended to be enriched in melanocyte cells. Using this multi-factor experimental design, we identified 417 differentially expressed lncRNAs between tumor and control samples, using an absolute log2 fold-change (|lFC|) > 1.5, and an adjusted *p*-values (p_{adj}) < 0.05 (see methods) (Figure 1b). From a cross-check of the DE analysis, we found that the *MDM2* proto-oncogene, shown to be recurrently

gained in human non-cutaneous melanomas [28], was almost four times more highly expressed in canine oral melanoma tumors than in controls (lFC = 1.96; p_{adj} = 0.02). Similarly, we observed a significantly lower expression of the *BUB1* gene in our cohort of canine melanomas (lFC = −1.06; p_{adj} = 0.02), in accordance with recent findings showing recurrent deletions of *BUB1* in mucosal melanomas using a cross-species strategy, including human, horse, and dog samples [3]. Among the 417 DE lncRNAs, 272 lncRNAs were down-regulated and 145 were up-regulated.

(a)   (b)
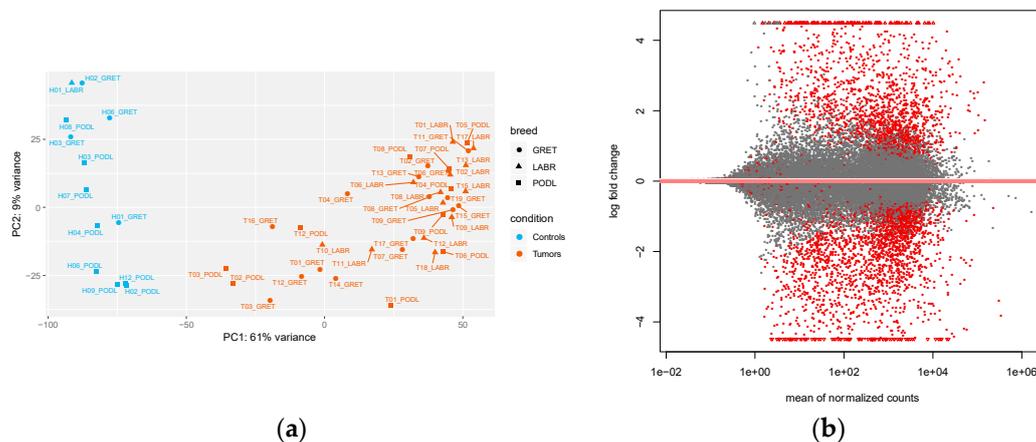
Figure 1. Expression analysis of the 52 oral melanoma samples. (**a**) Principal component analysis (PCA) of the 52 samples, based on gene-normalized counts, with control and tumor samples in blue and orange, respectively; (**b**) M-A plot representing log2-fold gene changes between tumors and controls over the mean of the normalized counts, with red points corresponding to significantly DE genes, with an adjusted *p*-value < 0.05, and without a log-fold change threshold; genes falling outside of the window are plotted as open triangles.

Although most canine lncRNAs have not yet been functionally characterized, one notable exception was given by the lncRNA *ZEB2-AS*, transcribed in an antisense orientation to the *ZEB2* mRNA, which was almost 14 times more highly expressed in tumors compared to normal tissues (lFC = 3.79, p_{adj} = 2.7×10^{-8}). Interestingly, this lncRNA has been shown to be involved in the regulation of *ZEB2* mRNA during the epithelial–mesenchymal transition (EMT) in human cell lines [29].

3.3. Comparative Genomics of Canine Differentially Expressed lncRNAs

Previous comparative genomic analysis [6] allowed us to identify a set of orthologous lncRNAs between human and dog, using a synteny-based approach. Here, we sought to annotate novel orthologous lncRNAs between dog and human by directly mapping DE lncRNAs sequences onto the human genome, using the minimap2 program [23] (see Methods). With the human genome assembly version GRCh38 defined as the target sequence, we aligned 33% of canine DE lncRNAs (*n* = 140) with a minimum identity of 70%. Amongst those, 26 matched to an already GENCODE-annotated [24] non-coding gene (Table 1). Most notably, we showed that several cancer-associated annotated lncRNAs in human are differentially expressed in canine mucosal melanomas. This included the down-regulation of *SOX21-AS1* (lFC = −2.97, p_{adj} = 0.003) (Figure 2a), already shown to be silenced in oral cancers [30], and the overexpression of the *CASC15* gene (Cancer Susceptibility Candidate 15) (lFC = 3.3, p_{adj} = 2.8×10^{-5}) (Figure 2b), whose RNA level has also been linked to cutaneous melanoma and phenotype switching in humans [31]. This analysis also shed light on 114 canine DE lncRNAs, which aligned to the human genome (identity > 70%) but without any known annotated transcripts by GENCODE, potentially highlighting novel human lncRNAs (Supplementary Materials Table S3).

