# Remus: a web-based tool for building a tissue-specific map of regulatory elements for a set of genes

Paweł Sztromwasser[1,*], Damian Skrzypczak[1,3], Wojciech Fendler[1,2]

1. Department of Biostatistics and Translational Medicine, Medical University of Lodz, Łódź, Poland
2. Department of Radiation Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, USA
3. Wroclaw University of Enviromental and Life Sciences, Wrocław, Poland
(*) pawel.sztromwasser@umed.lodz.pl

## Background

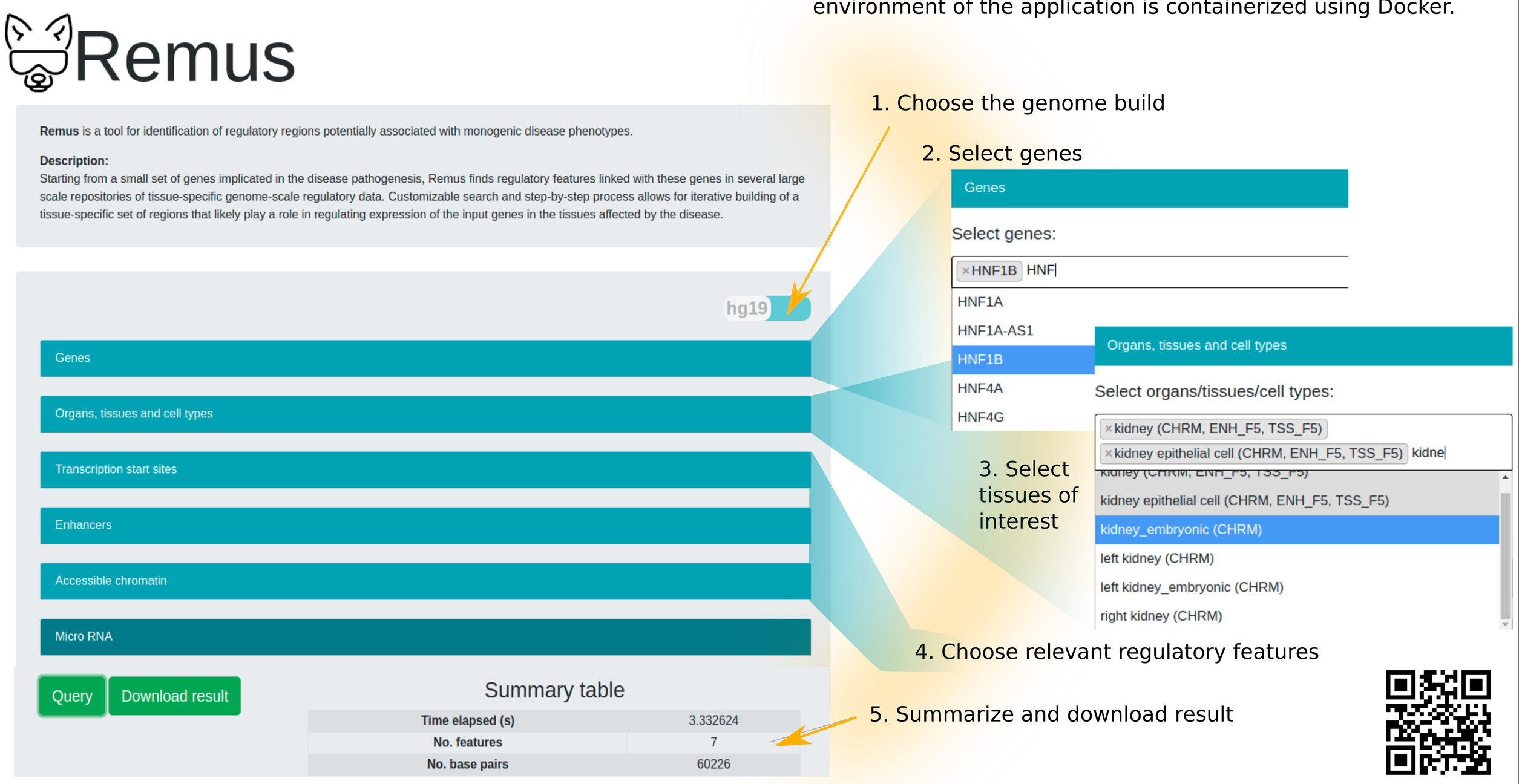
Frequent somatic mutations are the hallmark of cancer genomes. The main focus thus far has been on somatic mutations recurring in the coding sequence of genes. Recently however, several studies have also identified recurrent non-coding somatic mutations. The most notable example is the TERT promoter [1], but mutations in regulatory elements of other genes have also been reported [2].

Changes of non-coding regions can be specific to certain types of cancers and tissues, may contribute to tumorgenesis by acting in concert with changes of the coding regions, and potentially provide better insight into prognosis when evaluated together with gene mutations [2]. Finding non-coding mutations with a causative role in tumorgenesis is difficult though, mainly due to a large number of such events and limited understanding of their functional impact. One of the key challenges is linking mutations of the non-coding regions with their potential effector genes, thus predicting the outcome of such variants.

## Materials and Methods

Coordinates of regulatory regions were extracted from ENCODE [3] and FANTOM5 [4] repositories of tissue-specific regulatory data generated by genome-wide assays, such as ChIP-seq, CAGE-seq, and DNase-seq. After aggregating samples originating from the same tissues and cell-types, coordinates of regulatory regions categorized into promoters, enhancers, and accessible chromatin are provided for user operations in the application. Information about accessible chromatin *loci* is also used to filter non-tissue-specific interactions, such as microRNA - transcript interactions from miRTarBase [5] (experimental) and miRWalk [6] (predicted). The connection between genes and regulatory features is made based on experimental evidence of interaction or user-tunable distance in the genome.

The Web application is implemented in Python using Flask, and genomic coordinate arithmetics in pybedtools. Computational environment of the application is containerized using Docker.




## Results

Remus is a Web application that facilitates identification of regulatory regions potentially associated with the expression of input genes. Although the tool is primarily developed for searching pathogenic variants in rare-disease studies, it can be useful in cancer research, for instance to identify coordinates of tissue-specific regulatory features of oncogenes.

Starting from a small set of input genes implicated in pathogenesis of a disease, Remus allows creating a list of regulatory features active in a set of tissues of interest. Customizable search and step-by-step process allows for iterative building of a list of coordinates representing genomic locations of elements that likely play a role in regulating expression of the input genes in the user-selected tissues. The coordinates can subsequently be used to filter variants in search for mutations with pathogenic potential.

## Acknowledgements

Project funded by NCN Polonez grant no 2016/23/P/NZ2/04251. The project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 665778.

## References

[1] Huang FW. *et al*. (2018) Science
[2] Zhang W. *et al.* (2018) Nature
[3] The ENCODE Project Consortium (2012), Nature
[4] The FANTOM Consortium (2014) Nature
[5] Chou CḦ. *et al.* (2018) Nucleic Acid Research
[6] Sticht C. *et al.* (2018) PloS One