Biochimie 140 (2017) 122–132

Contents lists available at ScienceDirect

Biochimie

journal homepage: www.elsevier.com/locate/biochi

Research paper

Genomic landscape of copy number variation and copy neutral loss of heterozygosity events in equine sarcoids reveals increased instability of the sarcoid genome

Klaudia Pawlina-Tyszko [a, *], Artur Gurgul [a], Tomasz Szmatoła [a], Katarzyna Ropka-Molik [a], Ewelina Semik-Gurgul [a], Jolanta Klukowska-Rötzler [b, c], Christoph Koch [d], Kathrin Mählmann [d, e], Monika Bugno-Poniewierska [a]

[a] Laboratory of Genomics, Department of Animal Genomics and Molecular Biology, National Research Institute of Animal Production, Krakowska 1, 32-083, Balice, Poland
[b] Division of Pediatric Hematology/Oncology, Department of Clinical Research, University of Bern, Murtenstrasse 35, 3008, Bern, Switzerland
[c] Department of Emergency Medicine, University Hospital Bern, Inselspital, 3010, Bern, Switzerland
[d] Swiss Institute of Equine Medicine ISME, Faculty of Veterinary Medicine, University of Bern and Agroscope, Länggassstrasse 124c, Postfach 8466, CH-3001, Bern, Switzerland
[e] Equine Clinic: Surgery and Radiology, Department of Veterinary Medicine, Free University of Berlin, Oertzenweg 19b, 14163, Berlin, Germany

ARTICLE INFO

Article history:
Received 12 April 2017
Accepted 20 July 2017
Available online 23 July 2017

Keywords:
CNV
cnLOH
Sarcoid
Horse
Microarrays

ABSTRACT

Although they are the most common neoplasms in equids, sarcoids are not fully characterized at the molecular level. Therefore, the objective of this study was to characterize the landscape of structural rearrangements, such as copy number variation (CNV) and copy neutral loss of heterozygosity (cnLOH), in the genomes of sarcoid tumor cells. This information will not only broaden our understanding of the characteristics of this genome but will also improve the general knowledge of this tumor and the mechanisms involved in its generation. To this end, Equine SNP64K Illumina microarrays were applied along with bioinformatics tools dedicated for signal intensity analysis. The analysis revealed increased instability of the genome of sarcoid cells compared with unaltered skin tissue samples, which was manifested by the prevalence of CNV and cnLOH events. Many of the identified CNVs overlapped with the other research results, but the simultaneously observed variability in the number and sizes of detected aberrations indicated a need for further studies and the development of more reliable bioinformatics algorithms. The functional analysis of genes co-localized with the identified aberrations revealed that these genes are engaged in vital cellular processes. In addition, a number of these genes directly contribute to neoplastic transformation. Furthermore, large numbers of cnLOH events identified in the sarcoids suggested that they may play no less significant roles than CNVs in the carcinogenesis of this tumor. Thus, our results indicate the importance of cnLOH and CNV in equine sarcoid oncogenesis and present a direction of future research.

1. Introduction

In the genomes of cells undergoing neoplastic transformation, alterations at different levels of DNA organization occur. One of these alterations is a type of structural rearrangements called copy number variation (CNV), which comprises deletion or amplification events from 1 kbp to several Mbp [1,2]. CNVs can influence cell function through a variety of mechanisms, such as changes of gene dosage or disruptions of regulatory elements, e.g., enhancers [3–5]. CNVs may also lead to altered tissue specificity of genes [6] or induce further structural rearrangements [7]. As a result, such considerable genome rearrangements may cause recessive variants to reveal or even inactivate genes.

* Corresponding author.
E-mail addresses: klaudia.pawlina@izoo.krakow.pl (K. Pawlina-Tyszko), artur.gurgul@izoo.krakow.pl (A. Gurgul), tomasz.szmatola@izoo.krakow.pl (T. Szmatoła), katarzyna.ropka@izoo.krakow.pl (K. Ropka-Molik), ewelina.semik@izoo.krakow.pl (E. Semik-Gurgul), jolanta.klukowska-roetzler@insel.ch (J. Klukowska-Rötzler), christoph.koch@vetsuisse.unibe.ch (C. Koch), Kathrin.Maehlmann@fu-berlin.de (K. Mählmann), monika.bugno@izoo.krakow.pl (M. Bugno-Poniewierska).

http://dx.doi.org/10.1016/j.biochi.2017.07.006
0300-9084/

Numerous studies investigating CNV identification enabled elucidation of the influence of copy number changes on genetic diversity between individuals, the genetic basis of behavior, complex traits or susceptibility to diseases [8]. Furthermore, cancer genetics research often employs CNV association studies. Approximately 40% of genes engaged in oncogenesis are subjected to copy number changes [9], and almost all neoplasms are characterized by increased numbers of copy number aberrations [10]. Moreover, existing research suggests that not only CNV but also copy neutral loss of heterozygosity (cnLOH) may be of great importance in neoplastic transformation processes. This aberration occurs when a DNA fragment inherited from one parent is lost, and the other is duplicated. Acquired cnLOH is commonly observed in both solid and hematological tumors and constitutes 20–80% of LOH events identified in human tumors [11–14]. In cancer cells, cnLOH is considered a biological equivalent to the second hit in the Knudson hypothesis [15] because it may reveal mutated alleles when non-aberrant alleles are lost.