Table 1. List of DE lncRNAs conserved with human GENCODE non-coding genes. Genes are ordered by ascending log2-fold change (lFC).

| canfam3.1+_id	Dog EnsEMBL ID	Dog Gene Biotype	lFC	Human Gene Name	Dog/Human Identity
RLOC_00034858	NA	lincRNA	−3.693	*AC016903.1*	0.770
RLOC_00028807	NA	antisense	−3.409	*AC010503.4*	0.740
RLOC_00001518	NA	lincRNA	−3.055	*EPHA1-AS1*	0.717
RLOC_00010776	NA	lincRNA	−2.974	*SOX21-AS1*	0.732
RLOC_00026330	NA	lincRNA	−2.669	*MIR29B2CHG*	0.804
RLOC_00030709	NA	antisense	−2.659	*LINC02586*	0.709
RLOC_00012258	NA	lincRNA	−2.258	*TOB1-AS1*	0.717
RLOC_00019548	NA	lincRNA	−2.214	*AL049536.1*	0.768
RLOC_00015465	NA	lincRNA	−2.201	*LINC01588*	0.794
RLOC_00011768	NA	antisense	−2.050	*AC005821.1*	0.714
RLOC_00011720	NA	antisense	−1.597	*LINC02079*	0.969
RLOC_00014809	NA	lincRNA	1.698	*AC062015.1*	0.923
RLOC_00002398	NA	antisense	1.782	*NR2F1-AS1*	0.820
RLOC_00032616	NA	lincRNA	1.868	*HOXD-AS2*	0.754
RLOC_00023326	NA	antisense	1.989	*RASSF8-AS1*	0.713
RLOC_00032620	NA	antisense	2.030	*HAGLR*	0.786
RLOC_00020381	NA	lincRNA	2.188	*TRAM2-AS1*	0.745
RLOC_00024264	NA	lincRNA	2.634	*AC133644.3*	0.979
RLOC_00021953	NA	lincRNA	3.359	*CASC15*	0.942
RLOC_00011077	NA	lincRNA	3.673	*LINC01301*	0.780
RLOC_00008433	ENSCAFG00000028700 (ZEB2-AS1)	lincRNA	3.796	*ZEB2-AS1*	0.839
RLOC_00013073	NA	lincRNA	3.797	*AC006450.3*	0.746
RLOC_00018365	NA	lincRNA	4.403	*AC090692.1*	0.755
RLOC_00022953	NA	antisense	4.910	*HOXC-AS3*	0.757
RLOC_00002254	NA	antisense	4.958	*STARD4-AS1*	0.903
RLOC_00025419	NA	lincRNA	5.758	*SNAP25-AS1*	0.987
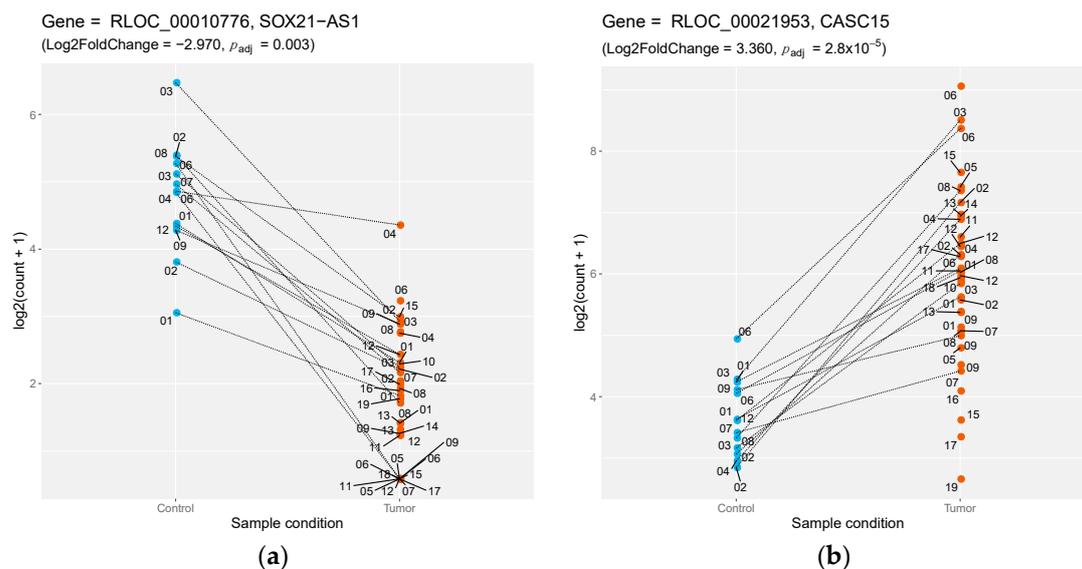
Figure 2. Differential expression of dog–human-conserved lncRNAs. (**a**) Down-regulation of the *SOX21-AS1* lncRNA between control samples (blue) versus tumor samples (orange), with the log2 of normalized counts on the *y*-axis; lines connect matched samples from the same individuals. (**b**) Same representation for the up-regulation of the lncRNA *CASC15*.

3.4. Inferring Functions of Differentially Expressed lncRNAs

We conducted an unsupervised expression analysis of lncRNAs, utilizing a WGCNA [26] based on the 52 RNA-Seq samples. The advantage of WGCNA is that it transforms gene expression data into co-expression modules, providing insights into signaling networks that may be responsible for the development and progression of oral melanomas. We included protein-coding genes ($n = 21,810$) to identify coexpressed modules that reveal relationships between lncRNAs and mRNAs, suggest common biological roles, and inform potential roles for lncRNAs.

