Zachary Lee Skidmore

Senior Bioinformatician & Data Scientist

Data Science | Machine Learning | Oncology R & D | Pipeline/Workflow Development | Data Visualization | Bioinformatics | Statistics

St. Louis, MO | ORCiD | GitHub | https://zlskidmore.github.io

Professional Summary:

Senior bioinformatics and quantitative scientist with 12+ years spanning Tier 1 academic research and industry. Proven applied experience across data science, machine learning, statistical inference, and pipeline/software development within bioinformatics and oncology. Author of 50+ high-impact publications and patent co-inventor. Mentor and instructor, developed and taught multiple bioinformatics courses. Seeking a senior role to apply deep technical expertise and background.

Skill Summary *(See Page 2 for a complete list)*:

Education:

University of Illinois - Chicago; Master of Engineering in Bioinformatics — May 2013

Work Experience:

Delfi Diagnostics — Remote

Position: Senior Quantitative Scientist, Translational Science, Data Science - Research — Aug 2022 - Oct 2025

2023 - 2025 Highlights

- *Co-inventor for provisional patent to differentiate histological subtypes of lung cancer using high-dimensional cfDNA features (WO202417327A2)*
- *Increased sensitivity of nucleosome depletion features via generalized additive models, improving feature performance*
- *Built and validated dimensionality reduction workflows (PCA) and projected incoming data on learned space for sample subtyping*
- *Applied unsupervised learning approaches and hierarchical clustering to identify and prioritize candidate biomarkers off of NDR features.*
- *Lead author and bioinformatic lead for development and application of ctDNA monitoring on a prospective phase III clinical trial (CAIRO5) using fragmentomics including survival analysis and hypothesis testing PMID: 39433569*
- *Maintained and iteratively improved a production level random forest model to predict ctDNA fraction from plasma*
- *Designed probabilistic binomial models to quantify the impact of sequencing coverage and ploidy on feature*
- *Engaged with drug development companies in the application and analysis of ctDNA monitoring across multiple cancer types*
- *Developed and cross-validated machine learning models using engineered features*
- *Delivered and presented interpretable analytical results to external collaborators and cross-functional stakeholders, supporting multi-million-dollar partnerships*
- *Executed 15+ end-to-end applied data science and translational research projects, from inception through publication, resulting in multiple domestic and international conference presentations and manuscripts*

McDonnell Genome Institute, Washington University Department of Medicine — St. Louis, MO

Position: Senior Staff Scientist - Griffith Lab — Nov 2013 - Aug 2022

2013 - 2022 Highlights

- *Lead author and analyst for a study of non-cirrhotic hepatocellular carcinoma (WGS/WES/Custom Capture) PMID: 35568002*
- *Developed a protocol for scTCRseq and scRNAseq applied to canine samples PMID: 38649520*
- *Implemented modifications to existing H. sapiens workflows to be compatible with multiple species PMID: 33556121*
- *Built a custom workflow to detect viral integration sites from dual indexed viral reads PMID: 34019588*
- *Designed, cross-validated, and implemented a machine learning model to classify and refine somatic variant calls using XGB models*
- *Refactored the RegTools analysis software package to be faster and less memory intensive obtaining a 20x boost in performance*
- *Developed custom Common Workflow Language (CWL) bioinformatic pipelines for institute wide use: bioinformatic, analysis-workflows*
- *Created and maintained Docker images for bioinformatic analysis: https://hub.docker.com/u/zlskidmore*
- *Trained and mentored staff scientists, postdocs, and graduate students in various bioinformatic and technical skill sets*
- *Created static websites and subsequent material for genomic visualization, precision medicine, and RNAseq for training purposes*
- *Performed analysis on TCR clonotype repertoire, HLA prediction, neoantigen prediction, ctDNA, WGS, WES, RNAseq PMIDs: 31390566, 31618044, 31467059, 31685621, 31902496, 32665297, 34036230, 33556121, 33547198*
- *Prototyped shiny application for interrogation of saturation mutagenesis assays: saturation-mutagenesis-viz*
- *Led development and publication of Genomic Visualization in R (GenVisR) library on bioconductor PMIDs: 27288499, 34506690*
- *Consulted on design and implementation of Clinical Interpretations of Variants in Cancer (CIViC) database; Co-organized first Biohackathon for CIViC at the Netherlands Cancer Institute; Submitted and curated database entries PMID: 28138153*
- *Assisted in the maintenance and improvement of institute project management and analysis software PMID: 26645048*
- *Conducted statistical and bioinformatic analysis for next generation sequencing data (Benchmarking Alignment, Clonotype Expansion, Cancer Genomic Alterations, etc.): PMID 26531824, 26563128, 27058228, 27029710, 27181063, 27681435*
- *Performed next generation sequencing analysis and presented results for cancer genomic tumor board n-of-1 cases*

