Alex Amlie-Wolf, PhD

Ashland, MA | bitbucket.org/alexamlie | alexamlie.github.io | he/him

Highly collaborative computational biologist with 13 years of experience in genomics and transcriptomics, large-scale data integration, RNA biology, and human genetics. 5+ years of industrial target nomination and drug development experience in disease areas including neuroscience, repeat expansions, immunology, and oncology. Experienced in people management, cross-functional collaboration, and communication with a wide range of audiences. Passionate about developing and applying computational methods to derive actionable biological insights from complex datasets.

Professional Experience

ARRAKIS THERAPEUTICS **WALTHAM, MA**

Scientist II, Computational Biology — July 2021 – Present

- Cross-functional, matrixed collaboration with biology and chemistry teams for drug discovery in neurology/repeat expansion, oncology, immunology, and additional indications. This includes target RNA characterization and computational biology leadership around development of reproducible sequencing-based assays for screening of 250+ compounds and 15+ mouse PK/PD studies for DM1
- Supervised and mentored one co-op student (H1 2021) who then transitioned to a Senior Research Associate position as my direct report (2022-2024)
- Broadened skillset to include analysis of long read sequencing, proteomics, and RNA structure prediction tools to support integrative subtarget nominations and characterization for RNA-targeting small molecules, nominating candidate RNA structures across ~100 targets
- Facilitated maturation of internal data systems including documentation of computational pipelines and processes and supporting development of a data lake
- Incorporated human genetics data into target assessments along with continuously updating omics data sources to support platform approaches spanning multiple mRNA-centric mechanisms

Scientist, Computational Biology — June 2019 – June 2021

- Multiomic (RNA-seq, TSS-seq, ribo-seq) analysis including differential expression and splicing, pathway and regulatory factor analysis of leaderboard compounds to characterize on- and off-target effects in relation to expected MoA and supporting Go/NoGo decisions on 4 programs
- Developed reproducible cloud-based pipelines for sequencing data types including RNA-seq, TSS-seq, RACE-seq, polyA-seq & ribo-seq. To date, processed 500+ sequencing datasets
- Data mining of large-scale consortium (ENCODE, FANTOM5, CCLE, DepMap, GTEx), literature, and internal omics and genetics datasets for novel target characterization and prioritization based on RNA features for ~40 internal targets and ~60 targets across Roche and Amgen collaboration projects
- Developed Python-based object-oriented API for extracting and characterizing alternative splicing from short read RNA-seq data. Applied to 1,709 RNA-seq datasets spanning CCLE and internal data collections to identify splicing vulnerabilities for small molecule treatments
- Point person for academic collaboration and software licenses for computational biology requirements around RNA splicing assessment and pathway analysis (Qiagen)

UNIVERSITY OF PENNSYLVANIA **PHILADELPHIA, PA**

PhD Student in Genomics & Computational Biology — August 2013 – May 2019

- Thesis research in Li-San Wang's lab involved integrative computational and experimental approaches for characterizing the regulatory mechanisms underlying noncoding genetic variation
- Lead development of open-source INFERNO method for integrating hundreds of functional genomics datasets to INFER the molecular mechanisms of NOncoding genetic variants, with web server: http://inferno.lisanwanglab.org
- Applied INFERNO to Alzheimer's Disease where I uncovered novel lncRNA-mediated regulatory mechanisms and performed luciferase validation of enhancer activity

- Applied INFERNO to a variety of phenotypes: neurodegenerative diseases (Parkinson's, PSP, ALS, CBD, FTD), psychiatric traits (schizophrenia, ADHD), and 2,419 UK Biobank phenotypes to nominate regulatory mechanisms underlying disease
- Implemented HiPPIE2 pipeline for Hi-C data analysis, from raw reads to high-resolution interacting sites, with functional annotation and identification of enhancer-promoter interactions
- Throughout my thesis work, I gained experience in analyzing many types of 'omics data including RNA-seq and pathway analysis, ChIP-seq, DNA-seq, ATAC-seq, and Hi-C. I also applied statistical genetics approaches including colocalization, haplotype deconvolution, and Bayesian QTL mapping
- My collaborative nature led me to many projects inside and outside of my thesis lab, spanning fields including neurodegenerative diseases (with Jerry Schellenberg, Virginia Lee, John Trojanowski, Eddie Lee), epigenetics (with Shelley Berger), and health economics (with Zeke Emanuel)