3.4.1. Correlating Transcriptional Networks and Traits using Co-Expression Analysis

In the initial phase of the WGCNA, we identified 59 coexpression modules in an unsupervised manner. Hierarchical clustering analysis was performed, and a dendrogram was used to represent coexpression modules, as shown by color assignments (Figure 3). The coexpression modules included 121 lncRNAs on average (a range of 10 to 627).
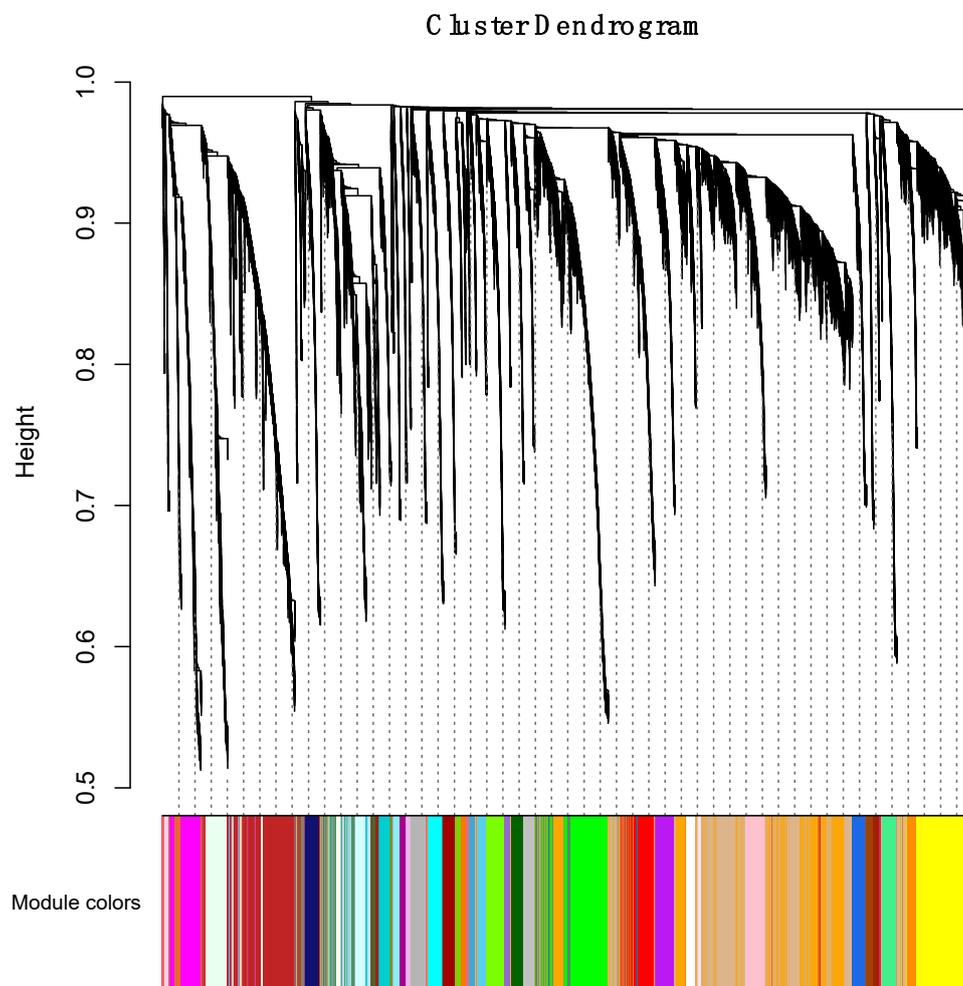
Figure 3. Clustering dendrogram. A total of 59 coexpression modules were constructed with assigned module colors at the bottom. The number of lncRNAs in the 59 modules is listed in Supplementary Materials Table S4.

We further performed the identification of coexpression modules that are associated with oral mucosal melanoma from all samples, through the calculation of Pearson's correlation coefficient (PCC). We further carried out intramodular analysis to identify genes with the highest significance association with MM, as well as a quantitative measure of membership in the module given by the correlation

of the eigengene module with the gene expression profile. We identified four modules (hereafter termed brown, medium-orchid, yellow, and tan) that were significantly associated with the melanoma status, with two modules being positively associated (PCC = +0.64, $p = 6 \times 10^{-7}$ ME yellow; PCC= +0.54, $p = 5 \times 10^{-5}$ for ME Tan), while the two other modules showed significant but opposite PCC associations with melanoma (ME brown, PCC = −0.90, $p = 8 \times 10^{-20}$; ME medium-orchid, PCC = −0.87, $p = 3 \times 10^{-16}$) (Figure 4).
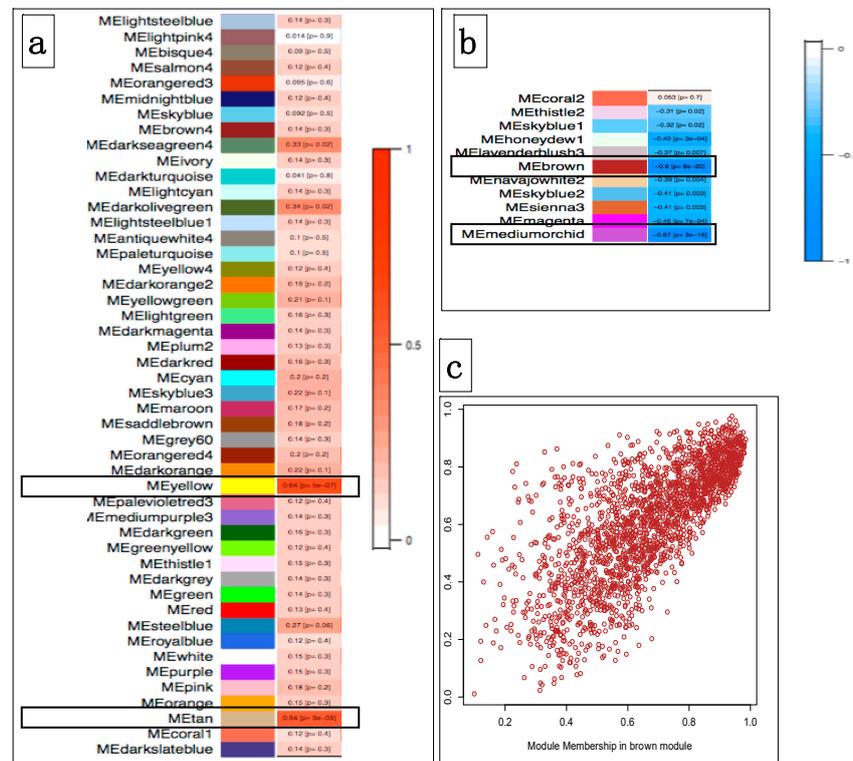
Figure 4. Module–trait associations. (**a**) Each row corresponds to a ME (module eigengene), and the column to the mucosal melanoma trait. Each cell contains the corresponding correlation and *p*-value with melanoma. Each correlation is color-coded according to the strength of the correlation, with a red gradient for positive correlations (red bar in 4.a). Modules Yellow and Tan are positively correlated ($p < 5 \times 10^{-5}$). (**b**) Modules with negative correlations according to the strength of the correlation; (blue bar in 4.b). Module Brown and Medium-orchid are the most significantly negatively correlated ($p < 1 \times 10^{-16}$). (**c**) Scatterplot of gene significance for mucosal melanoma status vs module membership for the Brown module. It shows a highly significant correlation between gene significance and Module membership in this module.

