C U R R I C U L U M   V I T A E

NAME
Protsvetkina Anastasia Vladimirovna

E-mail: procvetkina-nastya@mail.ru

Skills

- Solid mathematical background: algebra, probability & statistics
- R, Python (Django), SAS
- Linux
- Git
- Data structures & algorithms
- Bcftools, samtools, bwa-mem, plink, Beagle 5, Impute 5, haploview, Delly, CNVnator, LUMPY, MELT, GRIDSS, FastQC, MaxQuant, Andromeda, xQuest
- Principles of Next Generation Sequencing (I-III generations), Microarray Genotyping, Comparative hybridization, ChiP-Seq, Hi-C, RNA-Seq, LC-MS
- Pipelines creation with Snakemake
- Fundamentals of Neural Networks & Machine Learning
- Molecular biology fundamentals
- Latex

Employment and experience

10.2020 - 04.2021: Research engineer at Dobzhansky Center for Genome Bioinformatics, St.-Petersburg State University

01.2022 - till now: Bioinformatician, Institute of Systems Biology & Medicine, Moscow

My experience includes:

1. **Participation in the project devoted to studying genomic integrity of patients with cancer (since 09.2020 - 04.2021)**

   I was responsible for selection of suitable tools for structural variants detection and following pipelines creation with Snakemake. I worked with alignments & variant calls with samtools and bcftools. My colleagues & I worked on development of wavelet-based genomic instability metrics. I tried different coverage normalizations to see their impact on the metrics proposed. Also I analyzed Genome In a Bottle papers & metrics behaviour on their gold standard data. In one of my latest tasks I needed to analyse coverage wave-forms for patterns in certain types of genomic rearrangements for db creation.

2. **Teaching at St.-Petersburg Electrotechnical University "LETI" (since 09.2019 - till now)**

   Up to now I've been teaching 3 types of seminars for 1-3 year students. Namely, linear algebra, probability & statistics. 1-semester statistics course covers: rigorous point estimation (ML, REML), UMVUE theory &

construction, hypothesis testing, quadratic forms distribution, LRT, Wald test, UMP tests, confidence ellipsoids, LM, GLM, GLMM, GEE, ANOVA, contingency tables analysis, multiple testing issues, basics of epidemiological studies designs.

3. **Development of a new mathematically rigorous method to identify genomic regions associated with phenotypes in my Bachelor's final project. The tool is implemented in R (09.2020 - 05.2021)**

I had to estimate the correlation matrix of a multivariate normal vector constructed from partitioned chi-squared statistics from a sequence of contingency tables. First of all, I had to find such an orthogonal transformation so that the chi-squared statistic with $(n-1)(m-1)$ df is a sum of squares of $(n-1)(m-1)$ iid normal variables. For the needs of genetics 2x2 & 3x2 tables have been considered. Point estimate of the correlation matrix has been obtained. A method for controlling for PSD-property of the matrix has been proposed.

4. **Analysis of genotyping data in alcohol dependence study (09.2020-11.2020)**

Under supervision of a biologist & a mathematician worked with genotyping data of a St.-Petersburg cohort of alcohol dependent subjects. Tasks started from QC & preprocessing, followed by GWAS-analysis and ended with the results analysis and comparison with known associations from publicly available GWAS databases (DOI:10.1016/j.euroneuro.2021.08.262)

5. **Accurate imputation study (01.2022 - till now)**

Imputation of missing variants is a common practice in GWAS studies, though there's no procedure to control for imputation error. I'm studying factors affecting imputation accuracy & trying to come up with empirically-based standards which will allow to maintain a predefined error rate.

6. **Neural networks construction & training**

During the machine learning course I used Python language and Keras library for neural networks creation & training. I worked with dense, convolutional & recurrent NN for analysis of numeric and text data. Tasks included classification, regression and text generation.

Education

Bachelor of Science in Applied Mathematics and Informatics, 2021, St.-Petersburg Electrotechnical University "LETI". Pursuing Master's degree in Applied Mathematics (bioinformatics track) in St.-Petersburg Polytechnical University.

SUMMER SCHOOLS & WORKSHOPS

INSTITUTION AND LOCATION	SUBJECT
Bioinformatics Institute, St.-Petersburg (2020)	Bioinformatics Summer School
Mathematical Center in Academgorodok, Novosibirsk State University, Novosibirsk (2020)	Workshop in application of Systems Biology & Bioinformatics for Genome Analysis

Languages

Advanced level of English enables me to read almost anything in the scope of my interest, no matter whether it is a classic 19-th century novel or a textbook on statistics. In the past I participated in lots of linguistics olympiads. In the fall of 2020 worked as an assistant during masters' english-taught course in "Advanced Mathematical Methods". Basically, everything I know I can teach in English. Current English level – C1.

Current professional interests

I'm willing to go deeper in statistics & data analysis. Especially interested in biostatistics & bioinformatics. Not planning to quit teaching either, 'cause it allows me to get the full picture of the maths I love.