

## William Haynes Heaton, M.D., Ph.D.

840 Dakota Dr  
Auburn, AL 36832  
(256) 648-6432  
<https://github.com/wheaton5>  
[www.haynesheaton.com](http://www.haynesheaton.com)

### Education and Training

#### Cambridge University, Sanger Institute

- PhD Computational Biology 2017-2021
- Advisors: Richard Durbin and Mara Lawniczak
  - Projects:
    - \* Demultiplexing mixed-sample scRNAseq data by genotype without a genotype reference.
    - \* Haplotype resolved genome assembly via de novo phasing and long read haplotype binning.
    - \* Pedigree genome assembly strategies.

#### Brown University

- MD Brown Medical School 2007-2011
- ScB Computer Science 2003-2007
- Advisor: Sorin Istrail
- Student Researcher 2006-2007
- Computer vision/ML to classify malignancy of histology images in bladder cancer.

### Grants

Scott Furlan, **Haynes Heaton**. A compass for those with relapsed leukemia after transplant. NIH R01. Total Grant 1,753,638\$ over 5 years. Subaward for Haynes Heaton 247,077\$ over 3 years.

Tin Nguyen, **Haynes Heaton**. Innovative Approaches for Bias Correction & Systems-Level Analysis In Integrated Multi-omics data. 428,369\$ over 3 years.

### Papers

**Haynes Heaton**, Reza Behboudi, Collin Ward, Minindu Weerakoon, Sami Kanaan, Skylar Reichle, Nathan Hunter, Scott Furlan, “Collector: A tool to detect foreign genotype cells in scRNAseq data with applications in leukemia and microchimerism”. **Nature Communications** (submitted). 2026.

Minindu Weerakoon, Hai Vu, Reza Behboudi, **Haynes Heaton**, “Souporecell3: Robust Demultiplexing for High-Donor Single-Cell RNA-seq Datasets”. **Bioinformatics** (accepted). 2026.

Minindu Weerakoon, Sangjin Lee, Emily Mitchell, **Haynes Heaton**. “TopoQual polishes circular consensus sequencing data and accurately predicts quality scores”. **BMC Bioinformatics**. 2025.

Sandrine Eveline Nsango, Diego Ayala, Jean-Pierre Agbor, Harriet F Johnson, **Haynes Heaton**, ... Mara KN Lawniczak. “A chromosomal reference genome sequence for the malaria mosquito, *Anopheles nili*, Theobald, 1904”. **Wellcome Open Research**. 2024.

Lemondé BA Bouafou, Diego Ayala, Boris K Makanga, Nil Rahola, Harriet F Johnson, **Haynes Heaton**, ..., Mara KN Lawniczak. “A chromosomal reference genome sequence for the malaria

mosquito, *Anopheles maculipalpis*, Giles, 1902". **Wellcome Open Research**. 2024.

Boris K Makanga, Diego Ayala, Nil Rahola, Lemonde BA Bouafou, Harriet F Johnson, **Haynes Heaton**, ..., Mara KN Lawniczak. "A chromosomal reference genome sequence for the malaria mosquito, *Anopheles marshallii*, Theobald, 1903". **Wellcome Open Research**. 2024.

Minindu Weerakoon, Christopher T Saunders, **Haynes Heaton**. "LCSKPOA: Enabling banded semi-global partial order alignments via efficient and accurate backbone generation through extended lcsk++". **bioRxiv**. 2024. (pending BMC Bioinformatics review)

Tibebu Habtewold, Martin Wagah, Mgeni Mohamed Tambwe, Sarah Moore, Nikolai Windbichler, George Christophides, Harriet Johnson, **Haynes Heaton**, ..., Mara Lawniczak. "A chromosomal reference genome sequence for the malaria mosquito, *Anopheles gambiae*, Giles, 1902, Ifakara strain". **Wellcome Open Research**. 2024.

Diego Ayala, Ousman Akone-Ella, Pierre Kengne, Harriet Johnson, **Haynes Heaton**, ..., Mara Lawniczak. "The genome sequence of the malaria mosquito, *Anopheles funestus*, Giles, 1900". **Wellcome Open Research**. 2023.

Sandrine N Nsango, Jean-Pierre Agbor, Diego Ayala, Harriet F Johnson, **Haynes Heaton**, ..., Mara Lawniczak. "A chromosomal reference genome sequence for the malaria mosquito, *Anopheles moucheti*, Evans, 1925". **Wellcome Open Research**. 2023.

