

ALES VARABYOU, Ph.D.

ales.varabyou@jhu.edu | (443)531-2179 | [LinkedIn](#) | [GitHub](#) | [Personal Page](#) | Baltimore, MD

Genomics Research Scientist and Engineer developing open-source algorithms and pipelines for large-scale DNA and RNA sequencing analysis, with a focus on performance and reproducibility. My work covers genome annotation, multi-sample RNA-seq processing, protein discovery, and viral genome analysis. I am the lead developer of CHESS, the first evidence-based human gene catalog built from 20,000+ RNA-seq experiments now integrated into major databases. I currently lead the HIVAtlas initiative focused on creating the first reference annotation for thousands of viral genomes.

PROFESSIONAL EXPERIENCE

Research Scientist | Pertea Lab, Johns Hopkins University | 2023 - Present

- **Launched and led [HIV Atlas](#) collaboration:**
 - Designed algorithms for genome annotation transfer in hypervariable viral genomes (Vira, Python) and by-reference correction of spliced alignments (Snapper, Python).
 - Built first HIV-1 reference genome annotation and scaled to thousands of clinical isolates.
 - Created React-based single-page interface for interactive exploration of the data.
- **Improved spliced alignment of chimeric RNA:**
 - In collaboration with scientists from HJF/WRAIR/MHRP developed [IRIS](#) (Python, Rust) - a splice-aware chimeric RNA identification software.
 - Produced the first account of effects of viral integration on host gene expression during acute stages of viremia in SIV-infected macaques.
- Supervised graduate, undergraduate and high school students research into transcriptome quantification, transcriptome quantification, metagenomic database curation and genome annotation.
- Designed a SQL/Flask/React comparative transcriptomics [web platform](#) for creating custom genome annotations from multiple sources (CHESS, RefSeq, ENSEMBL, MANE, FANTOM).

Graduate Research Assistant | Salzberg Lab, Johns Hopkins University | 2017 - 2023

- **[CHESS](#) genome catalogue project lead for 7+ years:**
 - Designed a sample-to-annotation protocol for condition specific gene, transcript and protein finding.
 - Identified and characterized tens of thousands of conditions-specific transcriptional products from 20,000+ RNA-seq experiments across 54 tissue types.
- **Data compression and representation with [TieBrush](#) (Rust, C++):**
 - Designed and developed a suite of tools for multi-sample processing of TB-scale sequencing datasets, reducing processing time and storage requirements by 85%.
 - Identified key predictors of functional and noisy splicing signals in RNA-seq data.
 - Developed ML model to predict valid splice sites improving assembly specificity by over 40%.
- **Protein prediction with [ORFanage](#) (C++):**
 - Designed and developed an ultra-efficient and sensitive system for protein prediction in novel transcripts achieving >95% accuracy.
 - Predicted novel proteins and validated function with PhyloCSF and AlphaFold2.
- **Designed and implemented [Bolotie](#) (C++)** - the first algorithm capable of detecting recombination across millions of viral genomes using Markov Model trained on millions of sequences.

EDUCATION

Ph.D. in Computer Science (Genomics)

Johns Hopkins University | 2017 – 2023

Advisors: Dr. Steven Salzberg & Dr. Mihaela Pertea

Thesis: Computational Study of Transcriptional Landscapes from RNA-seq Data

M.Sc.Eng. in Computer Science | Johns Hopkins University | 2017 – 2020

Honors B.A. in Biology and Computer Science | Luther College | 2013 – 2017

SKILLS

Genomics & Bioinformatics:

- 10+ years developing custom algorithms for RNA-seq (bulk, single-cell, spatial) and DNA-seq analysis.
- Applied experience developing and deploying end-to-end bioinformatics workflows: from DNA and RNA alignment and assembly to differential expression, clustering and metagenomic classification.
- Extensive experience working with all major sequencing technologies, from short-read Illumina to long-read platforms (Oxford Nanopore, PacBio).
- Skilled in creating publication-quality data visualizations (matplotlib, ggplot) and interactive web interfaces (React, D3.js).

Programming: Python | C++ | Rust | shell | Unix | R | OpenMP | SQL | Docker | Singularity | Git | D3js

Languages: Belarusian (Native) | English (Native) | Russian (Native) | Polish (Conversational)

SELECTED PUBLICATIONS (Full list: [Google Scholar](#))

1. **Varabyou A.**, Artamonov M., Bolton D., Salzberg S., Pertea M. (2025). Comprehensive Transcriptome Annotation of Thousands of HIV-1 Genomes. *bioRxiv*, 2025-09
 2. **Varabyou A.**, Erdogdu B., Salzberg S., Pertea M. (2023). Investigating open reading frames in known and novel transcripts using ORFanage. *Nature Computational Science*, 10.1038/s43588-023-00496-1
 3. **Varabyou, A.**, Sommer, M. J., Erdogdu, B., Shinder, I., Minkin, I., ... & Pertea, M. (2023). CHESS 3: an improved, comprehensive catalog of human genes and transcripts based on large-scale expression data, phylogenetic analysis, and protein structure. *Genome biology*, 24(1), 249.
 4. **Varabyou A.**, Pockrandt C., Salzberg S., Pertea M. (2021). Rapid detection of inter-clade recombination in SARS-CoV-2 with Bolotie, *Genetics*, 10.1093/genetics/iyab074
-

SERVICE & LEADERSHIP

- **Invited Speaker:** Harvard University, Banbury Conference, Yale University, Medical School Intensive.
- **Peer Reviewer:** Nature, Genome Biology, ISMB/ECCB, MBE, NAR, OUP Bioinformatics and more.
- **Open-Source Contributor:** MUMmer4, gffread, gffcompare and more.
- **Technical advisor** consulting on variant calling and predictive modeling for Aevus Diagnostics.
- **JHU Transportation Initiative:** Designed and implemented a [real-time tracking system](#) to improve accountability and reliability for the JHU campus shuttle service.
- **Graduate Student Representative:** Represented graduate students on the Provost's Advisory Team on Healthcare, successfully advocating for improved university-wide health insurance policies.
- **Member** of the Belarus National UWC (United World Colleges) Committee.