Efficacy of ARACNE algorithm for inferring canine B-cell lymphoma gene regulatory network (GRN)

Article *in* Comparative Clinical Pathology · October 2016

DOI: 10.1007/s00580-016-2353-7

3 authors, including:

Ali Najafi
BUMS
111 PUBLICATIONS **637** CITATIONS
SEE PROFILE

Mohamad Zamani-Ahmadmahmudi
Shahid Bahonar University of Kerman
32 PUBLICATIONS **148** CITATIONS
SEE PROFILE

Comp Clin Pathol (2017) 26:121–125
DOI 10.1007/s00580-016-2353-7

ORIGINAL ARTICLE

Efficacy of *ARACNE* algorithm for inferring canine B-cell lymphoma gene regulatory network (GRN)

Arezoo Sharafi[1] · **Ali Najafi**[2] · **Mohamad Zamani-Ahmadmahmudi**[1]

Received: 7 August 2016 / Accepted: 3 October 2016 / Published online: 11 October 2016

Abstract Lymphoma is the most frequent hematopoietic cancer in dogs. Canine B-cell lymphoma has been proposed as an ideal model of human non-Hodgkin's lymphoma (NHL). Critical genes playing important roles in the cancer progression can be detected using the reconstruction and analysis of gene regulatory network (GRN). GRNs are inferred using various computational algorithms, where *ARACNE* is on the most important and efficient algorithms. Here, we evaluated the efficacy of *ARACNE* to reconstruct canine B-cell lymphoma GRN via different computational analyses. Hence, the gene expression profile of GSE43664 was downloaded from GEO database and differentially expressed genes were extracted using statistical analysis. Then, significant genes were subjected to reconstruct GRN using *ARACNE* algorithm. Our findings indicated that *ARACNE* inferred a logic biological network with 387 nodes (genes) and 845 edges (interactions). The inferred network followed a biological scale-free pattern, because many nodes had low numbers of interactions and a few nodes were highly connected. Additionally, node degree distribution showed a decreasing linear pattern. Although the network had 80 connected components, most nodes (71.5 %) contributed in a sub-network implying a strong biological network.

Keywords Gene regulatory network (GRN) · Canine B-cell lymphoma · *ARACNE* · Cancer

✉ Mohamad Zamani-Ahmadmahmudi
zamani_2012@alumni.ut.ac.ir

[1] Department of Clinical Science, Faculty of Veterinary Medicine, Shahid Bahonar University of Kerman, P. O Box: 76169133, Kerman, Iran

[2] Molecular Biology Research Center, Baqiyatallah University of Medical Sciences, Tehran, Iran

Introduction

Lymphoma is the most frequent hematopoietic cancer in dogs with the annual incidence of 13 to 24 cases per 100,000 dogs (MacEwen 1990; Marconato et al. 2013). Environmental factors and genetic susceptibility were considered as possible causes of the canine lymphoma. Hypercalcemia is the most common paraneoplastic symptom in canine lymphoma (Zandvliet 2016). Additionally, previous investigations proposed canine B-cell lymphoma as an ideal model of human non-Hodgkin's lymphoma (NHL), where both entities contain similar pathologic and molecular characteristics (McCaw et al. 2007; Richards et al. 2013; Zamani-Ahmadmahmudi et al. 2015; Zamani-Ahmadmahmudi et al. 2016). For example, genomic instability and clinical-pathologic and histologic features were found to be similar in canine B-cell lymphoma and human NHL. Furthermore, some recent cancer therapy agents including ABT526, GS-9219, and I-kappa kinase inhibitors could efficiently provide promising results in dogs with NHL (Marconato et al. 2013).

In the recent years, analysis of the gene expression data and network biology has been extensively tested to derive prognostic gene signatures in various cancers. The findings revealed that extracted gene signatures could robustly categorize tumor subtypes and predict outcome in patients with cancer (Alizadeh et al. 2000; Rosenwald et al. 2002; Lossos et al. 2004). One of the most important methods to detect critical (hub) genes in the biological pathways playing significant roles in cancer progression is the reconstruction of gene regulatory network (GRN) using the various computational algorithms (Bansal et al. 2007; Agnelli et al. 2011; de Matos et al. 2013; Emmert-Streib et al. 2014). Then, via different packages, the reconstructed GRNs will be explored to detect hub genes. Among different computational algorithms, *ARACNE*