Workshops, Instruction & Community Leadership:

- *Teaching assistant/content author for Cancer Research Institute bioinformatics workshop* — 2024, 2025
- *Instructor/contributing author for genomic visualizations workshop (https://genviz.org)*
  - *Physalia-courses, Freie Universität Berlin, Germany* — 2017, 2019
  - *Evomics - Workshop on Genomics 2019, Český Krumlov, Czech Republic* — 2019, 2020
- *Teaching assistant/contributing author for Precision Medicine workshop (https://pmbio.org)*
  - *PR Informatics, Glasgow, Scotland* — 2018
- *Contributing author for RNAseq workshop (https://rnabio.org)*

Selection of Scientific Expertise	
Advanced (***): Extensive multi-year experience with deep understanding. Have mentored, taught, or critically evaluated the topic. **Proficient (**):** Multi-year practical experience. Demonstrated use in published research or public repositories. **Familiar (*):** Working knowledge through guided exposure or light use.	
cfDNA Fragmentomics & Liquid Biopsy Data Analysis	Statistical Analysis
• *shallow-WGS Fragmentomic Profiling**** • *shallow-WGS Aneuploidy Detection**** • *Nucleosome Footprinting**** • *Early Cancer Detection and Modeling***	• *Hypothesis Testing (Parametric/Non-Parametric)**** • *Survival Analysis (Kaplan-Meier)**** • *Batch Effect Detection/Correction*** • *Power Analysis*** • *Distribution-based Modeling (e.g., binomial, negative binomial, poisson)***
Next-Generation Sequencing Analysis	Cancer Immunotherapy Data Analysis
• *Pre/Post Alignment QC (e.g. FastQC, Picard, samtools)**** • *Read Alignment (Burrows-Wheeler & K-mer Based)**** • *Somatic/Germline Variant Calling and interpretation**** • *Variant Annotation and Filtering (e.g. VEP)**** • *CNA/SV Detection**** • *LOH Detection**** • *Illumina-based WES/WGS/RNA-seq workflows**** • *UMI-Aware Processing (e.g. fgbio)***	• *Checkpoint Inhibitor Biomarker Profiling*** • *TCR Sequencing and Clonotype Repertoire Diversity*** • *HLA-Typing (e.g. OptiType)*** • *Neoantigen prediction (e.g. pVACtools)**** • *Immune Cell Deconvolution (e.g. CIBERSORT)*** • *TIL quantification from RNAseq***
Cancer Genomics & Tumor Biology Bioinformatics	Machine Learning and Modeling in Genomics
• *Liquid and Solid Tumors**** • *Tumor Heterogeneity and Clonal Evolution**** • *Pathway Analysis**** • *Genomic Visualization and Interpretation**** • *Tumor Purity Estimation**** • *Data Mining and Large Cohort Processing****	• *Random Forest*** • *Extreme Gradient Boosting*** • *Generalized Linear Models*** • *Cross-Validation*** • *Model Evaluation (ROC, AUC, Precision-Recall)***
RNAseq Analysis	Experimental Design
• *Differential Expression**** • *GSEA*** • *Transcript Quantification**** • *Alternative Splicing**** • *scRNA-seq***	• *Cohort Stratification*** • *Univariate and Multivariate Analysis**** • *Longitudinal Study Frameworks*** • *Prospective/Retrospective Study Design***

Selection of Technical Skills		
Advanced ***	**Proficient ****	**Familiar ***
• *R (inc. ggplot2, data.table, S4)* • *UNIX* • *GNU Utils (incl. awk, sed, vim)* • *Docker (incl. Image building)* • *Git/GitHub* • *High Performance Computing (e.g. LSF, Slurm)* • *Amazon Web Services (incl. EC2, S3)* • *Confluence/JIRA*	• *Python* • *Common Workflow Language and Pipeline Implementation* • *Perl* • *Jekyll/Static Site Building* • *HTML/Web Development* • *Markdown* • *Cancer Genomics Databases (TCGA, GTEx, cBioPortal, COSMIC, etc.)*	• *SQL* • *CSS/SASS* • *D3.js*