We considered only lncRNAs identified by the DE analysis in the coexpression analysis (*n* = 417) to overcome the heterogeneity bias between tumor and control cell types. A total of 215 DE lncRNAs (51.5%) also belonged to coexpressed modules with significant PCCs with melanoma status, such as the dog–human-conserved lncRNA *SOX21-AS1* which was found to be down-regulated in dog MM. In light of their correlations with cancer, dysregulated lncRNAs were classified into two categories; 30 belonged to modules with significant positive correlations, and 185 were in modules that yielded significant although opposite PCC.

3.4.2. Using Transcriptional Networks for Inferring lncRNA Functions

We used the lncRNA:mRNA correlated transcriptional networks constructed by WGCNA to infer the main functions of the lncRNAs, using the 'guilt-by-association' principle [32]. The functional implication of coexpressed mRNAs within the four modules (brown, medium-orchid, yellow, and tan)

that were significantly associated with MM was evaluated via gene set enrichment analysis, using the GSEA tool [27]. Our data showed that both positive and negative modules were significantly associated with specific but distinct GO biology process terms. As shown in Figure 5, genes involved in the positively associated module were enriched for GO terms involved in "cell cycle", "cell cycle process" or "mitotic cell cycle" for the yellow module, and "chromosome organization", "cellular response to stress", and "DNA metabolic process" for the tan module. These GO terms are connected with cancer, and implicated the replication and segregation of genetic material, and progression through the phases of the mitotic cell cycle.

(**a**)

(**b**)
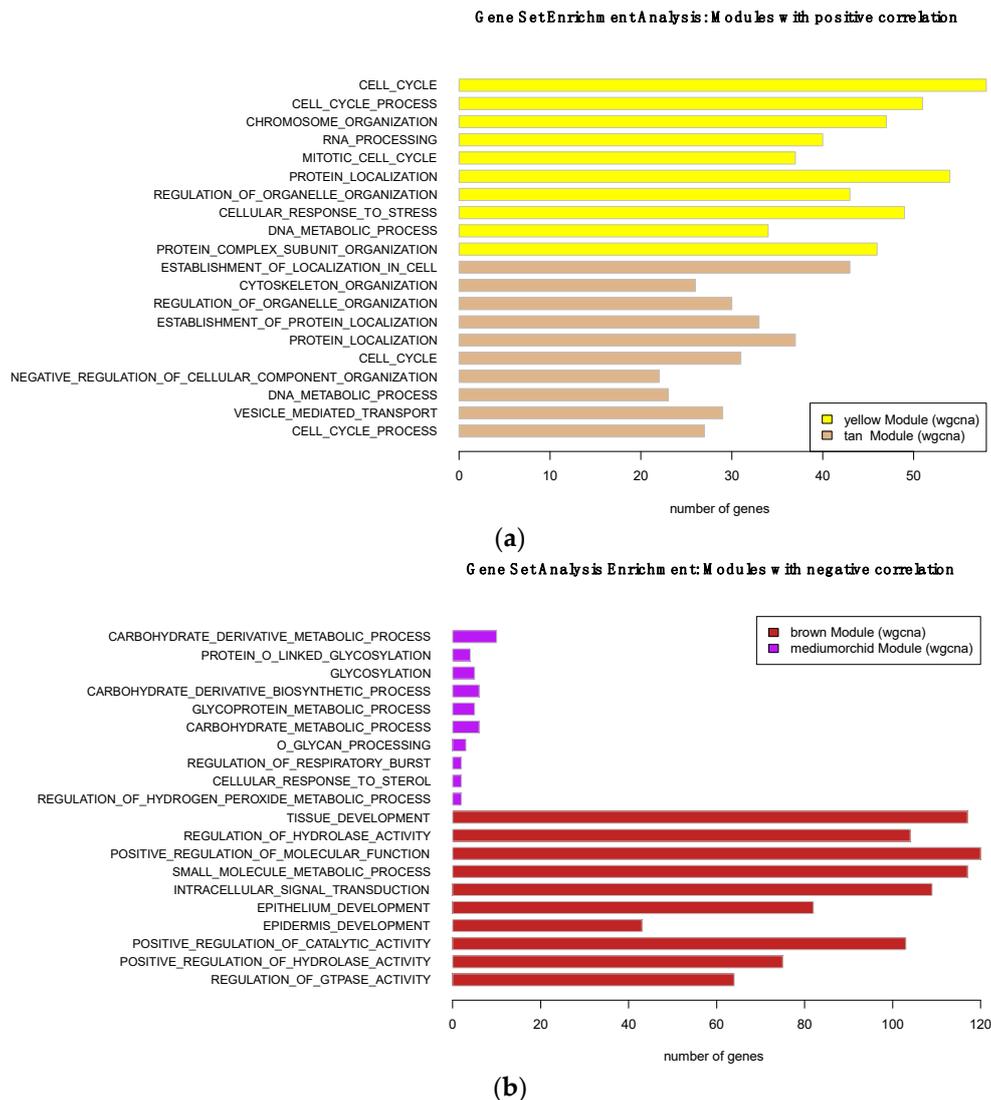
Figure 5. GO terms (Biological Process) enriched for (**a**) positively correlated and (**b**) negatively correlated modules with oral melanoma: the top ten enriched GO items are represented.