(Algorithm for the Reconstruction of Accurate Cellular Networks) as a member of the family of information-theoretic approaches is routinely used to infer GRNs (Butte and Kohane 2000; Margolin et al. 2006; Bansal et al. 2007). Via an information-theoretic structure, *ARACNE* infers direct relationships between target genes and transcriptional regulator proteins. This method can detect co-expression interactions in a gene expression dataset (microarray data and RNA-Seq data) with low inaccuracy rate. This approach facilitates clarifying functional mechanisms involving in various cellular processes (Margolin et al. 2006; Lachmann et al. 2016). However, this approach was majorly used in the cancer biology studies (Agnelli et al. 2011; Liang et al. 2012; Bae et al. 2013; de Matos et al. 2013). de Matos et al. (2013) that have inferred human B-cell lymphoma GRN using three mutual information-based GRN inference methods: *WGCNA, C3Net*, *BC3Net*, and *ARACNE*. Additionally, using a reconstruction of a transcriptional network by *ARACNE* approach, Agnelli et al. (2011) identified the most critical genes associated with poor prognosis in patients with multiple myeloma. There is rare information about GRNs of the different tumors in the veterinary pathology. Hence, in this study, we aimed to investigate *ARACNE* efficacy to infer GRN in canine B-cell lymphoma as one of the most important cancers in the comparative oncology.

Materials and methods

Microarray data

We downloaded gene expression data GSE43664 (platform: Affymetrix Canine Genome 2.0 Array), which has been provided and deposited by Richards et al. (2013) in GEO database. This dataset contains expression data of more than 43×10^3 probsets in 58 dog samples with B-cell lymphoma (each gene may have one or more probsets). Samples majorly comprised diffuse large B-cell lymphomas (DLBCL), histopathologically. More details about studied cases were provided in the GEO website (http://www.ncbi.nlm.nih.gov/geo). The raw dataset was downloaded at CEL file format and converted to the expression value using *affy* package (Gautier et al. 2004) in R program (http://www.r-project.org/). Then, dataset file was imported into *geWorkbench* (Floratos et al. 2010) software and then non-useful probsets detected and deleted. In this process, probsets without Entrez ID were deleted and in genes with the multiple probsets; genes with the highest expression variation were remained. Differentially expressed genes (DEGs) (three or more samples varying by at least fourfold from the median) were detected using method proposed by Richards et al. (2013) in *Cluster 3.0* (Eisen et al. 1998).

Inferring of gene regulatory network using *ARACNE* algorithm

GRN was reconstructed using *ARACNE* algorithm (Margolin et al. 2006) available as a *geWorkbench* plug-in. In this method, indirect interactions inferred by co-expression methods were removed by means of an information-theoretic approach (Margolin et al. 2006). To this purpose, different parameters were set as the following: mode: complete, algorithm: adaptive partitioning, threshold type: p value = 0.01 correct by no. of markers, and DPI tolerance: apply = 0.15.

Assessment of the reconstructed network

The reconstructed network was imported to the *Cytoscape* (Shannon et al. 2003) software package for subsequent analyses. The network was analyzed based on the topological parameters (i.e., number of nodes, diameter, radius, centralization, density, heterogeneity, number of connected components, number of the shortest paths, characteristic path length, and average number of neighbors) and central parameters (i.e., node degree distribution and neighborhood connectivity distribution) using *NetworkAnalyzer* plug-in. In this analysis, diameter is the highest distance between two nodes, where the distance is the minimum number of edges that connected

