J. Comp. Path. 2015, Vol. 152, 119–130

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

www.elsevier.com/locate/jcpa

NEOPLASTIC DISEASE

Reconstruction of Canine Diffuse Large B-cell Lymphoma Gene Regulatory Network: Detection of Functional Modules and Hub Genes

M. Zamani-Ahmadmahmudi*,†, A. Najafi‡ and S. M. Nassiri§

* *Department of Clinical Science, Faculty of Veterinary Medicine, Shahrekord University, Shahrekord,* † *Department of Clinical Science, Faculty of Veterinary Medicine, Shahid Bahonar University of Kerman, Kerman,* ‡ *Molecular Biology Research Center, Baqiyatallah University of Medical Sciences and* § *Department of Clinical Pathology, Faculty of Veterinary Medicine, University of Tehran, Tehran, Iran*

Summary

Lymphoma is one of the most common malignancies in dogs. Canine lymphoma is similar to human non-Hodgkin's lymphoma (NHL) with shared clinical presentation and histopathological features. This study reports the construction of a comprehensive gene regulatory network (GRN) for canine diffuse large B-cell lymphoma (DLBCL), the most common type of canine lymphoma, and performs analysis for detection of major functional modules and hub genes (the most important genes in a GRN). The canine DLBCL GRN was reconstructed from gene expression data (NCBI GEO dataset: GSE30881) using the STRING and MiMI interaction databases. Reconstructed GRNs were then assessed, using various bioinformatics programmes, in order to analyze network topology and identify major pathways and hub genes. The resultant network from both interaction databases had a logically scale-free pattern. Gene ontology (GO) analysis revealed cell activation, cell cycle phase, immune effector process, immune system development, immune system process, integrin-mediated signalling pathway, intracellular protein kinase cascade, intracellular signal transduction, leucocyte activation and differentiation, lymphocyte activation and differentiation as major GO terms in the biological processes of the networks. Moreover, bioinformatics analysis showed *E2F1*, *E2F4*, *PTEN*, *CDKN1A*, *PCNA*, *DKC1*, *MNAT1*, *NDUFB4*, *ATP5J*, *PRKDC*, *BRCA1*, *MYCN*, *RFC4* and *POLA1* as the most important hub genes. The phosphatidyl inositol signalling system, P53 signalling pathway, Rac CycD pathway, G1/S checkpoint, chemokine signalling pathway and telomere maintenance were the main signalling pathways in which the protein products of the hub genes are involved.

Keywords: dog; gene regulatory network; hub genes; lymphoma

Introduction

Canine lymphoma is one of the most common malignancies in dogs and most commonly has multicentric anatomical distribution (MacEwen, 1990). Canine lymphoma shares clinical and histopathological features with human non-Hodgkin's lymphoma (NHL) (Vail and MacEwen, 2000; McCaw *et al.*, 2007). Histopathological classification of canine lymphoma has been performed using three classification schemes the updated Kiel scheme, the Revised European and American Lymphoma (REAL) scheme and the World Health Organisation (WHO) scheme. These classifications consider epidemiological, clinical, clinicomorphological, genetic and phenotypic parameters. The most common form of canine lymphoma is diffuse large B-cell lymphoma (DLBCL) (Ponce *et al.*, 2010).

Cancer develops through dysregulation of normal cellular processes including apoptosis, cell mitosis, cell cycling, cell differentiation and DNA repair.

Correspondence to: M. Zamani-Ahmadmahmudi (e-mail: zamani_2012@alumni.ut.ac.ir).

0021-9975/$ - see front matter
http://dx.doi.org/10.1016/j.jcpa.2014.11.008

Neoplastic transformation results in activation of oncogenes, deactivation of tumour suppressor genes and/or genome instability (Croce, 2008; Emmert-Streib *et al.*, 2014) and various cell processes are involved in cancer (Hanahan and Weinberg, 2000). Cancer involves intricate interactions between different signalling components such as genes, proteins and metabolites, and as such interactions do not follow a simple chain-like pattern, it has proven challenging to understand the molecular interactions underlying oncogenesis (Emmert-Streib *et al.*, 2014).