D Ayala, O Akone-Ella, P Kengne, H Johnson, **H Heaton**, ..., M Lawniczak. "The genome sequence of the malaria mosquito, *Anopheles funestus*, Giles, 1900". **Wellcome Open Research**. 2022.

**Haynes Heaton**, Arthur M Talman, Andrew Knights, Maria Imaz, Richard Durbin, Martin Hemberg, Mara Lawniczak. "souporecell: Robust clustering of single cell RNAseq by genotype and ambient RNA inference without reference genotypes". **Nature Methods**. 2020.

#### **Most viewed paper in Genes 2019 -**

Sarah B Kingan\*, **Haynes Heaton\***, Juliana Cudini, Christine C Lambert, Primo Baybayan, Brendan D Galvin, Richard Durbin, Jonas Korlach, Mara KN Lawniczak. "A high-quality de novo genome assembly from a single mosquito using PacBio sequencing." **Genes**, 2019.

Virginia M Howick, Andrew JC Russell, Tallulah Andrews, **Haynes Heaton**, Adam J Reid, Kedar Natarajan, Hellen Butungi, Tom Metcalf, Lisa H Verzier, Julian C Rayner, Matthew Berriman, Jeremy K Herren, Oliver Billker, Martin Hemberg, Arthur M Talman, Mara KN Lawniczak. "The Malaria Cell Atlas: Single parasite transcriptomes across the complete *Plasmodium* life cycle." **Science**, 2019.

Patrick Marks, Sarah Garcia, Alvaro Martinez Barrio, Kamila Belhocine, Jorge Bernate, Rajiv Bharadwaj, Keith Bjornson, Claudia Catalanotti, Josh Delaney, Adrian Fehr, Ian T Fiddes, Brendan Galvin, **Haynes Heaton**, et al. "Resolving the full spectrum of human genome variation using Linked-Reads." **Genome Research**, 2019.

Mark JP Chaisson, Ashely Sanders, Xuefang Zhao, 19 authors, **Haynes Heaton**, 65 authors, Evan E Eichler, Charles Lee. "Multi-platform discovery of haplotype-resolved structural variation in human genomes." **Nature Communications**, 2019.

Justin Zook, Jennifer McDaniel, Hemang Parikh, **Haynes Heaton**, Sean A Irvine, Len Trigg, Rebecca Truty, Cory Y McLean, Francisco M De La Vega, Marc Salit, Genome in a Bottle Consortium. "An open resource for accurately benchmarking small variant and reference calls."

**Nature Biotechnology**, 2019.

Grace Zheng, Billy T. Lau, Michael Schnall-Levin, 17 authors, **Haynes Heaton**, 34 authors, Serge Saxonov, Kevin D Ness, Benjamin J Hindson, Hanlee P Ji. "Haplotyping germline and cancer genomes with high-throughput linked-read sequencing." **Nature Biotechnology** (2016).

Vineeta Tanwar, Jeffery B. Bylund, Jianyong Hu, Jingbo Yan, Joel M. Walthall, Amrita Mukherjee, **William H. Heaton**, Wen-Der Wang, Franck Potet, Meena Rai, Sabina Kupersmidt, Ela W. Knapik, and Antonis K. Hatzopoulos. "Gremlin 2 promotes differentiation of embryonic stem cells to atrial fate by activation of the JNK signaling pathway." **Stem Cells**, 2014.

## Patents

Sofia Kyriazopoulou-Panagiotopoulou, Patrick Marks, Michael Schnall-Levin, Xinying Zheng, Mirna Jarosz, Serge Saxonov, Kristina Giorda, Patrice Mudivarti, Heather Ordonez, Jessica Terry, **William Haynes Heaton**. "Systems and methods for determining structural variation and phasing using variant call data". **US Patent #10854315**. 2020.

Goldstein, Peter, **William Heaton**, Franco Preparata, and Eli Upfal. "Distance maps using multiple alignment consensus construction." U.S. Patent Application 14/212,458, filed March 14, 2014.

## Invited talks

Haynes Heaton. "Using scRNAseq for multiplexing and microchimerism, a case study in leukemia." Fred Hutchison Cancer center 2024.

Haynes Heaton. "New approaches to generate single insect haplotype assemblies." Plant and Animal Genomes conference 2020.