Education & Honors

PhD, Genomics & Computational Biology, University of Pennsylvania | 2013-2019 | Philadelphia, PA

- Coursework included Genetic Principles, Genome Regulation, Mathematical Statistics, Machine Learning, Statistical Methodology, Bayesian Statistics, and several Computational Biology courses
- Predoctoral Trainee in Age Related Neurodegenerative Diseases, NIH/NIA T32 AG00255
- Tutor for Genome Regulation course; TA and course material development for an Introduction to Bioinformatics course (https://github.com/greenelab/GCB535)

BA, Computer Science (High Honors) & Neuroscience, Oberlin College | 2009-2013 | Oberlin, OH

- Graduated with double major in Neuroscience and Computer Science with a thesis on algorithmic approaches to the multiagent traveling salesman problem. Graded and tutored for several CS courses
- Research in neurophysiology, computational psychology, evolutionary algorithms, neural networks, network routing algorithms, and wet and dry lab neurodegenerative disease research
- Summer undergraduate research assistant for two summers (2010, 2011) at the Center for Neurodegenerative Disease Research at UPenn with Virginia Lee and John Trojanowski.
- 2012 recipient of the competitive (~18%) DAAD RISE summer scholarship for network routing research at the Department of Computer Science, Humboldt-Universität zu Berlin, Berlin, Germany
- Recipient of the Nancy Robell Prize in Neuroscience and inducted into Phi Beta Kappa
- Recipient of the John F. Oberlin Merit Scholarship and the National Merit Scholarship for all 4 years

Technical Skills

Programming: R (tidyverse, Rshiny), Python (large-scale data processing and ML), Unix and bash, cloud computing (AWS, GCP), pipelining and reproducible research (Snakemake, Nextflow, Docker, Anaconda), Jupyter, github/bitbucket, Atlassian (Confluence, Jira), auto-documentation (Sphinx), databases (SQL)

Computational biology: genomics (WES/WGS), transcriptomics (RNA-seq, TSS/PolyA-seq), epigenomics (ChIP/CLIP-seq), long read sequencing (Nanopore), proteomics, integrative analysis of public (FANTOM5, DepMap, ENCODE, GTEx, UK Biobank) and proprietary datasets, differential expression and differential splicing analysis, network biology and pathway analysis (IPA, GSEA, WGCNA) with regulatory factor inference, knowledge graphs, ontologies, machine learning and statistics, RNA structure prediction

Genetics: GWAS, QTL mapping, post-GWAS analysis (colocalization, functional data integration), haplotype deconvolution, repeat expansion, RNA splicing, transcriptional regulation, noncoding elements

Drug development: target nomination and assessment, assay development and validation including sequencing-based technologies, biomarker discovery, PK/PD studies, mechanism of action-enabling studies, commercial/clinical assessment and patient stratification, RNA-targeting small molecules

Selected Presentations

1. Application of IPA for interpreting RNA-targeted small molecule treatment datasets. Qiagen IPA User Group Meeting, Waltham, MA (2022). Invited presentation.
2. Inferring the shared noncoding regulatory mechanisms underlying genetic susceptibility to Alzheimer's and Parkinson's diseases. International Conference on Alzheimer's and Parkinson's Diseases, Lisbon, Portugal (2019). Oral Presentation.

