Henry Bigelow

55A Brosnan St
San Francisco, CA, 94103

hrbigelow@gmail.com
+1.917.692.8394

OBJECTIVE

Software Engineer and Artificial Intelligence Research position

EDUCATION

- **Columbia University** New York, NY
  Ph.D. Biochemistry & Molecular Biophysics *April 2007*
  - **Thesis Title**: *Membrane protein structure prediction with Bayesian Networks*
    Awarded Distinction (top 10%)
    M.S. April 2002, M.Phil. April 2001
- **Oberlin College** Oberlin, OH
  B.A. Biology *May 1997*

ADVANCED SKILLS & AREAS OF STUDY

- **Programming Languages & API**: C99, C++, Python (TensorFlow, PyTorch), Rust, CUDA, POSIX Threads, Linux API, Oracle, PL/SQL, PostgreSQL, PL/pgSQL, MySQL, Docker, HTML5, Javascript, PHP, OCaml, Perl, Lex/Yacc, LATEX, OpenGL
- **Infrastructure & Services**: Cloud (AWS EC2/S3, Google Cloud SQL, Compute Engine, Cloud Storage), Docker, Isilon vs NetApp read/write speed tests, VM vs. Bare Metal server memory bandwidth and Ethernet card speed tests, large volume (300 TB) data management, REST Web Service design and maintainance, SSL Certificate generation, Wiki installation
- **Machine Learning**: Neural Networks, Bayesian Networks, Factorial Hidden Markov Models, Baum Welch, Maximum Likelihood, Belief propagation, Junction trees, Metropolis-Hastings, Multivariate Gradient-based minimization
- **Deep Learning Reading**: Deep Learning (Goodfellow et al.), Linear Algebra (Shilov), Elements of Information Theory (Cover & Thomas), Sparse Distributed Memory (Kanerva), Probability Theory (Jaynes), Variational Principles of Mechanics (Lanczos), Convex Optimization (Boyd et al.), The Elements of Statistical Learning (Hastie et al.)
- **Programming Reading**: CUDA By Example (Kandrot), Programming with POSIX Threads (Butenhof), The Linux Programming Interface (Kerrisk), Advanced Programming in the UNIX Environment (Stevens et al.), Introduction to Algorithms (Cormen et al.), The C Programming Language (Kernighan et al.), Handbook of Floating-Point Arithmetic (Miller et al.)
- **Computational Biology**: Ultradeep viral genome diversity estimation and rare variant detection, Automated genomic variant annotation, Membrane protein structure prediction, Transcription factor binding site detection, Comparative genomics

SOFTWARE WRITTEN

- **https://github.com/hrbigelow/ae-wavenet**: implementing WaveNet Autoencoder from https://arxiv.org/abs/1901.08810 in PyTorch. 100+ followers, 10 forks. Training on V100 and TPU with pytorch-xla
- **https://github.com/hrbigelow/lb-wavenet**: implemented WaveNet in TensorFlow using lookback technique for flexible batching of different-length wav files in parallel pipelined training. Uses tf.data API. Can run using eager or graph mode. Trained on Tesla K80
- **https://hrbigelow.github.io**: Machine Learning Blog
- **Deep learning in PyTorch**: Implemented MNIST digit autoencoder using Local Winner-Take-All neural competition mechanism.
- **https://github.com/hrbigelow/depngs**: Multi-threaded next-gen sequence library for BAM parsing, sequence pileup and probabilistic genotyping.
- **https://github.com/hrbigelow/wij-wemnan**: JavaScript FM-index implementation for client-side auto-suggest with large lists.
- **https://sourceforge.net/projects/maqweb**: Web-based read-mapping and variant annotation pipeline. Variants classified by protein coding effect, proximity to UTRs, ncRNAs, known SNPs. Used by over 50 labs.
- **ProfTMB**: HMM-based Bacterial transmembrane *beta*-barrel structure prediction.
- **https://sourceforge.net/projects/cisortho/**: Genome-wide transcription factor binding site search in nematode genomes using position weight matrices and comparative genomics.