Studies to characterize genomic aberrations of livestock species are still in the development phase [16]. Similar to human association research in this field, the involvement of CNVs in different phenotypes was reported, such as pea-comb phenotype in chickens, white coat in horses, osteopetrosis, copper toxicosis and intersexuality [16]. However, no such studies have been performed in sarcoids, which are the most commonly observed equine skin tumors. This locally malignant, benign tumor of fibrous tissue [17] is believed to constitute up to 40% of all diagnosed neoplasms [18–20]. Although sarcoids do not have the ability to metastasize, they may exhibit differences in aggressiveness [21–23] and recurrence if not treated properly [24]. Complete, constant remission is sporadically observed [25], whereas unsuccessful, prolonged treatment may result in the owner's decision to euthanize the equid suffering from sarcoid disease. Bovine papillomavirus (BPV) infection is generally accepted to be a crucial etiological factor for the development of sarcoids; however, the mechanism of BPV's oncogenic activity has not been deciphered to date [17,26,27].

Given the significant importance of equids in livestock production, recreation, and equine-assisted therapy and as working animals in poorly developed countries, it is essential to broaden our knowledge about equine genome organization and aberrations, which may underlie disease processes, such as sarcoid disease. Given that CNV, as well as cnLOH events, are commonly identified in a variety of cancer cells and influence the function of oncogenes and tumor suppressor genes, we hypothesized that the sarcoid cell genome can also be characterized by genome aberrations of potential functional importance. To address this issue, genotyping using SNP microarrays followed by signal intensity analysis was conducted. The results provide insight into the organization of the sarcoid cell genome and disruptions caused by neoplastic transformation that will facilitate our understanding of underlying aberrant mechanisms.

2. Materials and methods

2.1. Samples and SNP genotyping

Research material collected at the ISME Equine Clinic in Bern, Switzerland comprised sarcoid tissue samples and normal skin biopsy samples obtained from a tumor distant site (at the neck) of 16 horses. Samples were obtained from sarcoid-affected animals treated at the clinic. The use of excised tissue samples was approved by the Animal Experimentation Ethics Committee of the Canton of Berne, Switzerland (BE 30/11; 11 April 2011). After excision, the samples were stored in −80 °C. DNA was isolated using DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer protocol and stored in −20 °C until use. All DNA samples were genotyped using the Equine SNP64 BeadChip assay (Illumina, San Diego, CA, USA; NEOGEN, Lansing, MI, USA) following the standard Infinium Ultra protocol. Obtained data were assessed for quality parameters, such as call rate and log R ratio standard deviation. Samples with call rates less than 0.975 and samples with log R ratio deviation greater than 0.35 were excluded due to poor quality for CNV detection [28,29].

2.2. CNV and cnLOH calling

Data on genotypes and intensity parameters (LRR – log R Ratio and BAF – B allele frequency) were exported from GenomeStudio and subjected to analysis using three algorithms dedicated to CNV and cnLOH calling, namely, cnvPartition v3.2.0 (Illumina), PennCNV v2011Jun16 [30] and OncoSNP v2.1 (Beta) [31]. All analyses were performed using EquCab2.0 equine genome assembly. Control tissue data were analyzed using cnvPartition and PennCNV, and obtained results were subsequently analyzed using PLINK software v1.07 (http://pngu.mgh.harvard.edu/~purcell/plink/) to identify common aberrations for these algorithms, so-called consensus regions. Aberrations in sarcoid tissue were identified using OncoSNP software, which is designed for the analysis of cancer data on the basis of both control and tumor input files.

2.2.1. cnvPartition

The algorithm uses BAF and LRR parameters to identify systematic deviation of neighboring markers, which indicates the occurrence of different copy numbers. The BAF parameter indicates the relative quantity of one allele compared with the other, whereas the LRR parameter is defined as the log ratio of observed to expected SNP intensities [32].

The results were clustered with respect to the standard cluster positions according to Illumina Inc. The confidence threshold for CNVs was set to 35, whereas the minimal number of probes was set to 3. The sex chromosomes were excluded from the analysis, given the lack of information regarding the sex of the animals. Furthermore, GC correction implemented in GenomeStudio software was applied to exclude the influence of so called 'GC genomic waves', which can significantly reduce the ability of algorithms to detect CNVs [33].

2.2.2. PennCNV

To confirm the obtained results, we employed another CNV calling algorithm, namely, PennCNV, which applies hidden Markov models (HMM). The software includes a few sources of information, such as signal intensity and allele frequency of SNPs and the distance between neighboring SNPs.

The parameters of analysis were the same as above, including LRR standard deviation < 0.35 and the minimum of three probes within a CNV region. All files were prepared and analyzed according to instructions (http://www.openbioinformatics.org/penncnv/), including GC wave correction. The sex chromosomes were excluded from the analysis for the abovementioned reasons.

2.2.3. PLINK software

To increase the likelihood of predictions, the control tissue data obtained with the use of cnvPartition and PennCNV algorithms were analyzed using PLINK software. To this end, first, the map of the detected aberrations was generated using the command –cnv-make-map. Then, common regions were identified applying the –segment-group command.

2.2.4. OncoSNP

To identify aberrations in the sarcoid tissue, we used the cancer-