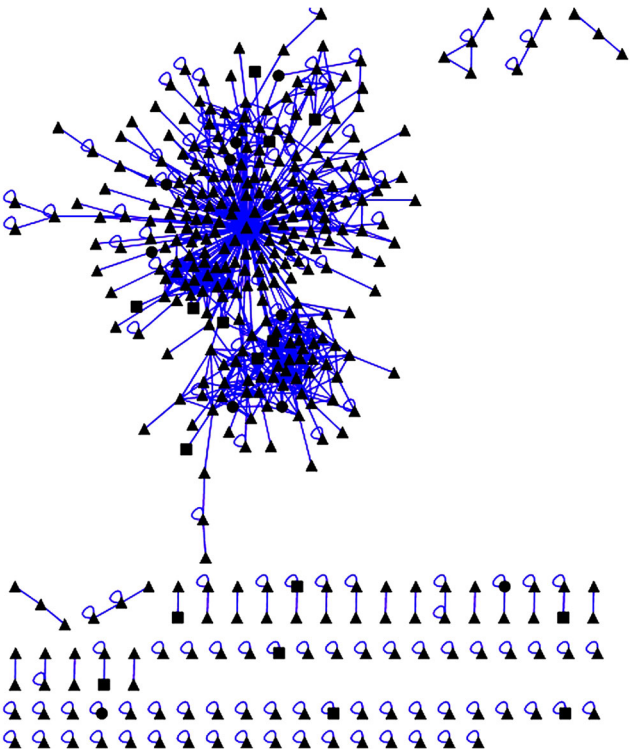

Fig. 1 Topological representation of the reconstructed canine B-cell lymphoma GRN using ARACNE algorithm

Table 1 Simple topological parameters of GRNs resulted from *ARACNE* algorithm

Parameter	Value
Number of nodes	387
Network diameter	8
Network radius	1
Network centralization	0.36
Network density	0.009
Network heterogeneity	2.25
Number of connected components	80
Number of the shortest paths	76,528 (51 %)
Characteristic path length	3.06
Average number of neighbors	3.57

two nodes, radius is the smallest distance between two nodes, and centralization indicates how the network topology resembles a star structure. In a value between 0 and 1, density indicates how densely the network is populated with edges, heterogeneity indicates the affinity of a network to contain hub markers, number of connected components is the number of sub-networks that constitute a network, number of the shortest paths is the minimum numbers of edges that form a path, and characteristics path length is the average shortest path length.

Results and discussion

Our analysis revealed that 1108 out of 43,035 probsets (genes) were differentially expressed in the investigated samples. Expression data of these genes were subjected to reconstruct GRN. Then, GRN was inferred using *ARACNE* algorithm. Topological representation of the reconstructed GRN was presented in Fig. 1. The network contained 387 nodes (genes) and 845 edges (interactions) (Fig. 1 and Table 1).

The inferred network showed a biological scale-free pattern (Barabási and Oltvai 2004), as many nodes had low numbers of interactions and a few nodes were highly connected (Fig. 1). This feature was quantitatively confirmed, where node degree distribution presented a linear descending pattern as shown in Fig. 2a. The neighborhood connectivity distribution was also presented in Fig. 2b. The neighborhood connectivity of a node *n* is defined as the average connectivity of all neighbors of *n* (Maslov and Sneppen 2002). The neighborhood connectivity distribution of our network was in a decreasing way indicating edges between low connected and highly connected nodes prevailed in the network (Maslov and Sneppen 2002).

Lower number of connected component implies a stronger network (Shannon et al. 2003) (http://med.bioinf.mpi-inf.mpg.de/netanalyzer/index.php). Although the inferred GRN included 80 connected components, our network showed a strong connectivity because 277 out of 387 nodes (71.5 %) participated in only one sub-network (Fig. 1 and Table 1). Additionally, on average, each node in the network had 3.5 neighbors. The network density, as a normalized version of average number of neighbors, was calculated 0.009 indicating relatively loose network. The network centralization, indicating how the network topology resembles a star structure (Diestel 2006), was calculated 0.366. It seems that higher number of connected components prevented the network to have a centralization value close to 1. Furthermore, the network clustering coefficient was 0.164 (normal range 0–1) (Table 1). Because nodes with less than two neighbors (number of connected component = 80) having a clustering coefficient of 0 (Watts and Strogatz 1998) were prevalent in our network (Fig. 1), the network clustering coefficient was close to 0.

In this study, we tried to infer GRN of the canine B-cell lymphoma. Based on topological pattern and subsequent analyses, a logic biological network was reconstructed implying acceptable efficacy of the *ARACNE* algorithm.