One of the most useful approaches in providing a global overview of cancer pathways has been reconstruction of 'gene regulatory networks' (GRNs) using various algorithms or interaction databases (Blais and Dynlacht, 2005; Bansal *et al.*, 2007; Emmert-Streib *et al.*, 2014). GRNs can yield valuable information about major gene ontology modules and are able to identify major genes ('hub genes') involved in development of a specific tumour (Basso *et al.*, 2005, 2010; Emmert-Streib *et al.*, 2014). Through reconstruction and analysis of a breast cancer GRN, Emmert-Streib *et al.* (2014) characterized the major cellular processes contributing to this form of neoplasia, including cell cycling, cell adhesion, translation, organelle fission, the immune response and mitosis. Using a reconstruction of a transcriptional network, Agnelli *et al.* (2011) identified the most critical genes associated with poor prognosis in patients with multiple myeloma.

These methods have not yet been applied widely to the investigation of canine cancer; however, microarray data exist for canine mammary gland tumours, canine lymphoma and canine haemangiosarcoma (http://www.ncbi.nlm.nih.gov/gds/?term=canine%20lymphoma). The aim of the current study was to reconstruct a canine DLBCL GRN and evaluate it using different analytical methods. Specifically, the study defined major functional modules (GO), hub genes (highly connected nodes or the most important genes in a GRN) and the functional biological processes associated with the identified hub genes.

Materials and Methods

Data Collection and Primary Processing

Gene expression data were obtained from the NCBI Gene Expression Omnibus GEO dataset (GSE30881, platform: GPL3738) (Mudaliar *et al.*, 2013). Data included the gene expression profile of 23 canine DLBCLs and normal lymph nodes from 10 healthy dogs. Data were downloaded in the CEL file format and converted to expression values by the Affy package (Gautier *et al.*, 2004) in *R* program, version 3.0.2 (http://www.r-project.org/). As part of this process, data were transformed logarithmically. Then, the data (probe sets) were entered into the geWorkbench 2.5.1 package (Floratos *et al.*, 2010) and non-useful data were filtered based on two criteria: data without Enterz identification were omitted and markers with multiple probset identifications were filtered, while probsets with the highest mean expression value were retained. Upregulation or downregulation of the genes ($P < 0.01$) was determined by use of a *t* test provided in the geWorkbench package (correction method: just alpha, group variance: equal).

k-means Clustering and Principle Component Analysis

Genes were clustered through two algorithms including k-means clustering and principle component analysis (PCA). Using k-means clustering, data were partitioned into specified numbers of groupings (k) in which each observation belonged to the cluster with the nearest mean. PCA reduced dimensionality through generating few linear combinations of all data that are named 'principle components'. Principle components can extract clusters from the original data. Both analyses were performed in MATLAB 7.8.0 (R2009a) (MathWorks, Natick, Massachusetts, USA).

Network Reconstruction and Analysis

The GRN was constructed using two major interaction databases, including MiMI (Michigan Molecular Interactions) (Tarcea *et al.*, 2009) and Search Tool for the Retrieval of Interacting Genes/Proteins (STRING; Franceschini *et al.*, 2013). Briefly, the list of genes was submitted online or through a related plug-in within Cytoscape software (Cline *et al.*, 2007) and then interactions between genes were retrieved and displayed as an interactions network. MiMI collects real interaction data from interaction databases such as the Swiss Protein Database (SwissProt), the Human Protein Reference Database (HPRD), the Biomolecular Interaction Network Database (BIND), the Database of Interacting Proteins (DIP), Reference Sequence (RefSeq) and the International Protein Index (IPI). Interaction data from STRING originate from four major sources including high-throughput experiments, genomic context, co-expression and previous knowledge. Constructed networks were imported to Cytoscape and analyzed using various plug-ins for different aspects of network features.

The networks obtained from these databases were analyzed by NetworkAnalyzer plug-in (Assenov *et al.*, 2008) in order to determine topological parameters and centrality measures, including the network