Haynes Heaton. "De Novo phasing: toward haploid assemblies from diploid organisms without a trio." Darwin Tree of Life consortium workshop. 2019.

Haynes Heaton. "Novel genetic variation and validation using Linked Reads." Genome in a bottle consortium workshop. 2016.

## Poster presentations

Haynes Heaton. "Alignment and Variant Calling in Segmental Duplications with Linked-Read Data". Genome Informatics.

## Teaching

**Assistant Professor** Computational Biology: Genomics and transcriptomics Spring 2022, 2023, 2024, 2025, 2026.

**Assistant Professor** Graph Algorithms Fall 2022, 2023.

**Assistant Professor** Introduction to Algorithms Fall 2024, 2025, Spring 2025, 2026.

**Head Teaching Assistant** Computational Molecular Biology Fall 2006, 2007

**Teaching Assistant** Introduction to Computer Systems Fall 2005

**Head Teaching Assistant** Introduction to Scientific Computing Spring 2005, 2006

**Teaching Assistant** Introduction to Scientific Computing Spring 2004

## Industry Experience

**Senior Scientist**, 10X Genomics 2014-2017

Algorithm and software development on genomics platform bringing long range genetic information to nextgen short-read sequencing. We use a microfluidic system to attach the same barcode to every read originating from a long DNA molecule while different molecules get different barcodes with high probability.

- Developed new linked-read aligner “Lariat” which takes into account the molecule information when finding its mapping. This produces fewer mismapped reads and is able to map into many repeat regions of the genome with high confidence. (lead on this project)
- Invented novel phasing probabilistic model which is able to filter variants that do not haplotype consistent.
- Head of short variant calling and data analysis, metrics, ground truth analysis of short variants.
- Created a system of haploid variant calling to improve sensitivity.
- Worked with biochemists and chemists to create data metrics that allow them to continuously improve the data quality.

**Senior Software Engineer: Scientific Computing**, GNS Healthcare, Cambridge MA 2013 - 2014

Part of a team working on causal bayesian network machine learning with MCMC to sample graph structure of the bayesian nets.

- Contributed to methods in post learning simulation, analysis, and clustering.
- Developed and supported Amazon EC2 execution of our platform using Starcluster.
- Created distribution process for our post learning simulation using Hadoop on Amazon Elastic map reduce.

**Computational Scientist**, Nabsys, Providence RI 2011 - 2013

Developed algorithms for a nano channel DNA mapping startup. Collaborated with Biochemistry and Electrical Engineering teams as well as consulting CS professors to make novel methods addressing data produced by a unique assay – long (10s-100s of kilobases) DNA fragments with tag molecules attached to sequence specific sites are run through a solid-state nano-detector, creating data analogous to ordered restriction maps (or bionano genomics data).

- Developed genetic distance map de novo assembly software with computational biologist Peter Goldstein.
- Created novel multiple alignment using a probabilistic, graph theoretic approach. This allowed us to reduce error through averaging distances and consensus voting over multiple measurements. (Lead on this project. Patented, with Peter Goldstein, Computational Biologist Nabsys)
- Wrote signal processing software employing standard EE methods as well as HMMs and watershed algorithm for feature extraction. Created interactive data visualization package. (Lead on this project managing two employees.)

## Software Engineering

	Expert	Proficient
Languages	Rust, Python, R, Go, Java	C/C++, Javascript (D3.js), Matlab/Octave

## Techniques and skills

*Machine Learning* - linear/logistic with various regularizers, SVMs/SVRs with kernels, neural nets (DNN, CNN, RNN, deep autoencoders etc), causal Bayesian nets, decision/regression trees/random forrest, Gaussian Processes, tensorflow

*Dimensionality Reduction* - PCA, SVD, random projections/sparse random projections, tSNE, deep autoencoders

*Clustering* - K-means, hierarchical, K-nearest neighbors, mixture of gaussians, graph clustering (spectral, markov clustering), fuzzy c-means, GMMs / other mixture models including sparse mixture

models.

*Temporal/Series Pattern Recognition* - HMMs (viterbi, forward/backward, Baum-Welch), Fourier transform *Combinatorial optimization* - Metropolis Hastings, MCMC, simulated annealing, dynamic programming, branch and bound *Statistical inference* - variational inference, MCMC with STAN, tensorflow probability, edward