PROFESSIONAL EXPERIENCE

- **Software Startup (Undisclosed)** San Francisco, CA
  First employee, Chief Software Architect reporting to CEO *June 2019 - Oct 2019*
  - **Product Backend Design and Implementation**: Designed and implemented the entire backend for the product. This included PostgreSQL & PL/pgSQL DataMart connecting to a Rust web server supporting GraphQL Queries and Mutations, email-based new user registration flow, and PubSub-based messages for application events.
  - **High-level Product conceptual design**: Numerous discussions with CEO and Founder to diagram the product and define requirements in a reliable, scalable, and maintainable way.
- **Deep Learning Independent Study** San Francisco, CA
  Self Study *June 2017-present*
  - **Foundational Reading and Coursework**: Studied Deep Learning Literature full time independently. Thoroughly read Deep Learning (Goodfellow) twice, Linear Algebra (Shilov), Sparse Distributed Memory (Kanerva) and Probability Theory (Jaynes). Completed Andrew Ng's Coursera course. Watched all of Hinton's Youtube video lectures.
  - **Literature and Implementations**: Implemented WaveNet in TensorFlow. Implemented autoencoder in PyTorch. Studied papers on CNN, RNN, LSTM, MDL, Autoencoders, Variational approximations, Bayesian NNs. Attended Deep Learning Meetup reading group, met authors of Image Transformer.
- **Amgen** South San Francisco, CA
  Genome Analysis Unit *June 2010 - Dec 2016*
  Scientist Level 5 for Dr. Sasha Kamb
  - **Genomic Analysis Software**: Wrote software to perform SNP and indel mutational distance analysis of Leukemic B cells across pairs of samples in BlinCyto responder/non-responder study. Analyzed ploidy and zygosity across whole genomes of 30 TB of WGS data (35 samples). Attempted to discover genetic mechanisms of resistance. Selected slides presented to CEO.
  - **LIMS Data Mart**: Designed and maintained the group's central sample tracking / LIMS Oracle database. Implemented PL/SQL views and ETL procedures, Tomcat/JDBC web service and HTML frontend for generating Excel reports.
  - **Gene Expression Data Mart**: Designed Oracle PL/SQL datamart for viewing gene expression data from RNA-Seq, Fluidigm (qPCR) and GeneLogic microarray, identifying differentially and stably expressed genes. Wrote Spotfire visualizations regularly used by many colleagues in Oncology and other Therapeutic Areas.
- **Broad Institute** Cambridge, MA
  Computational R&D *May 2009 - June 2010*
  Computational Biologist for Dr. David Jaffe
  - **Mutational Inference**: Implemented Bayesian causal model for inferring nucleotide composition at diverse intrahost viral loci from aligned sequence reads. Estimates base composition with confidence intervals by sampling from the posterior distribution.
- **Columbia University** New York, NY
  Department of Biochemistry and Molecular Biophysics *May 2007 - May 2009*
  Postdoctoral Research Scientist for Dr. Oliver Hobert
  - **CisOrtho Transcription Factor Binding Site Search**: Implemented multi-genome transcription factor binding site search software for C. elegans and related nematode genomes. Allows flexible specification of binding site patterns, clustering criteria, filtering based on homolog/ortholog conservation and existing genome annotation data. Implements prefix-tree based ungapped pattern search supporting multiplexing and approximate match.
- **University of Chicago** Chicago, IL
  Ben May Department for Cancer Research *June 1997 - Nov 1998*
  Research Technician for Dr. Akira Imamoto
  - **Src Knockout Study**: Engineered transgenic mice to study effects of Src on neural differentiation. Designed and constructed a cre-lox switchable vector, transfected into mouse Embryonic Stem (ES) cells, confirmed genomic integrants via Homologous Recombination by PCR of cultured ES cells and activity by biochemical assays. Restriction mapped several large genomic constructs.