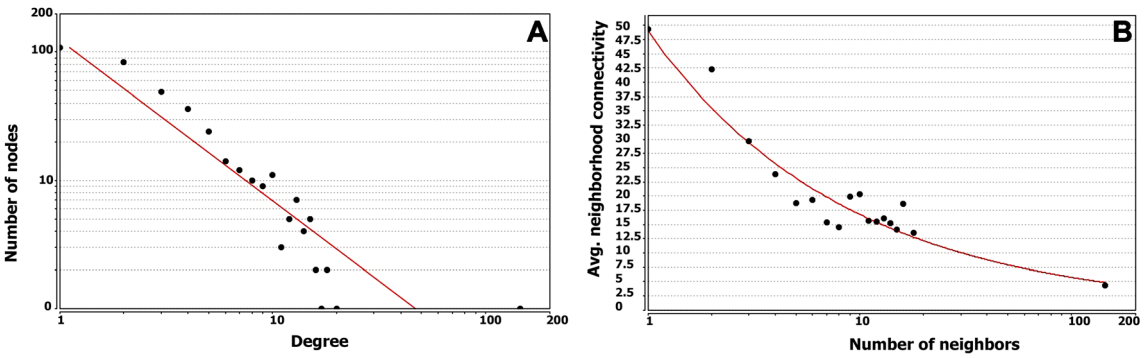

Fig. 2 Node degree distribution (**a**) and neighborhood connectivity distribution (**b**) of the inferred GRN. Both distributions followed a descending linear pattern. In **a**, Y and X axes presented number of nodes and number of neighbors, respectively. Furthermore, in **b**, Y and X axes presented the average of the neighborhood connectivities and number of neighbors, respectively. *Red oblique line* shows fitted power law

Using "fake" gene expression data generated by a computer model of gene regulation, Bansal et al. (2007) have tested performance of the various GRN algorithms. Their results revealed that *ARACNE* could efficiently predict GRNs of 10, 100, and 1000 genes with the acceptable sensitivity and positive predictive value (PPV). The previous investigation showed that *ARACNE* could be reliably used to reconstruct networks in the steady-state data and data with few experiments, as compared with the number of genes (similar to our samples) (Bansal et al. 2007). However, this algorithm cannot perform well for short time series data. Additionally, it is more reliable to run *ARACNE* on minimum 100 samples to providing more realistic networks (Margolin et al. 2006).

Our study could be a start point to employ new computational/bioinformatics approaches exploring animal cancers in more advance ways. The previous investigations have majorly focused on experimental methods/approaches to study cancer biology in animal tumors (Kiupel et al. 1999; McCaw et al. 2007; Mudaliar et al. 2013; Richards et al. 2013) and bioinformatics analyses have had minor role in the study workflows. At current study, using *ARACNE*, we could infer a logic biological GRN for canine B-cell lymphoma. In the next step, the future studies should identify important (hub) genes and critical sub-networks (modules) of the inferred network. These hub genes and modules are playing important roles in various cancer mechanisms (Bansal et al. 2007; Agnelli et al. 2011; Zamani-Ahmadmahmudi et al. 2015). Additionally, the future studies could evaluate efficacy of the ARACNE comparing to other developed algorithms to infer GRN in canine B-cell lymphoma.

To the best of our knowledge, gene expression databases such as GEO and ArrayExpress contain rare and incomplete expression datasets on canine cancers comparing to the human counterparts. Using datasets with the larger and complete samples, future studies will provide more reliable and stronger gene regulatory networks of the various cancers in dog as an ideal animal model for human oncology investigations.

Acknowledgments

The authors wish to thank Mr. Pedram Amouzadeh, who assisted in the proofreading of the manuscript.

Conflict of interest statement The authors declare that they have no conflict of interest.

Ethical approval All applicable international, national, and/or institutional guidelines for the care and use of animals were followed.

References

Agnelli L, Forcato M, Ferrari F, Tuana G, Todoerti K, Walker BA, Morgan GJ, Lombardi L, Bicciato S, Neri A (2011) The reconstruction of transcriptional networks reveals critical genes with implications for clinical outcome of multiple myeloma. Clin Cancer Res 17: 7402–7412

Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO, Staudt LM (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 403:503–511

Bae T, Rho K, Choi JW, Horimoto K, Kim W, Kim S (2013) Identification of upstream regulators for prognostic expression signature genes in colorectal cancer. BMC Syst Biol 7:1–10

Bansal M, Belcastro V, Ambesi-Impiombato A, di Bernardo D (2007) How to infer gene networks from expression profiles. Mol Syst Biol 3

Barabási A-L, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nat Rev Genet 5:101–113

Butte AJ, Kohane IS (2000) Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. Pac Symp Biocomput:418–429

de Matos SR, Dehmer M, Emmert-Streib F (2013) B-cell lymphoma gene regulatory networks: biological consistency among inference methods. Front Genet 4:281

