

## William Haynes Heaton, M.D.

Grange Rd. Selwyn College  
Cambridge, Cambridgeshire, CB3 9DQ  
+44 7955 343268  
<https://github.com/wheaton5>  
[www.haynesheaton.com](http://www.haynesheaton.com)

**Education** PhD (current) Sanger Institute, Cambridge University, Cambridge, UK  
MD May 2011 Brown Medical School, Providence, RI  
Computer Science May 2007 Brown University, Providence, RI

**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. This data type is not dissimilar from Moleculo, complete genomics LFR, or illumina CPT seq but since it is droplet based instead of plate based, we are able to have millions of different partitions/barcodes instead of thousands.

- Developed new linked-read aligner “Lariat” based on BWA-mem 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 segregate correctly on haplotype lines.
- 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.

**Sr 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**

**Languages**

**Expert**

**Proficient**

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

**Techniques and skills**

**Machine Learning**

*Classification/Regression* - 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

*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)

*Temporal/Series Pattern Recognition* - HMMs (viterbi, forward/backward, Baum-Welch), Fourier transform

**Bioinformatics tools**

*Visual analysis* - IGV, dot plots, UCSC browser

*tools* - samtools, bcftools, bedtools, vcftools, vcfallelicprimitives, freebayes, gatk, picard, pysam, pyVcf, bwa, blat, blast, STAR, hisat2, cellranger, longranger, supernova, among many others

**Academic Research**

**Student Researcher**, Brown University

2006 - 2007

- Algorithmic Cancer Diagnosis. Computer vision techniques and machine learning to classify histology images of bladder cancer into normal, low malignancy, and high malignancy. (with Sorin Istrail, Professor of CS, Brown)

**Visiting Researcher**, Vanderbilt University

Summers 2005-2008

- Researched Bone Morphogenic Protein antagonist regulation of differentiation of embryonic stem cells in to various cardiomyocyte lineages. Wet lab genetics including cell culture, PCR, western blot, immunohistochemistry.

**Teaching**

**Teaching Assistant** Introduction to Scientific Computing

Spring 2004

**Head Teaching Assistant** Introduction to Scientific Computing

Spring 2005, 2006

**Teaching Assistant** Introduction to Computer Systems

Fall 2005

**Head Teaching Assistant** Computational Molecular Biology

Fall 2006, 2007

**Papers**

Zheng, Grace XY, Billy T. Lau, Michael Schnall-Levin, Mirna Jarosz, John M. Bell, Christopher M. Hindson, Sofia Kyriazopoulou-Panagiotopoulou et al. "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*.

Mark JP Chaisson et al. "Multi-platform discovery of haplotype-resolved structural variation in human genomes." Preprint. *bioRxiv*, 2017, 193144.

Patrick Marks, Sara Garcia, Alvaro Martinez Barrio, Kamila Belhocine, Jorge Bernate, Rajiv Bharadwaj, Keith Bjornson, Claudia Catalanotti, Josh Delaney, Adrian Fehr, Brendan Galvin, Haynes Heaton et al. "Resolving the Full Spectrum of Human Genome Variation using Linked-Reads." Preprint. *bioRxiv*. 2017, 230946.

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. "Reproducible integration of multiple sequencing datasets to form high-confidence SNP, indel, and reference calls for five human genome reference materials." Preprint. *bioRxiv*, 2018, 281006.

#### **Patents**

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.

William Heaton, Patrick Marks. "Handling Non-heterozygous variants in haplotype phasing and filtering." (Pending)

Kyriazopoulou-Panagiotopoulou, S., Marks, P., Schnall-Levin, M., Zheng, X., Jarosz, M., Saxonov, S., ... & Heaton, W. H. (2016). U.S. Patent Application No. 15/019,928.

#### **Posters and invited talks**

Haynes Heaton, Patrick Marks, Matt Sooknah, Sofia Kyriazopoulou-Panagiotopoulou, Sarah Garcia, Brendan Galvin, Deanna Church, Michael Schnall-Levin. "Alignment and Variant Calling in Segmental Duplications with Linked-Read Data". *Genome Informatics*.

Haynes Heaton, Patrick Marks, Deanna Church. "Novel genetic variation and validation using Linked Reads." *Genome in a bottle consortium workshop*. 2016.