CanIsoNet: A Database to Study the Functional Impact of Isoform Switching Events in Cancer

Tülay Karakulak[1,2,3], Damian Szklarczyk[1,3], Holger Moch[2,4], Christian von Mering[1,3], Abdullah Kahraman[1,2,3,*]

[1] Institute of Molecular Life Sciences, University of Zurich, Zurich, Switzerland [2]Department of Pathology and Molecular Pathology, University Hospital Zurich, Zurich, Switzerland [3]Swiss Institute of Bioinformatics, Lausanne, Switzerland [4]Faculty of Medicine, University of Zurich, Zurich, Switzerland

*corresponding author

e-mail: Abdullah.Kahraman@usz.ch

Abstract

Motivation: Alternative splicing, as an essential regulatory mechanism in normal mammalian cells, is frequently disturbed in cancer. Switches in the expression of alternative isoforms can alter protein interaction networks of associated genes giving rise to cancer progression and metastases. We have recently analyzed the pathogenic impact of switching events in 1209 cancer samples covering 24 different cancer types. Here, we are presenting CanIsoNet (Cancer Isoform specific interaction Network), a database to view, browse and search these isoform switching events. CanIsoNet is the first webserver that incorporates isoform expression data with STRING interaction networks and COSMIC annotations to predict the pathogenic impact of isoform switching events in various cancer types.

Results: Data in CanIsoNet can be browsed by cancer types or searched by genes or isoforms in annotation rich data tables. Various annotations for 11,041 isoforms and 31,748 unique isoform switching events are provided across 24 cancer types, including proximity information to COSMIC cancer census genes, network density data for each cancer-specific isoform, PFAM domain IDs of disrupted interactions, domain structure visualization of transcripts and expression data of switched isoforms for each sample.

Availability: CanIsoNet is freely available at https://caniso.net under a Creative Common License. The source codes can be found at https://github.com/KarakulakTulay/CanIsoNet_Web

1 Introduction

Alternative splicing is an essential mechanism to regulate the generation of various mature mRNA transcripts from a single gene[1]. Dysregulation of this mechanism can lead to the overexpression of alternative isoforms and downregulation of the canonical isoform causing an isoform switching event in cancer (Vitting-Seerup and Sandelin, 2017; Kahraman et al., 2020) . Depending on the length and composition of the alternative isoform, the switch event can lead to the loss or gain of protein-protein interaction domains of the affected gene. They result in protein interaction network disruption which can have severe functional impact (Climente-González et al., 2017). The IsoformSwitchAnalyzeR software has been implemented to probe isoform switch events in RNA-seq data and report the functional loss or gain of protein domains (Vitting-Seerup and Sandelin, 2019). More recently the DIGGER database was introduced as a tool to study the interaction network of protein isoforms and protein domains at an isoform and exon level (Louadi et al., 2021) . However, there is no current database or webserver that provides cancer-specific isoform switching data for diverse cancer types. Here, we describe the CANcer-specific ISOform interaction NETwork (CanIsoNet) database that merges cancer specific isoform data with STRING (Szklarczyk et al., 2015) protein interaction networks to visualize and identify the functional impact of isoform switching events across 24 cancer types.

