Conor R. Walker

Bioinformatics Scientist
Moderna, 325 Binney Street, Cambridge, MA 02142, USA
conor@conorwalker.net • Conor R. Walker • conorwalker • conor-walker

Summary

Bioinformatics scientist with 8+ years of experience specializing in developing statistical and machine learning models for sequence analysis and interpreting human 'omics data. I am highly motivated by advancing human genetic research with real world impact.

Research experience

Moderna *Aug. 2024–Present*

Scientist, Oncology Bioinformatics

- I am conducting research and developing analysis pipelines to support the individualized neoantigen therapy programme

New York Genome Center & Columbia University, USA *Jan. 2022–Aug. 2024*

Postdoctoral Research Associate

- Developed statistical methods and a deep learning pipeline (PyTorch) and for assessing privacy risks associated with releasing single-cell RNA-Seq datasets
- Implemented multiple classification models (scikit-learn) to predict genotypes from single-cell RNA-Seq, revealing privacy risks across many publicly available datasets
- Collaboratively designed a privacy-preserving bulk RNA-Seq quantification algorithm using homomorphic encryption, and processed hundreds of bulk RNA-Seq samples for benchmarking (Snakemake & AWS) – *MS in preparation*
- As part of a multidisciplinary team, I validated the activity of cis- and trans-regulatory elements in various cancer types that were identified using CRISPR screens (R) – *MS in preparation*

European Bioinformatics Institute (EMBL-EBI), UK *Oct. 2017–Jan. 2022*

Predoctoral Researcher

- Designed and implemented (C++) hidden Markov models to identify small-scale rearrangements in sequencing data, providing the first statistical method for identifying <50nt DNA rearrangements
- Applied my models to thousands of whole-genome human sequencing datasets to evaluate the evolutionary, population, cancer, and *de novo* landscape of template switch mutagenesis
- Implemented convolutional neural networks (Keras & PyTorch) to accurately detect positive selection within protein-coding sequences, validated by simulating evolution across tens of millions of genes
- Collaborated on multiple projects to characterize global SARS-CoV-2 genome evolution, analyzing thousands of samples (Python/Snakemake) and co-authoring 4 peer-reviewed papers

Newcastle University, UK *Nov. 2016–Sep. 2017*

Postgraduate Student Research

- Implemented multi-objective genetic algorithms for to optimize DNA sequences for use in computing systems, producing molecules with multiple desirable biochemical properties for *in vitro* validation

Education

University of Cambridge, UK *2017–2021*

PhD Computational Biology

- Thesis: Statistical analysis of short template switch mutations in human genomes

Newcastle University, UK *2016–2017*

MSc Bioinformatics · Distinction (83%), highest grade in cohort

- Thesis: Optimising nucleic acid sequences for DNA strand displacement systems

Liverpool John Moores University, UK *2013–2016*

BSc (Hons) Zoology · First-class honours (80%), highest grade in cohort

- Thesis: The genetics of insecticide resistance in the blackfly *Simulium vittatum*

Skills

Programming	Python, Bash, C++, R, AWK, LaTeX
Libraries	Numpy, SciPy, TensorFlow, pandas, matplotlib, seaborn, Dask, Biopython, pysam, scikit-allel, Scanpy, scvi-tools
Bioinformatics	Variant calling, sequence analysis, genome assembly, phasing & imputation, bulk & single-cell RNA-Seq analysis, eQTL mapping, phylogenetic inference, GWAS
Statistics & machine learning	Linear models, classical statistics, supervised/unsupervised learning (scikit-learn, XGBoost), deep learning (Keras, PyTorch)
Coding practices	Git, Nextflow, Snakemake, unit testing, Docker, Singularity, Jupyter, Vim
Cloud computing & HPC	AWS, Google Cloud, Slurm, LSF

Publications

Private information leakage from single-cell count matrices.
Walker C. R., Li X., Chakravarthy M., Lounsbery-Scaife W., Choi Y. A., Singh R., Gürsoy G.
Cell, 187, 1–13 (2024).

phastSim: Efficient simulation of sequence evolution for pandemic-scale datasets.
De Maio N., Boulton W., Weilguny L., **Walker C. R.**, Turakhia Y., Corbett-Detig R., Goldman N.
PLOS Computational Biology, 18, e1010056 (2022).

Mutation rates and selection on synonymous mutations in SARS-CoV-2.
De Maio N., **Walker C. R.**, Turakhia Y., Lanfear R., Corbett-Detig R., Goldman N.
Genome Biology and Evolution 13, evab087 (2021).

Short-range template switching in great ape genomes explored using pair hidden Markov models.
Walker C. R., Scally A., De Maio N., Goldman N.
PLOS Genetics 17, e1009221 (2021).

Stability of SARS-CoV-2 phylogenies.
Turakhia Y., De Maio N., Thornlow B., Gozashti L., Lanfear R., **Walker C. R.**, Hinrichs A. S., Fernandes J. D., Borges R., Slodkowicz G., Weilguny L., Haussler D., Goldman N., Corbett-Detig R.
PLOS Genetics 16, e1009175 (2020).

A phylodynamic workflow to rapidly gain insights into the dispersal history and dynamics of SARS-CoV-2 lineages.
Dellicour S., Durkin K., Hong S. L., Vanmechelen B., Martí-Carreras J., Gill M., Meex C., Bontems S., André E., Gilbert M., **Walker C. R.**, De Maio N., Faria N., Hadfield J., Hayette M., Bours V., Wawina-Bokalanga T., Artesi M., Baele G., Maes P.
Molecular Biology and Evolution 38, 1608–1613 (2020).

Masking strategies for SARS-CoV-2 alignments.
De Maio N., **Walker C. R.**, Borges R., Weilguny L., Slodkowicz G., Goldman N.
virological.org/t/masking-strategies-for-sars-cov-2-alignments/480 (2020).

Issues with SARS-CoV-2 sequencing data.
De Maio N., **Walker C. R.**, Borges R., Weilguny L., Slodkowicz G., Goldman N.
virological.org/t/issues-with-sars-cov-2-sequencing-data/473 (2020).