Conversely, genes of the negatively correlated modules were mainly enriched in "tissue development", "epithelium development", and "epidermis development" for the brown module, and mostly in "carbohydrate metabolic process" for the medium-orchid module. These categories reflect processes whose specific outcomes are the progression of a tissue over time, from its formation to its mature structure, and many pathways involving carbohydrate derivatives.

3.4.3. Breed-Specific lncRNAs Associated with Oral Melanoma

The design of our study, which included three distinct breeds predisposed to MM, made it possible to integrate both the coexpression module analysis and the differentially expressed lncRNAs, for each separate breed. Given the low number of control samples for Labrador retrievers, we focused our analysis on the pairwise comparisons between golden retrievers and poodles. Using WGCNA, the analysis of the poodle breed produced a significant correlation for eight modules (six with PCC > 0.8, $p < 2 \times 10^{-12}$ and two PCC < -0.8, $p < 2 \times 10^{-15}$) (Supplementary Materials Figure S3). From these modules, the gene set enrichment analysis showed that the GO terms (biological process) that were most significantly enriched were "regulation of gene expression", "chromatin organization", and "chromatin modification" (orange module, Supplementary Materials Figure S4). Complementary to this analysis, we refined the DESeq2 experimental design, which previously computed the global melanoma effect while controlling for differences due to the breeds, to search for DE lncRNAs only in poodles, and not in golden retrievers (see Methods). Our analysis identified a panel of 11 lncRNAs that were significantly DE only in poodles ($|$lFC$| > 1.5$ and $p_{\text{adj}} < 0.05$), and which belong to WGCNA modules associated to poodles (Supplementary Materials Table S5). For instance, we observed that the most significant DE lncRNA (*RLOC_00005829*) is down-regulated in poodles (lFC $= -5.99$, $p_{\text{adj}} = 8.1 \times 10^{-7}$), while its expression is not significantly altered in golden retrievers ($p_{\text{adj}} = 0.61$) (Supplementary Materials Figure S5). Concordantly, this lncRNA was not considered as being DE ($p_{\text{adj}} = 0.54$) in the first design when the tumor effect was controlled for differences due to breeds. Finally, we mapped the *RLOC_00005829* sequence on the human genome, and showed that it clearly aligned to the *COLCA1* gene (identity = 61.1%), a GENCODE-annotated antisense lncRNA [24] that was already associated with human colorectal cancer by GWAS [33].

For golden retrievers, the coexpression analysis produced significant correlations for four modules (PCC < -0.8, $p < 3 \times 10^{-20}$). Similarly, the DE analysis identified a panel of seven lncRNAs only found to be differentially expressed in golden retriever samples and not in poodles, but these were not identified with WGCNA (Supplementary Materials Table S5).

4. Discussion and Conclusions

Long non-coding RNAs (lncRNAs) are key regulators in many biological processes and they are often dysregulated in cancers [34], including cutaneous melanoma [13]. We investigated lncRNAs of the canine model as being potential cancer markers for mucosal melanomas in humans. Our findings show the existence of a genetic basis and expression variation involving long non-coding RNAs in oral mucosal melanomas in dogs from three breeds (golden retrievers, Labrador retrievers, and poodles), with an increased risk of developing oral mucosal melanomas.

In this study, bioinformatic analyses identified more than 400 dysregulated lncRNAs that discriminated canine oral melanoma tumors from control samples. We further pinpointed one down-regulated (*SOX21-AS*) and two up-regulated (*CASC15* and *ZEB2-AS*) DE lncRNAs (inferred as "onco-lncRNAs") [35,36] in canine oral melanoma, that were significantly conserved with humans, and already associated with human cancers. These results provide a novel resource for candidate biomarkers, for which further in vitro and in vivo experimental validations will be required.