3. Inferring enhancer and noncoding RNA dysregulation underlying 2,419 UK Biobank Phenotypes. American Society of Human Genetics, San Diego, CA (2018). Conference Platform Talk.
4. Using INFERNO to characterize the effects of noncoding AD variants. Alzheimer's Disease Genetics Consortium (ADGC), Philadelphia, PA (2018). Consortium Meeting Presentation.
5. Integrative functional analysis identifies enhancers and lncRNAs perturbed by LOAD-associated genetic variants. ADGC, Cleveland, OH (2017). Consortium Meeting Presentation.
6. Integrative analysis identifies immune-related enhancers and lncRNAs perturbed by genetic variants associated with Alzheimer's disease. UPenn Institute on Aging Retreat, Philadelphia, PA (2017). Received first place poster award in basic science category.

Publications

* denotes first or co-first author; full information and poster abstracts at https://orcid.org/0000-0002-2073-1519

1. Kuksa PP, ..., **Amlie-Wolf A**, ..., Wang LS. FILER: a framework for harmonizing and querying large-scale functional genomics knowledge. NAR Genomics and Bioinformatics (2022)
2. **Amlie-Wolf A*,** ..., Wang LS. Using INFERNO to Infer the Molecular Mechanisms Underlying Noncoding Genetic Associations. Functional Analysis of Long Non-Coding RNAs (2021, Book Chapter)
3. Nativio R, ..., **Amlie-Wolf A,** ..., Berger S. An integrated multi-omics approach identifies epigenetic alterations associated with Alzheimer's disease. Nature Genetics (2020)
4. Kuksa PP, ..., Amlie-Wolf A, ..., Wang LS. SparkINFERNO: a scalable high-throughput pipeline for inferring molecular mechanisms of non-coding genetic variants. Bioinformatics (2020)
5. Kuksa PP, **Amlie-Wolf A***, ..., Wang LS. HIPPIE2: a method for fine-scale identification of physically interacting chromatin regions. NAR Genomics and Bioinformatics (2020)
6. Wheeler JM, ..., **Amlie-Wolf A**, ..., Schellenberg GD, Kraemer B. Activity of the poly(A) binding protein MSUT2 determines susceptibility to pathological tau in the mammalian brain. Science Translational Medicine (2019)
7. Liu EY, ..., **Amlie-Wolf A**, Lee EB. Loss of Nuclear TDP-43 Is Associated with Decondensation of LINE Retrotransposons. Cell Reports (2019)
8. Kunkle BW, ..., **Amlie-Wolf A**, ..., Schellenberg GD, Lambert JC, Pericak-Vance MA. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. Nature Genetics (2019)
9. **Amlie-Wolf A*,** ..., Wang LS, Schellenberg GD. Inferring the Molecular Mechanisms of Noncoding Alzheimer's Disease-Associated Genetic Variants. Journal of Alzheimer's Disease (2019)
10. **Amlie-Wolf A*,** ..., Brown CD, Schellenberg GD, Wang LS. INFERNO: INFERring the molecular mechanisms of NOncoding genetic variants. Nucleic Acids Research (2018)
11. Kuksa PP, **Amlie-Wolf A**, ..., Leung YY. SPAR: small RNA-seq portal for analysis of sequencing experiments. Nucleic Acids Research (2018).
12. Nativio R, ..., **Amlie-Wolf A**, ..., Berger S. Dysregulation of the epigenetic landscape of normal aging in Alzheimer's disease. Nature Neuroscience (2018)
13. Kuksa PP, **Amlie-Wolf A**, ..., Wang LS, Leung YY. DASHR 2.0: integrated database of human small non-coding RNA genes and mature products. Bioinformatics (2018)
14. Leung YY, Kuksa PP, **Amlie-Wolf A**, ..., Wang LS. DASHR: database of small human noncoding RNAs. Nucleic Acids Research (2016)
15. **Amlie-Wolf A*,** ..., Trojanowski JQ, Lee VM, Wang LS, Lee EB. Transcriptomic Changes Due to Cytoplasmic TDP-43 Expression Reveal Dysregulation of Histone Transcripts and Nuclear Chromatin. PLoS ONE (2015)