Jenny Leopoldina Smith, MSc

Email: JennyL.Smith12@gmail.com **Phone:** (603) 801-9876
Code: github.com/jennylsmith **LinkedIn:** linkedin.com/in/jenny-leopoldina-smith

Bioinformatician and R programmer applying methods in data science and data engineering to biomedical research questions, with expertise in genomics and next generation sequencing (NGS) data.

Research and Work Experience

Computational Biologist, Senior Seattle Children's, *Seattle, WA*
April 2022 – Sept 2023 *Research Scientific Computing Dept.*

- Developed and adapted reproducible genomics workflows using Nextflow and nf-core tools for RNA-seq quantification, Cut&Run, ATAC-seq, and PacBio Isoseq; enabled use of containerized software with singularity/apptainer and the PBSpro executor on a high-performance compute cluster (HPC).
- Analysis and visualization of multi-sample scRNA-seq and scATAC-seq datasets with Cellranger quantification, followed by doublet detection, ambient DNA correction, dataset integration (SCANVI and Seurat v4/v5). Production of multi-omics analyses for bulk transcriptomic and single-cell datasets using dplyr and tidyverse principles.
- Generation of complex visualizations include heatmaps, circos plots, oncoprints, 3D scatter plots, genomic tracks, network graphs, and volcano plots, among others, to identify actionable insights from multivariate biological datasets from NGS and public databases (ensembl, UCSC, Genomic Data Commons).
- Engagement in bioinformatics support by providing guidance and troubleshooting assistance at office hours and teaching internal courses.

Bioinformatics Analyst Fred Hutchinson Cancer Center, *Seattle, WA*
January 2017 – March 2022 *Principal Investigator:* Dr. Soheil Meshinchi

- Profiled pediatric AML subtypes defined by gene fusions and mutations using RNA-sequencing with supervised and unsupervised clustering ML algorithms, statistical regression and classification, in a multidisciplinary environment with biologists, bioinformaticians, clinicians, and lead investigators.
- Created an R package (DeGSEA) using Roxygen and 'usethis' framework to streamline the association of clinical covariates and RNA-seq expression data.
- Initiated a lab github organization at github.com/Meshinchi-Lab.
- Developed Nextflow pipelines using docker for cloud computing on AWS and with singularity for on-prem HPC; processed 1000's of RNA-seq samples for gene expression quantification and fusion detection.

Postbaccalaureate Fellow National Institutes of Health, *Bethesda, MD*
June 2014 - June 2016 *Principal Investigator:* Dr. Maria Morasso

- Investigated the function of homeodomain protein DLX3 in keratinocytes at NIAMS. Bioinformatics analysis of ATAC-seq and ChIP-seq (peak calling, motif enrichment, NGS.plot, Deeptools) from knock-out murine models.

Science Teacher Agua Fria High School, *Avondale, AZ*
June 2012 - May 2014

- Teach For America Phoenix corps member; certified for secondary biology and chemistry education. Courses taught: general biology, AP biology, general chemistry and integrated science.

Experience with NGS Data

- RNA-seq – Illumina, PacBio
- scRNA-seq (single-cell)
- scATAC-seq
- ATAC-seq
- miRNA-seq
- Cut&Run and ChIP-seq

Experience with Software

- R programming: base R, Tidyverse, Bioconductor
- R package development
- Bash/unix shell scripting
- Python programming
- Quarto and Rmarkdown
- Git version control
- Github and bitbucket
- Nextflow and nf-core
- Snakemake
- PBS/SLURM HPC schedulers
- Containerization: docker, singularity, apptainer
- Amazon Web Services: AWS S3 and Batch
- Jira and agile framework, Atlassian products
- CI/CD with Bamboo

Bioinformatics Analysis

- Single-cell analysis for scRNA-seq and scATAC-seq
- Differential expression: Limma, EdgeR, DESeq2
- Fusion detection for RNA-seq: STAR-Fusion, CICERO, TransAbyss
- Isoseq3 pipeline for long-read RNA-seq (PacBio) data
- Motif analysis for Cut&Run, ChIP-seq, ATAC-seq

Jenny Leopoldina Smith, MSc

Email: JennyL.Smith12@gmail.com **Phone:** (603) 801-9876
Code: github.com/jennylsmith **LinkedIn:** linkedin.com/in/jenny-leopoldina-smith