Although we used bulk RNA-Seq to analyze dysregulated lncRNAs in canine oral melanomas, we adopted an enrichment-based computational analysis to control for covariates such as cell-type heterogeneity between samples. Importantly, the xCell program, which was used to compute these enrichments, includes melanoma-related cell-types from the Tirosh et al. single-cell RNA-Seq study [37], such as malignant, immune, and endothelial cells. In melanomas, the distinct subtypes that most likely harbor multiple cell types and high genetic heterogeneity are thought to play a role in the development and progression of tumors. Future directions to explore the distinct genotypic and phenotypic states of the tumors will involve directly performing single-cell RNA sequencing on oral mucosal melanomas. Furthermore, the expression of lncRNAs is highly tissue- and cancer-specific [8,34] and this is particularly relevant for studying cells of the same tumor and/or tissue that exhibit transcriptional

heterogeneity [38]. In our study, we also observed that the tissue specificity, measured from canine normal tissues [6], was significantly higher for DE lncRNAs than for DE mRNAs (*p*-value= 2.8×10^{-9}, Wilcoxon's rank-sum test), reinforcing the attractiveness of lncRNAs as potential biomarkers of oral melanomas.

We also report a weighted gene coexpression network analysis (WGCNA) that constructed 59 modules by an unsupervised analysis of gene expression profiles. The WGCNA method was further used to detect the relationship between the lncRNA expression profiles and the melanoma status. WGCNA has many advantages over other clustering methods, since the analysis uses a 'soft-thresholding procedure' to avoid the selection of an arbitrary cut-off. It also focuses on the association between coexpression modules and clinical features, and the results have robust reliability and biological significance. Genes in the same module are considered to be related with each other by their functions. We identified four coexpression modules that are related to oral melanoma for all breeds studied, and specific DE lncRNAs for the poodle and golden retrievers breeds. Thus, this study led to the identification of biologically relevant modules and hub lncRNAs that could serve as biomarkers for the detection of mucosal melanomas [39].

To give biological meaning to identify lncRNAs, we conducted a gene set enrichment analysis. These analyses showed clear differences in enriched GO (BP) terms between the different modules, which were largely associated with different functions. As a result, modules containing up-regulated genes were found to be mainly enriched in cancer-associated pathways, implicating the replication and segregation of genetic material, and its progression through the mitotic cell cycle phases. The dysregulated lncRNAs of these modules could possibly have a role in cell cycle or cell proliferation. Modules with down-regulated genes were largely involved in carbohydrate metabolic processes. Carbohydrates and glucose can have important effects on the proliferation of tumor cells. It has been reported that most malignant cells are dependent on the availability of glucose in the blood for their energy, and that they are not able to metabolize it, especially in case of mitochondrial dysfunction [40].

Gene expression profiling is actively investigated as a clinical biomarker and diagnostic tools to detect multiple cancer types and distinct stages. However it is challenging to take into account the variability of gene expression, and thus the underlying functions of genes in populations of different ethnic origins [41]. Here, we used the unique features of the dog model, and its diversity and breed structure, to study the expression variations of lncRNAs that are associated with mucosal melanomas between breeds. We have identified lncRNAs that are differentially expressed only in melanomas sampled in poodles, such as the antisense lncRNA *COLCA1*. Therefore, the variation in lncRNA expression identified in dog breeds may help to better characterize the observed disparities and heterogeneity of mucosal melanomas in humans.

Identifying the dysregulation of lncRNA expression in mucosal melanomas provides novel tools and resource that can serve as diagnostic and therapeutic targets. Here, we show by the identification of conserved dog–human lncRNAs, that they can also provide key markers in human mucosal melanomas.

Author Contributions: Conceptualization, T.D. and C.H.; methodology, C.L.B., V.W., C.H. and T.D.; software, C.L.B., V.W., C.H., T.D.; validation, C.L.B., V.W., A.P. (Aline Primot), E.C., A.P. (Anaïs Prouteau), N.B., B.H.; investigation, C.L.B., V.W., A.P. (Aline Primot), E.C., A.P. (Anaïs Prouteau), N.B., B.H.; resources, N.B., E.C. and C.A.; data curation, C.L.B., V.W., A.P. (Aline Primot), E.C., A.P. (Anaïs Prouteau), N.B., B.H.; writing—original draft preparation, T.D., C.H.; writing—review and editing, C.H., C.L.B., V.W., B.H., K.L.-T., C.A., T.D.; supervision, T.D., C.H.; funding acquisition, C.A., K.L.-T.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

References

1. Gillard, M.; Cadieu, E.; de Brito, C.; Abadie, J.; Vergier, B.; Devauchelle, P.; Degorce, F.; Dréano, S.; Primot, A.; Dorso, L.; et al. Naturally occurring melanomas in dogs as models for non-UV pathways of human melanomas. *Pigment Cell Melanoma Res.* **2014**, *27*, 90–102. [CrossRef] [PubMed]

2. Hernandez, B.; Adissu, H.; Wei, B.-R.; Michael, H.; Merlino, G.; Simpson, R. Naturally Occurring Canine Melanoma as a Predictive Comparative Oncology Model for Human Mucosal and Other Triple Wild-Type Melanomas. *Int. J. Mol. Sci.* **2018**, *19*, 394. [CrossRef] [PubMed]

3. Wong, K.; van der Weyden, L.; Schott, C.R.; Foote, A.; Constantino-Casas, F.; Smith, S.; Dobson, J.M.; Murchison, E.P.; Wu, H.; Yeh, I.; et al. Cross-species genomic landscape comparison of human mucosal melanoma with canine oral and equine melanoma. *Nat. Commun.* **2019**, *10*, 353. [CrossRef] [PubMed]