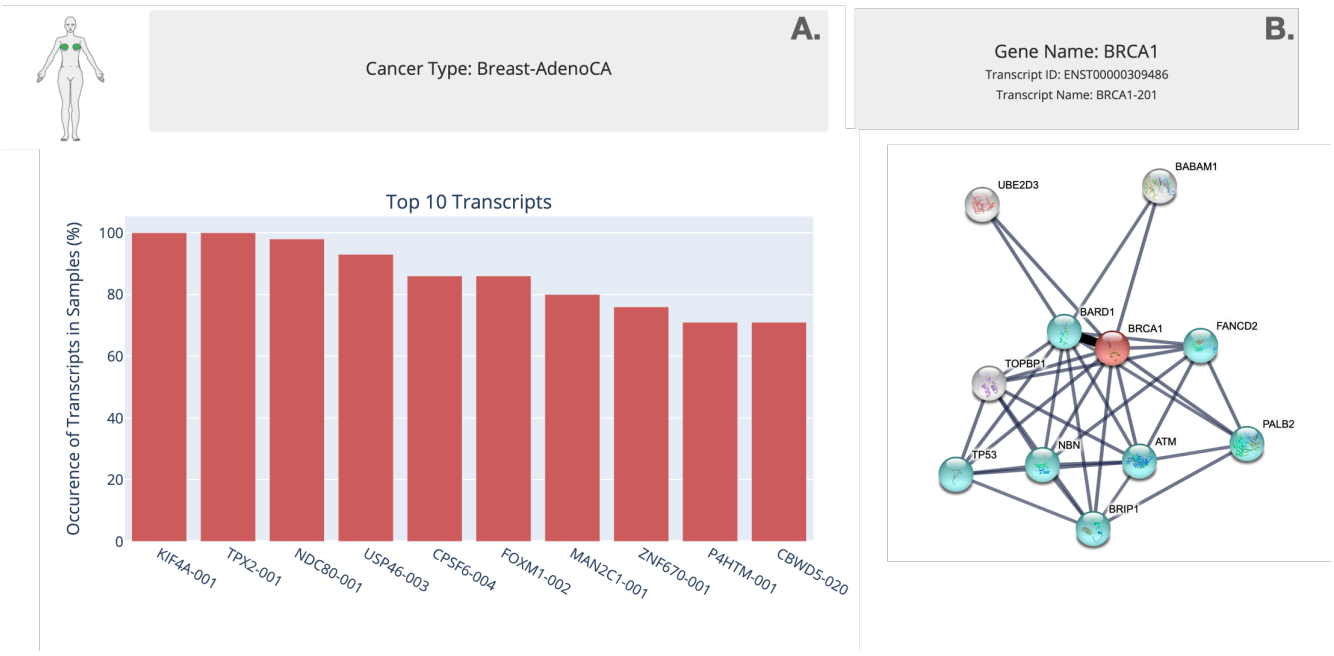

Figure 1. Example case from CanIsoNet: **A**. Top 10 cMDTs, which are most frequently observed in Breast-Adenocarcinoma samples. **B**. Interaction network of the isoform BRCA1-201. Red sphere shows the protein of interest, spheres in turquoise highlight Cosmic cancer census genes. Black line represents the lost interaction from a cMDT, while gray lines represent the remaining interactions.

2 Implementation

CanIsoNet is built upon a MySQL database with a Python Flask back-end and JavaScript/AJAX extensions to the front-end. The Python Plotly library is used for plotting. All plots are zoomable and downloadable. The STRING network is constructed using the STRING API and graphically adjusted to visualize interaction disruptions and COSMIC cancer census genes using JavaScript. Domain structures of transcripts were generated by the wiggleplotr R package (Kaur Alasoo, 2020). Anatograms were produced by the gganatogram Rshiny app (Maag, 2018). Data in CanIsoNet can be searched or browsed via cancer types or subsequently by gene or isoform names. To construct CanIsoNet, we integrated the STRING database (version 10), with the domain-domain interaction database 3did (version 2018_04), the cancer mutation database COSMIC (release v94, 28th May 2021), the genome database Ensembl (version 75) and the protein domain family database Pfam (version 32).

Expression data of normal samples (GTEx version 4) and cancer samples were provided by the PCAWG project (see (Kahraman et al., 2020) for more details).

3 Results

CanIsoNet stores splicing information on a total of 7144 genes, 11,041 isoforms and 31,748 unique isoform switching events across 24 cancer types from 1209 cancer samples. For each cancer type, we list all detected cancer-specific Most Dominant Transcripts (cMDT) and highlight the top 10 most frequent ones (Figure 1.A). Among the top frequent cMDTs could be transcripts that are expressed in all samples of a cancer type, pointing out potential biomarker candidate, such as KIF4A-001, TPX2-001 in Breast Adenocarcinoma (Figure 1A).

