MARINA VABISTSEVITS

Doctoral student at the University of Bristol with a strong background in bioinformatics and data science, working in the interdisciplinary research area of applying data mining methods to answer epidemiological questions

T	➡ EDUCATION
2023 2019	 PhD, Data mining in Epidemiology University of Bristol Pristol, UK
	• Thesis: "Data mining breast cancer epidemiological relationships"
2017 2015	MSc, Bioinformatics University of Copenhagen Orenhagen Orenhagen
	• Master's thesis (RNA-Seq analysis): "Identification of autophagy signatures in breast cancer using The Cancer Genome Atlas data"
2015	BSc, Biochemistry
2011	University of Bath VBath, UK
	• Final project (structural bioinformatics): "Structural analysis of substrate binding in <i>Sulfolobus solfataricus</i> 2-keto-3-deoxygluconate aldolase variants"
<u>[</u>	RESEARCH EXPERIENCE
2023	 Doctoral Student MRC Integrative Epidemiology Unit, University of Bristol P Bristol, UK
2019	 Mini-project 1: Implemented a generalisable R workflow of performing multivariate correlation analysis (metaCCA) on GWAS summary statistics data from OpenGWAS to identify pleiotropic variants in the measured traits Mini-project 2: Conducted a Mendelian randomization (MR) study to investigate the mechanism mediating the effect of early life BMI on breast cancer risk (see <i>Publications</i>) PhD project: Using EpiGraphDB (a Neo4j knowledge graph) as a data mining tool to investigate causal relationships and facilitate evidence triangulation in breast cancer epidemiology research. Analysing MR-EvE (MR Everything-vs-Everything) causal associations between OpenGWAS traits and looking for supporting evidence and mechanistic detail in literature-mined semantic relationships linked to phenotypes.
2017 I	Visiting Researcher / Master's Thesis Student Ocopenhagen Depmark
2016	 For my master's project, I was invited to DCRC as a visiting researcher after building positive working relationships during collaborations earlier in my master's degree. Explored TCGA breast cancer gene expression RNA-Seq data to identify the involvement of autophagy-related genes in certain disease subtypes. Derformed an extensive exploratory data analysis in D. followed by
	differential expression analysis using Bioconductor packages and enrichment analysis, allowing me to find over-represented autophagy genes
	• Built an efficient workflow in R, and identified and applied a new method to this field, which was enthusiastically adopted by other group members

CONTACT

marina.vabistsevits
 @bristol.ac.uk
 marina_vab
 github.com/mvab
 marinalearning.netlify.com
 in linkedin.com/marina-

vabistsevits-82590443

FECH SKILLS

R / tidyverse		
Python		
Git		
Bash / awk		
Neo4j / Cyph	er	
SQL		
Docker		
Snakemake		
AWS		

OTHER SKILLS

Data visualisation	
Reproducible research	
Documentation writing	
Remote collaboration	

CV made with the R package pagedown.

Code available at github.com/mvab/cv.

Last updated on 2022-12-22.

INDUSTRY EXPERIENCE

2021	•	Data Scientist (Senior Executive Officer)				
 2020		Public Health England QUK				
		 Completed a 6-month part-time secondment at PHE during my PhD, working within the Data Surveillance and Contact Tracing team to assist with COVID-19 transmission data analysis 				
		 Worked with a large group of epidemiologists and R developers on processing and analysing the daily cases and contacts data to extract valuable insights into the pandemic development in the UK 				
		 Contributed to the investigation of the <i>alpha</i> variant identified in December 2020, which was published as a technical briefing for the government 				
2020 2017	•	Bioinformatician				
		Living DNA \mathbf{Q} Frome, UK				
		 Led the research work on improving ancestry reference panels used by the core pipeline behind the company's direct-to-consumer ancestry genetics test, bringing considerable improvement to results accuracy 				
		Gained experience in working with a legacy codebase through maintaining and contributing to the in house ninglines (Dythen)				
		 Honed my R programming skills by switching to the tidyverse approach and advanced my data visualisation skills 				
2017	•	Student research assistant in the Big Data group				
2016		3Shape ♀ Copenhagen, Denmark				
2016		 Performed data preparation and visualisation tasks in Python, gaining practical experience of programming in a professional environment Used deep learning framework Caffe2 to develop a neural network training pipeline for scan image classification tasks 				
2014		Placement Student in Bioinformatics team				
		Oxford Gene Technology Q Oxford, UK				
2013		 Became responsible for a multitude of exome- and RNA-Seq projects, running in-house data analysis pipelines and performing custom analysis for different projects 				
		PUBLICATIONS				
2022	•	Integrating Mendelian randomization and literature-mined evidence for breast cancer risk factors				
		<u>M. Vabistsevits</u> , T. Robinson, B. Elsworth, Y. Liu, T. R. Gaunt; MedRxiv				
2021	•	Deciphering how early life adiposity influences breast cancer risk using Mendelian randomization				
		<u>M. Vabistsevits</u> , G. D. Smith, E. Sanderson, T. G. Richardson, B. Lloyd- Lewis, and R. C. Richmond; Communications Biology				
2020		EpiGraphDB: a database and data mining platform for health data science				
		Y. Liu, B. Elsworth, P. Erola, V. Haberland, G. Hemani, M. Lyon, J. Zheng, O. Lloyds, <u>M. Vabistsevits</u> , T.R. Gaunt; Bioinformatics				
		Other publications are available in Google Scholar profile				