4. Hendricks, W.P.D.; Zismann, V.; Sivaprakasam, K.; Legendre, C.; Poorman, K.; Tembe, W.; Perdigones, N.; Kiefer, J.; Liang, W.; DeLuca, V.; et al. Trent, Somatic inactivating PTPRJ mutations and dysregulated pathways identified in canine malignant melanoma by integrated comparative genomic analysis. *PLoS Genet.* **2018**, *14*, e1007589. [CrossRef] [PubMed]

5. Wucher, V.; Legeai, F.; Hédan, B.; Rizk, G.; Lagoutte, L.; Leeb, T.; Jagannathan, V.; Cadieu, E.; David, A.; Lohi, H.; et al. FEELnc: A tool for long non-coding RNA annotation and its application to the dog transcriptome. *Nucleic Acids Res.* **2017**, *45*, e57. [CrossRef] [PubMed]

6. Le Béguec, C.; Wucher, V.; Lagoutte, L.; Cadieu, E.; Botherel, N.; Hédan, B.; de Brito, C.; Guillory, A.-S.; André, C.; Derrien, T.; et al. Characterisation and functional predictions of canine long non-coding RNAs. *Sci. Rep.* **2018**, *8*, 13444. [CrossRef] [PubMed]

7. Carninci, P. The Transcriptional Landscape of the Mammalian Genome. *Science* **2005**, *309*, 1559–1563. [CrossRef] [PubMed]

8. Derrien, T.; Johnson, R.; Bussotti, G.; Tanzer, A.; Djebali, S.; Tilgner, H.; Guernec, G.; Merkel, A.; Gonzalez, D.; Lagarde, J.; et al. The GENCODE v7 catalogue of human long non-coding RNAs: Analysis of their structure, evolution and expression. *Genome Res.* **2012**, *22*, 1775–1789. [CrossRef]

9. Djebali, S.; Davis, C.A.; Merkel, A.; Dobin, A.; Lassmann, T.; Mortazavi, A.; Tanzer, A.; Lagarde, J.; Lin, W.; Schlesinger, F.; et al. Landscape of transcription in human cells. *Nature* **2012**, *489*, 101–108. [CrossRef]

10. Lanzós, A.; Carlevaro-Fita, J.; Mularoni, L.; Reverter, F.; Palumbo, E.; Guigó, R.; Johnson, R. Discovery of Cancer Driver Long Noncoding RNAs across 1112 Tumour Genomes: New Candidates and Distinguishing Features. *Sci. Rep.* **2017**, *7*, 41544. [CrossRef] [PubMed]

11. Hosono, Y.; Niknafs, Y.S.; Prensner, J.R.; Iyer, M.K.; Dhanasekaran, S.M.; Mehra, R.; Pitchiaya, S.; Tien, J.; Escara-Wilke, J.; Poliakov, A.; et al. Oncogenic Role of THOR, a Conserved Cancer/Testis Long Non-coding RNA. *Cell* **2017**, *171*, 1559–1572. [CrossRef] [PubMed]

12. Wang, Z.; Yang, B.; Zhang, M.; Guo, W.; Wu, Z.; Wang, Y.; Jia, L.; Li, S.; Xie, W.; Yang, D.S. lncRNA epigenetic landscape analysis identifies EPIC1 as an oncogenic lncRNA that interacts with MYC and promotes cell-cycle progression in cancer. *Cancer Cell* **2018**, *33*, 706–720. [CrossRef] [PubMed]

13. Leucci, E.; Coe, E.A.; Marine, J.-C.; Vance, K.W. The emerging role of long non-coding RNAs in cutaneous melanoma. *Pigment Cell Melanoma Res.* **2016**, *29*, 619–626. [CrossRef] [PubMed]

14. Khaitan, D.; Dinger, M.E.; Mazar, J.; Crawford, J.; Smith, M.A.; Mattick, J.S.; Perera, R.J. The melanoma-upregulated long noncoding RNA SPRY4-IT1 modulates apoptosis and invasion. *Cancer Res.* **2011**, *71*, 3852–3862. [CrossRef] [PubMed]

15. Leucci, E.; Vendramin, R.; Spinazzi, M.; Laurette, P.; Fiers, M.; Wouters, J.; Radaelli, E.; Eyckerman, S.; Leonelli, C.; Vanderheyden, K.; et al. Melanoma addiction to the long non-coding RNA SAMMSON. *Nature* **2016**, *531*, 518–522. [CrossRef] [PubMed]

16. Lindblad-Toh, K.; Members, B.S.P.; Wade, C.M.; Mikkelsen, T.S.; Karlsson, E.K.; Jaffe, D.B.; Kamal, M.; Clamp, M.; Chang, J.L.; Kulbokas, E.J.; et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **2005**, *438*, 803–819. [CrossRef] [PubMed]

17. Djebali, S.; Wucher, V.; Foissac, S.; Hitte, C.; Corre, E.; Derrien, T. Bioinformatics Pipeline for Transcriptome Sequencing Analysis. *Methods Mol. Biol.* **2017**, *1486*, 201–219. [CrossRef]

18. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**, *29*, 15–21. [CrossRef]

19. Li, B.; Dewey, C.N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **2011**, *12*, 323. [CrossRef]

20. Anders, S.; Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **2010**, *11*, R106. [CrossRef]

21. Aran, D.; Hu, Z.; Butte, A.J. xCell: Digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **2017**, *18*, 220. [CrossRef] [PubMed]

22. Zhu, A.; Ibrahim, J.G.; Love, M.I. Heavy-tailed prior distributions for sequence count data: Removing the noise and preserving large differences. *Bioinformatics* **2018**, *35*, 2084–2092. [CrossRef]

23. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **2018**, *34*, 3094–3100. [CrossRef] [PubMed]

24. Harrow, J.; Frankish, A.; González, J.M.; Tapanari, E.; Diekhans, M.; Kokocinski, F.; Aken, B.L.; Barrell, D.; Zadissa, A.; Searle, S.; et al. GENCODE: The reference human genome annotation for The ENCODE Project. *Genome Res.* **2012**, *22*, 1760–1774. [CrossRef]

25. Quinlan, A.R.; Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841–842. [CrossRef] [PubMed]

26. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **2008**, *9*, 559. [CrossRef] [PubMed]

27. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles 2005. Available online: www.pnas.orgcgidoi10.1073pnas. 0506580102 (accessed on 15 April 2019).