Websites dedicated for cMDT, show an isoform-specific interaction network where disruptions are featured in a STRING interaction network (Figure 1.B). Furthermore, different statistics are provided such as the number of disrupted STRING interactions, the relative network density score of a gene, the type of domain-domain interactions that are lost due to isoform switch events or a list of COSMIC cancer census gene interactors. The integration of the COSMIC database is a unique feature of CanIsoNet, which allows users to assess the pathogenicity of a cMDT. Furthermore, for each PCAWG sample, the relative expression values of each cMDT and the median expression value of the MDTs in the matched normal tissue are shown on a sample specific page.

Lastly, tables of all cancer-type-specific cMDTs and all isoform-specific interactions can be browsed and downloaded via dedicated drop-down menus or the download page.

CanIsoNet is the first isoform-specific interaction network webserver for cancer-specific isoforms covering a total of 24 cancer types. With its user-friendly interface as well as rich annotations, CanIsoNet supports the discovery of functional and pathogenic isoform switch events.

4 Case Study

The proto-oncogene SRC protein is a non-receptor tyrosine and it is a hub protein interacting with proteins having role in many functions in the cell, including cell growth, cell migration, angiogenesis and survival. Its dysregulation has been reported in many cancer types (Wheeler et al., 2009). In CanIsoNet, we observe that a transcript of SRC, SRC-202, has been found as cMDT in Hepatocellular

carcinoma. The network density of SRC, indicating the density of STRING interactions in the local neighborhood, is 93%. This transcript losses ~20% of protein interaction (33 out of 167 interaction having domain-domain information). The reason is that the SH3 (between sequence 90-137) domain in SRC-202, enabling binding to other proteins, is lost in the transcript. The loss of this domain cause disruption in interaction with many cancer census genes such as MAPK1, ITK, RAF1 and STAT3. All these information can be found in the CanIsoNet.

Acknowledgements

We thank all members of the Moch lab, the Clinical Computational Group at University Hospital Zurich, and the von Mering Lab at the University of Zurich for their valuable feedbacks and constant support.

Funding

This project was funded by Krebsliga Zürich.
Conflict of Interest: none declared.

References

Climente-González, H., Porta-Pardo, E., Godzik, A., and Eyras, E. (2017). The Functional Impact of Alternative Splicing in Cancer. *Cell Reports* 20, 2215–2226. doi:10.1016/j.celrep.2017.08.012.

Kahraman, A., Karakulak, T., Szklarczyk, D., and von Mering, C. (2020). Pathogenic impact of transcript isoform switching in 1,209 cancer samples covering 27 cancer types using an isoform-specific interaction network. *Sci Rep* 10, 14453. doi:10.1038/s41598-020-71221-5.

Kaur Alasoo (2020). *wiggleplotr: Make read coverage plots from BigWig files. R package version 1.14.0.*

Louadi, Z., Yuan, K., Gress, A., Tsoy, O., Kalinina, O. V., Baumbach, J., et al. (2021). DIGGER: exploring the functional role of alternative splicing in protein interactions. *Nucleic Acids Research* 49, D309–D318. doi:10.1093/nar/gkaa768.

Maag, J. L. V. (2018). gganatogram: An R package for modular visualisation of anatograms and tissues based on ggplot2. *F1000Res* 7, 1576. doi:10.12688/f1000research.16409.2.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al. (2015). STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Research* 43, D447–D452. doi:10.1093/nar/gku1003.

Vitting-Seerup, K., and Sandelin, A. (2017). The Landscape of Isoform Switches in Human Cancers. *Mol Cancer Res* 15, 1206–1220. doi:10.1158/1541-7786.MCR-16-0459.

Vitting-Seerup, K., and Sandelin, A. (2019). IsoformSwitchAnalyzeR: analysis of changes in genome-wide patterns of alternative splicing and its functional consequences. *Bioinformatics* 35, 4469–4471. doi:10.1093/bioinformatics/btz247.

Wheeler, D. L., Iida, M., and Dunn, E. F. (2009). The Role of Src in Solid Tumors. *The Oncol* 14, 667–678. doi:10.1634/theoncologist.2009-0009.