Diestel R (2006) Graph theory. Springer, New York

Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95:14863–14868

Emmert-Streib F, de Matos SR, Mullan P, Haibe-Kains B, Dehmer M (2014) The gene regulatory network for breast cancer: integrated regulatory landscape of cancer hallmarks. Front Genet 5:15

Floratos A, Smith K, Ji Z, Watkinson J, Califano A (2010) geWorkbench: an open source platform for integrative genomics. Bioinformatics 26:1779–1780

Gautier L, Cope L, Bolstad BM, Irizarry RA (2004) Affy-analysis of Affymetrix Gene Chip data at the probe level. Bioinformatics 20: 307–315

Kiupel M, Teske E, Bostock D (1999) Prognostic factors for treated canine malignant lymphoma. Vet Pathol 36:292–300

Lachmann A, Giorgi FM, Lopez G, Califano A (2016) ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. Bioinformatics 32:2233–2235

Liang Y, Wu H, Lei R, Chong RA, Wei Y, Lu X, Tagkopoulos I, Kung S-Y, Yang Q, Hu G, Kang Y (2012) Transcriptional network analysis identifies BACH1 as a master regulator of breast cancer bone metastasis. J Biol Chem 287:33533–33544

Lossos IS, Czerwinski DK, Alizadeh AA, Wechser MA, Tibshirani R, Botstein D, Levy R (2004) Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. N Engl J Med 350:1828–1837

MacEwen EG (1990) Spontaneous tumors in dogs and cats: models for the study of cancer biology and treatment. Cancer Metastasis Rev 9: 125–136

Marconato L, Gelain ME, Comazzi S (2013) The dog as a possible animal model for human non-Hodgkin lymphoma: a review. Hematol Oncol 31:1–9

Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, Califano A (2006) ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics 7(Suppl 1):S7

Maslov S, Sneppen K (2002) Specificity and stability in topology of protein networks. Science 296:910–913

McCaw DL, Chan AS, Stegner AL, Mooney B, Bryan JN, Turnquist SE, Henry CJ, Alexander H, Alexander S (2007) Proteomics of canine lymphoma identifies potential cancer-specific protein markers. Clin Cancer Res 13:2496–2503

Mudaliar MAV, Haggart RD, Miele G, Sellar G, Tan KAL, Goodlad JR, Milne E, Vail DM, Kurzman I, Crowther D, Argyle DJ (2013) Comparative gene expression profiling identifies common molecular signatures of NF-κB activation in canine and human diffuse large B cell lymphoma (DLBCL). PLoS One 8:e72591

Richards KL, Motsinger-Reif AA, Chen H-W, Fedoriw Y, Fan C, Nielsen DM, Small GW, Thomas R, Smith C, Dave SS, Perou CM, Breen M, Borst LB, Suter SE (2013) Gene profiling of canine B-cell lymphoma reveals germinal center and postgerminal center subtypes with different survival times, modeling human DLBCL. Cancer Res 73:5029–5039

Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, Gascoyne RD, Muller-Hermelink HK, Smeland EB, Giltnane JM, Hurt EM, Zhao H, Averett L, Yang L, Wilson WH, Jaffe ES, Simon R, Klausner RD, Powell J, Duffey PL, Longo DL, Greiner TC, Weisenburger DD, Sanger WG, Dave BJ, Lynch JC, Vose J, Armitage JO, Montserrat E, López-Guillermo A, Grogan TM, Miller TP, LeBlanc M, Ott G, Kvaloy S, Delabie J, Holte H, Krajci P, Stokke T, Staudt LM, Lymphoma/Leukemia Molecular Profiling Project (2002) The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. N Engl J Med 346:1937–1947

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504

Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393:440–442

Zamani-Ahmadmahmudi M, Najafi A, Nassiri SM (2015) Reconstruction of canine diffuse large B-cell lymphoma gene regulatory network: detection of functional modules and hub genes. J Comp Pathol 152:119–130

Zamani-Ahmadmahmudi M, Najafi A, Nassiri SM (2016) Detection of critical genes associated with overall survival (OS) and progression-free survival (PFS) in reconstructed canine B-cell lymphoma gene regulatory network (GRN). Cancer Investig 34:70–79

Zandvliet M (2016) Canine lymphoma: a review. Vet Q 36:76–104