- **UCLA** Los Angeles, CA
  Department of Biochemistry and Molecular Biology *Feb 1999 - June 2000*
  Research Technician for Dr. Jay Gralla and Dr. Robert Clubb
  - **Protein Structure studies**: Worked towards solving structure of $\sigma54$ transcription initiation factor domains. Engineered, cloned, expressed and affinity purified several deletion constructs. Collected and analyzed Nuclear Magnetic Resonance and Circular Dichroism spectra and performed initial crystallization trials towards solving 3D structure. Performed Electrophoretic Mobility Shift and other binding assays to determine binding partners.

Honors & Awards

Distinction for Ph.D. Dissertation (top 10%)
98^{th} Percentile in Math Section of GRE exam
Predoctoral fellowship from Columbia University

Talks

Workshop: Whole Genome Sequencing for Mutant Identification. Presented at: The 17^{th} International *C. elegans* Meeting. June 27, 2009. Grand Horizon Ballroom, UCLA.

Papers

Bigelow H, Doitsidou M, Sarin S, Hobert O. MAQGene: software to facilitate C. elegans mutant genome sequence analysis. *Nature Methods* **6(8)**, 549 (2009).

Sarin S, Bertrand V, Bigelow H, Boyanov A, Doitsidou M, Poole R, Narula S, Hobert O. Analysis of Multiple Ethyl Methanesulfonate-Mutagenized Caenorhabditis elegans Strains by Whole-Genome Sequencing. *Genetics* **185(2)**, 417-430 (2010).

Henn MR, Lennon NJ, Newman R, Charlebois P, Boutwell C, OhAinle M, Berlin AM, Ryan E, Malboeuf C, Macalalad A, Casali M, Erlich R, Bigelow HR, Green L, Gnerre S, Young S, Levin J, Nusbaum C, Walker BD, Diamond MS, Kramer LD, Ebel GD, Harris E, Allen TA, Birren BW. Sensitive population profiling and genome assembly of HIV and Flaviviruses using ultra-deep sequencing technologies. *Genome Biology* **11(Suppl 1)**, 18 (2010).

OMeara MM, Bigelow H, Flibotte S, Etchberger JF, Moerman DG, Hobert O. Cis-regulatory mutations in the Caenorhabditis elegans homeobox gene locus cog-1 affect neuronal development. *Genetics* **181(4)**, 1679-1686 (2009).

Punta M, Forrest LR, Bigelow H, Kernytsky A, Liu J, Rost B. Membrane Protein Prediction Methods. *Methods* **41(4)**, 460-474 (2007).

Bigelow H, Rost B. PROFtmb: a web server for predicting bacterial transmembrane beta barrel proteins. *Nucleic Acids Res* **1;34** 186-188 (2006).

Bigelow HR, Petrey DS, Liu J, Przybylski D, Rost B. Predicting transmembrane beta-barrels in proteomes. *Nucleic Acids Res* **32(8)**, 2566-2577 (2004).

Bigelow HR, Wenick AS, Wong A, Hobert O. CisOrtho: a program pipeline for genome-wide identification of transcription factor target genes using phylogenetic footprinting. *BMC Bioinformatics* **12;5** 27 (2004).

Bigelow HR, Rost B. Online tools for predicting integral membrane proteins. *Membrane Proteomics: Methods and Protocols, from Methods in Molecular Biology* **528**, 3-23 (2009).

B Rost, J Liu, D Przybylski, R Nair, KO Wrzeszczynski, HR Bigelow and Y Ofran. Prediction of protein structure through evolution. *Handbook of Chemoinformatics - From Data to Knowledge* 1789-1811 (2003).

Other Interests

4 years Japanese study (95% of JLPT Kanji, Dictionary of Basic/Intermediate Japanese Grammar, radio listening)

Violin making and other woodworking, Violin, Piano, Clarinet, Tennis, Golf, Basketball
Violinist in the Columbia University Orchestra for four years. Performed in two Commencements
Violin section leader in two University productions - West Side Story (2005) & Guys and Dolls (2006)
Member of New York Academy of Sciences, International Society for Computational Biology