28. Hayward, N.K.; Wilmott, J.S.; Waddell, N.; Johansson, P.A.; Field, M.A.; Nones, K.; Patch, A.M.; Kakavand, H.; Alexandrov, L.B.; Burke, H.; et al. Whole-genome landscapes of major melanoma subtypes. *Nature* **2017**, *545*, 175–180. [CrossRef]

29. Beltran, M.; Puig, I.; Peña, C.; García, J.M.; Alvarez, A.B.; Peña, R.; Bonilla, F.; de Herreros, A.G. A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. *Genes Dev.* **2008**, *22*, 756–769. [CrossRef]

30. Yang, C.-M.; Wang, T.-H.; Chen, H.-C.; Li, S.-C.; Lee, M.-C.; Liou, H.-H.; Liu, P.-F.; Tseng, Y.-K.; Shiue, Y.-L.; Ger, L.-P.; et al. Aberrant DNA hypermethylation-silenced SOX21-AS1 gene expression and its clinical importance in oral cancer. *Clin. Epigenet.* **2016**, *8*, 129. [CrossRef]

31. Lessard, L.; Liu, M.; Marzese, D.M.; Wang, H.; Chong, K.; Kawas, N.; Donovan, N.C.; Kiyohara, E.; Hsu, S.; Nelson, N.; et al. The CASC15 Long Intergenic Noncoding RNA Locus Is Involved in Melanoma Progression and Phenotype Switching. *J. Investig. Dermatol.* **2015**, *135*, 2464–2474. [CrossRef]

32. Guttman, M.; Amit, I.; Garber, M.; French, C.; Lin, M.F.; Feldser, D.; Huarte, M.; Zuk, O.; Carey, B.W.; Cassady, J.P.; et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **2009**, *458*, 223–227. [CrossRef] [PubMed]

33. Closa, A.; Cordero, D.; Sanz-Pamplona, R.; Solé, X.; Crous-Bou, M.; Paré-Brunet, L.; Berenguer, A.; Guino, E.; Lopez-Doriga, A.; Guardiola, J.; et al. Identification of candidate susceptibility genes for colorectal cancer through eQTL analysis. *Carcinogenesis* **2014**, *35*, 2039–2046. [CrossRef] [PubMed]

34. Iyer, M.K.; Niknafs, Y.S.; Malik, R.; Singhal, U.; Sahu, A.; Hosono, Y.; Barrette, T.R.; Prensner, J.R.; Evans, J.R.; Zhao, S.; et al. The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* **2015**, *47*, 199–208. [CrossRef] [PubMed]

35. Chiu, H.-S.; Somvanshi, S.; Patel, E.; Chen, T.-W.; Singh, V.P.; Zorman, B.; Patil, S.L.; Pan, Y.; Chatterjee, S.S.; Sood, A.K.; et al. Pan-Cancer Analysis of lncRNA Regulation Supports Their Targeting of Cancer Genes in Each Tumor Context. *Cell Rep.* **2018**, *23*, 297–312. [CrossRef] [PubMed]

36. Esposito, R.; Bosch, N.; Lanzós, A.; Polidori, T.; Pulido-Quetglas, C.; Johnson, R. Hacking the Cancer Genome: Profiling Therapeutically Actionable Long Non-coding RNAs Using CRISPR-Cas9 Screening. *Cancer Cell* **2019**, *35*, 545–557. [CrossRef] [PubMed]

37. Tirosh, I.; Izar, B.; Prakadan, S.M.; Wadsworth, M.H.; Treacy, D.; Trombetta, J.J.; Rotem, A.; Rodman, C.; Lian, C.; Murphy, G.; et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **2016**, *325*, 189–196. [CrossRef]

38. Liu, S.J.; Nowakowski, T.J.; Pollen, A.A.; Lui, J.H.; Horlbeck, M.A.; Attenello, F.J.; He, D.; Weissman, J.S.; Kriegstein, A.R.; Diaz, A.A.; et al. Single-cell analysis of long non-coding RNAs in the developing human neocortex. *Genome Biol.* **2006**, *17*, 67. [CrossRef] [PubMed]

39. Qi, P.; Zhou, X.; Du, X. Circulating long non-coding RNAs in cancer: Current status and future perspectives. *Mol. Cancer* **2016**, *15*, 39. [CrossRef] [PubMed]

40. Klement, R.J.; Kämmerer, U. Is there a role for carbohydrate restriction in the treatment and prevention of cancer? *Nutr. Metab.* **2011**, *8*, 75. [CrossRef] [PubMed]

41. Rawlings-Goss, R.A.; Campbell, M.C.; Tishkoff, S.A. Global population-specific variation in miRNA associated with cancer risk and clinical biomarkers. *BMC Med. Genomics* **2014**, *7*, 53. [CrossRef]