Education

Applied Bioinformatics and Genomics, M.Sc. June 2016 – September 2017
University of Oregon, *Eugene, OR*

Secondary Education in Science, M.Ed. August 2012 – May 2014
Arizona State University, *Phoenix, AZ*

Biology, B.A., *magna cum laude* August 2008 – May 2012
University of San Diego, *San Diego, CA*

Publications

- Bertrums, E. J. M.*, **Smith, J. L.*,** Harmon, L.*, … Meshinchi, S. Comprehensive molecular and clinical characterization of NUP98 fusions in pediatric acute myeloid leukemia. *Haematologica* **2023**. DOI:10.3324/haematol.2022.281653
- Farrar, J. E., **Smith, J. L**., Othus, M., … Meshinchi, S. Long Noncoding RNA Expression Independently Predicts Outcome in Pediatric Acute Myeloid Leukemia. *Journal of Clinical Oncology* **2023**. DOI:10.1200/jco.22.01114
- Le, Q., Hadland, B., **Smith, J. L.,** Leonti, A., … Meshinchi, S. CBFA2T3-GLIS2 model of pediatric acute megakaryoblastic leukemia identifies FOLR1 as a CAR T cell target. *Journal of Clinical Investigation* **2022**. DOI:10.1172/jci157101
- Huang, B. J., **Smith, J. L**., Farrar, J. E., … Meshinchi, S. Integrated stem cell signature and cytomolecular risk determination in pediatric acute myeloid leukemia. *Nature Communications* **2022**. DOI:10.1038/s41467-022-33244-6
- Huang, B, **Smith, JL,** … Meshinchi, S. *CBFB-MYH11* Fusion Transcripts Distinguish Acute Myeloid Leukemias with Distinct Molecular Landscapes and Outcomes. *Blood Advances* **2021.** DOI: 10.1182/bloodadvances.2021004965
- **Smith, JL,** Ries, RE, Hylkema, … Meshinchi, S. Comprehensive Transcriptome Profiling of Cryptic *CBFA2T3-GLIS2* Fusion-positive AML Defines Novel Therapeutic Options – A COG and TARGET Pediatric AML Study. *Clinical Cancer Research* **2019**. DOI: 10.1158/1078-0432.CCR-19-1800
- **Smith, JL,** Maden SK, Lee, D, Buie, R, … Busby, B. Consensus Machine Learning for Gene Target Selection in Pediatric AML Risk. *bioRxiv* **2019**. DOI: 10.1101/632166

Additional Publications: https://orcid.org/0000-0003-0402-2779

Presentations

- **American Society for Hematology,** poster on ETS Family Transcription Factor Fusions in Childhood AML: Distinct Expression Networks and Clinical Implications (Virtual, 2021).
- **Molecular Medicine Tri-conference**, panelist on AI in Genomics and Precision Medicine (San Francisco, CA, 2020).
- **American Society for Hematology**, oral presentation on Comprehensive Transcriptome Profiling of Cryptic *CBFA2T3-GLIS2* Fusion-Positive AML Defines Novel Therapeutic Options — a COG and Target Pediatric AML Study (San Diego, CA, 2018).
- **American Society for Hematology**, oral presentation on The LSC17 Leukemic Stem Cell Signature Predicts Outcome in Pediatric Acute Myeloid Leukemia. Abstract Achievement Award Recipient (Atlanta, GA, 2017).

- Differential peak binding and accessibility analysis: DiffBind, edgeR
- DNA variant calling: GATK, Ensembl Variant Effect Predictor (VEP), SnpEff
- Genomic alignments: STAR-aligner, BWA, Bowtie
- NGS quality control: MultiQC, FastQC, Deeptools, Picard

Continuing Education and Volunteering

- International Society for Computational Biology (ISCB) member, 2023
- RLadies event organizer, Seattle, WA, 2018-2023
- Bioconductor conference organizer, Seattle, WA, 2022
- Mentor, UO Bioinformatics and Genomics Graduate Program, 2021- 2022
- SnpReportR, Carnegie Mellon and DNAnexus Hackathon, 2021
- Pacific Biosciences Isoseq3 analysis training, 2020
- Nextflow workflow development training, 2020
- Women In Biology, MAPS Mentorship Group, 2017-2019
- ConsensusML, NCBI Hackathon, 2018
- Mentor, Fred Hutch High School Internships, 2018